

KANT AND MORAL RESPONSIBILITY

CARL H. HILDEBRAND

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the MA degree in Philosophy

Department of Philosophy
Faculty of Arts
University of Ottawa

© Carl H. Hildebrand, Ottawa, Canada, 2012

ABSTRACT

This project is primarily exegetical in nature and aims to provide a rational reconstruction of the concept of moral responsibility in the work of Immanuel Kant, specifically in his *Critique of Pure Reason (CPR)*, *Groundwork of the Metaphysics of Morals (GR)*, and *Critique of Practical Reason (CPrR)*. It consists of three chapters – the first chapter interprets the idea of freedom that follows from the resolution to the Third Antinomy in the *CPR*. It argues that Kant is best understood here to be providing an unusual but cogent, compatibilist account of freedom that the author terms meta-compatibilism. The second chapter examines the *GR* and *CPrR* to interpret the theory of practical reason and moral agency that Kant develops in these works. This chapter concludes by evaluating what has been established about Kant’s ideas of freedom and moral agency at that point in the project, identifying some problems and objections in addition to providing some suggestions for how Kantian ethics might be adapted within a consequentialist framework. The third chapter argues that, for Kant, there are two necessary and jointly sufficient conditions (in addition to a compatibilist definition of freedom) that must obtain for an individual to qualify as responsible for her actions.

There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy.

William Shakespeare in *Hamlet* (Act 1: Scene 5)

Two things fill the mind with ever new and increasing admiration and reverence, the more often and more steadily one reflects on them: the starry heavens above me and the moral law within me. I do not need to search for them and merely conjecture them as though they were veiled in obscurity or in the transcendent region beyond my horizon; I see them before me and connect them immediately with the consciousness of my existence. The first begins from the place I occupy in the external world of sense and extends the connection in which I stand into an unbounded magnitude with worlds upon worlds and systems of systems, and moreover into the unbounded times of their periodic motion, their beginning and their duration. The second begins from my invisible self, my personality, and presents me in a world which has true infinity but which can be discovered only by the understanding, and I cognize that my connection with that world (and thereby with all those visible worlds as well) is not merely contingent, as in the first case, but universal and necessary. ...It infinitely raises my worth as an intelligence by my personality, in which the moral law reveals to me a life independent of animality and even of the whole sensible world, at least so far as this may be inferred from the purposive determination not restricted to the conditions and boundaries of this life but reaching into the infinite.

Immanuel Kant in *The Critique of Practical Reason* (5:161-2)

TABLE OF CONTENTS

INTRODUCTION	1
CHAPTER I	7
1.1 The Third Antinomy	8
1.1.1 <i>Outline of the Third Antinomy</i>	8
1.1.2 <i>Resolution of the Third Antinomy</i>	13
1.1.3 <i>Problems with the Third Antinomy</i>	17
1.2 Meta-Compatibilism: An Interpretive Solution	22
1.2.1 <i>An Outline of Meta-Compatibilism</i>	22
1.2.1.1 <i>Practical Freedom and Causal Determinism Together</i>	24
1.2.1.2 <i>Limiting Cognitive Capacities</i>	26
1.2.1.3 <i>Rationally Licensed Beliefs</i>	28
1.2.2 <i>Meta-Compatibilism and the Problems of the Third Antinomy</i>	32
1.3 A Significant Objection to Meta-Compatibilism	36
CHAPTER II	40
2.1 The Groundwork and the Idea of the Moral Law	41
2.2 From Meta-Compatibilism to the Moral Law	55
2.2.1 <i>Freedom and Choice</i>	56
2.2.2 <i>Awareness of Freedom Through the Moral Law</i>	60
2.2.3 <i>Freedom and Morality: Transition From the CPR to the GR and CPrR</i>	64
2.3 Evaluating Kant	68
2.3.1 <i>Objection to Freedom as a Useful Concept</i>	68
2.3.2 <i>A Consequentialist Adaption of Kant</i>	73
2.3.3 <i>The Problem of the Unknowability of Maxims</i>	81
2.3.4 <i>Self-Interest and Morality</i>	83
CHAPTER III	86
3.1 Introducing the Conditions of Accountable Agency	86
3.2 Internal Motivation	91
3.2.1 <i>Autonomy and the Moral Law</i>	92
3.2.2 <i>Internal and Rational Incentives</i>	98
3.3.3 <i>Some Objections to Kant's Account of Internal Motivation</i>	103
3.3 Objective Justification	107
3.3.1 <i>The Form of Universality</i>	108
3.3.2 <i>Externally Evaluable Maxims</i>	110
3.3.3 <i>Some Objections to Objective Justification in Kant</i>	112
3.4 Conclusion	114

INTRODUCTION

The aim of this project is to achieve an understanding of the concept of moral responsibility as it is found in Immanuel Kant's work in the *Critique of Pure Reason* (*CPR*), *Groundwork of the Metaphysics of Morals* (*GR*), and *Critique of Practical Reason* (*CPrR*). I will argue that in attempting to assemble a coherent concept of moral responsibility therein, we encounter a fruitful account of the conditions for moral responsibility that effectively permits us (human beings) to regard ourselves as rational and moral agents who are accountable to one another and capable of interacting with one another accordingly. I suggest that with this concept of human agency and moral responsibility, Kant aims to preserve a traditional, practical, and common sense idea of agency and moral responsibility, ultimately tying them in with the three postulates of God, freedom, and immortality that he points to in the preface to the second edition of the *CPR*.¹

I am not the first to argue for this sort of interpretation of Kant. In a chapter on responsibility in personal relations, Christine Korsgaard points out that this suggestion that Kant ties the idea of responsibility to these three postulates is evident in the *CPrR* and she develops an interpretation of his idea of moral responsibility that fits

¹ Immanuel Kant, *Critique of Pure Reason*, trans. and eds. Paul Guyer and Allan Wood (Cambridge: Cambridge University Press, 2008), Bxxx. All citations of the *CPR* will use the standard "A" and "B" page numbers. Citations for all of Kant's work will use the volume and page numbers of the Prussian Academy edition rather than the page numbers of the particular translation. This includes the *GR* and *CPrR* in addition to the *CPR*. Also, I will not use the bold or italics that is found in the translations of Kant's work, so any italics or bold print in quotations of Kant's work have been put there by myself in order to highlight aspects of his thought salient to my argument.

accordingly.² I need to admit a significant debt to Korsgaard for clarifying Kant's overall idea of moral responsibility for me, which was especially important in light of the problem of maxims that I wrestled with in writing Chapters II and III. However, though the idea of a composite concept of assessment that I suggest in Section 2.3.2 of Chapter II is inspired by her and aligns with her practical account of responsibility in Kant, the rest of this project is quite separate from hers and, I hope, as original as I intuited it before encountering her analysis of responsibility. Nonetheless, it remains important that I mention this debt. Stephen Engstrom argues for a similar interpretation of Kant in his recent book *The Form of Practical Knowledge*. He locates Kant in what he calls a "cognitivist tradition in practical philosophy"³ alongside such figures as Thomas Aquinas.⁴ He claims: "standard interpretations have exaggerated the extent of Kant's break with his precursors... [because] Kant never departs from the idea that practical reason is a capacity for knowledge of the good."⁵ In a display of roundabout admiration, Allen Wood also affirms that though it seems counter-intuitive at first, it is pleasantly surprising to find that "Kant's compatibilism involves no greater revision of our commonsense view of our agency than it does."⁶ Wood also argues that Kant "wants to reserve for himself a more Aristotelian [rather than Humean] notion of causal efficacy, to be ascribed to free agents as members of the intelligible world."⁷ I owe a debt to Wood

² Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), 208.

³ Stephen Engstrom, *The Form of Practical Knowledge* (Cambridge, MA: Harvard University Press, 2009), vii.

⁴ *Ibid.*, xi.

⁵ *Ibid.*

⁶ Allen Wood, "Kant's Compatibilism" in *Self and Nature in Kant's Philosophy* ed. Allan Wood (Ithaca: Cornell University Press, 1984), 76.

⁷ *Ibid.*, 88.

for first showing me how compatibilism may be understood within Kant before developing my own interpretation of Kant's theory of freedom in Chapter I. I too would like to locate Kant's theory of practical reason, agency, and moral responsibility within this more traditional and, I suggest, common sense lineage.

Over the course of this project, I will define three necessary and jointly sufficient conditions for moral responsibility found in Kant's *CPR*, *GR*, and *CPrR*. The first of these is a compatibilist account of free will/freedom. I understand this to be a necessary precondition for anything like an intuitive or common sense idea of agency to obtain in the first place and I believe that Kant did as well. In Kant's case, the concern with freedom arises as a result of the causal determinism he believes is necessary for the unity of consciousness.⁸ The second of these conditions for moral responsibility is internal motivation—in short, the idea that human beings are agents who possess the capacity to reason and act on their own abilities, free from alien or external forces of compulsion. This second condition is consonant with and inspired by Aristotle's idea in the *Nicomachean Ethics* that a morally responsible agent is one who possesses the capacity for deliberation and decision in addition to the capacity for voluntary action.⁹ This condition also aligns with the idea of voluntary agency that Thomas Aquinas communicates in the *Summa Theologica*. Here voluntary agency consists in the knowledge of and internal capacity to act in accordance with an end.¹⁰ The third and final

⁸ This concern is evident in the Second Analogy (*CPR*, B234).

⁹ Aristotle, *Nicomachean Ethics Book III. Chapters 1-5* as cited in Andrew Eshlemen, "Moral Responsibility" in *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/moral-responsibility/>.

¹⁰ Thomas Aquinas, *Summa Theologica, First Part of the Second Part, Question VI* as cited in Derk Pereboom ed., *Free Will* (Indianapolis: Hackett Publishing Company, 2009), 43.

condition for morally responsible agency in Kant is what I call objective justification—the awareness of an objective, inter-personal moral standard to which all morally responsible agents are accountable. To conduct an extensive comparison between these conditions for moral responsibility in Kant and these other thinkers would be beyond the scope of the current project, so in the meantime, I only suggest this similarity, admit it as a source of inspiration, and leave the reader to contemplate the extent of its correspondence.

It is necessary to clarify early on that Kant's work in the *CPR*, *GR*, and *CPrR* does not admit of a precise, fact-based concept of moral responsibility as attributability as defined by Andrew Eshlemen in his entry "Moral Responsibility" in the *Stanford Encyclopedia of Philosophy*. This is because for Kant the question of moral responsibility is not focused narrowly on placing praise and blame in particular circumstances. As Christine Korsgaard points out, for Kant, moral responsibility arises from the first-person point of view, which means that it is more an issue of taking responsibility than of assigning it.¹¹ Onora O'Neill affirms this idea as well in her earlier book on Kant *Acting on Principle: An Essay on Kantian Ethics*. She openly admits that at least as of 1975 (when the book was written), there existed no solution to "the problem of relevant [act] descriptions" but goes on to say "Kant's theory of right can at least be used in contexts of decision and action. And it is these contexts which are of most importance for the moral life."¹² These circumstances do not seem to have changed.

¹¹ Christine Korsgaard, *Creating the Kingdom of Ends*, 189.

¹² Onora O'Neill, *Acting on Principle: An Essay on Kantian Ethics* (New York: Columbia University Press, 1975), 143.

Interestingly, the concept of attributability¹³ of (one could say, moral responsibility for) specific acts does not occupy a central role in Kant's moral theory. Instead, his focus is on the specific conditions that must be in place in order for an agent to qualify as morally responsible in the more general sense. He is concerned with what it is about human beings that makes them morally responsible agents of the sort that can hold each other accountable in the first place. Christine Korsgaard highlights this feature of Kant's thought and contrasts it with the tradition of British Empiricism.

In the British Empiricist tradition, the concept of responsibility has been closely associated with the ideas of praise and blame, and these in turn have played a central role in its moral philosophy. In the theories of Hutcheson, Hume, and Smith, the approval and disapproval of others is the fundamental moral phenomenon, from which all our moral ideas spring. There is something obviously unattractive about taking the assessment of others as the starting point in moral philosophy. One of the appealing things about Kant's ethics, by contrast, is that in it moral thought is seen as arising from the perspective of the agent who is deciding what to do. Responsibility in the first instance is something taken rather than something assigned. And this fact about the structure of his view is complemented by a fact about its content. Kant is not very interested in praise and blame and seldom mentions them.¹⁴

So it is obvious that Kant is not concerned with the issues of praise and blame in particular circumstances but rather in the idea of morally responsible agency in general and the fact that we can regard ourselves and each other as having that sort of agency. The idea that we can do this is what I mean when I say that Kant *licenses* the ideas of freedom and moral responsibility. For example, against the background of his metaphysical and epistemological project of transcendental idealism laid out in the *CPR*, his account of agency does not *prove* the fact of our freedom—it is not even decisive as

¹³ As outlined and contrasted with 'accountability' by Andrew Eshlemen: Andrew Eshlemen, "Moral Responsibility" in *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/moral-responsibility/>.

¹⁴ Christine Korsgaard, *Creating the Kingdom of Ends*, 189.

to whether there is a fact of freedom to be known. This is because the framework of transcendental idealism does not grant that we can know anything about the fact of our freedom in the first place, let alone our responsibility for particular acts.

However, it is evident that Kant's views in the later *GR* and *CPrR* change on the matter. In these later works devoted to practical reason and morality, he expounds the view that we can in fact know about the moral law—that it exists in the noumenal realm and that we can act freely in accordance with it. This is not so much a troubling inconsistency as a development that Kant undergoes when turning his hand to practical reason and morality, and away from the concerns of the First Critique (the *CPR*) with speculative philosophy. Despite this shifting view on freedom in the later *GR* and *CPrR*, I contend that from all three works a coherent picture of agency and the conditions for moral responsibility does emerge. It is a picture that upholds and furthers human dignity and the respect we ought to show one another and therefore worthwhile to understand.

CHAPTER I

[The] intelligible ground does not touch the empirical questions at all, but may have to do merely with thinking in the pure understanding; and, although the effects of this thinking and acting of the pure understanding are encountered among appearances, these must nonetheless be able to be explained perfectly from their causes in appearance, in accord with natural laws, by following its merely empirical character as the supreme ground of explanation; and the intelligible character, which is the transcendental cause of the former, is passed over as entirely unknown.¹⁵

The Third Antinomy of the *CPR* provides the crucial theoretical background to Kant's metaphysics of freedom. Here he places what he believes are two airtight arguments—one for causality in accordance with freedom and one for causality in accordance with complete causal determinism—in opposition to one another in order to make room for the possibility of free-will in terms of agent causation. Over the course of Chapter I here, I will offer a detailed discussion of the Third Antinomy and the role that it plays in Kant's metaphysics of freedom. This will include a close look at the structure of the Third Antinomy, how the metaphysical and epistemological features of transcendental idealism contribute to Kant's solution of the Third Antinomy and a number of strong objections to that solution. Following that, I will argue that a compatibilist reading of the solution to the Third Antinomy provides the best account of Kant's resolution to the Third Antinomy and his broader theory of freedom that follows from it. I will outline and defend a unique form of compatibilism, called meta-compatibilism, that I believe is attributable to Kant. I will argue that meta-compatibilism provides the most satisfying answers available to a series of objections held against Kant's solution to the Third Antinomy. Following that, I will defend meta-compatibilism from one significant

¹⁵ *CPR*, A545-6/B573-4.

objection that I anticipate some readers may have. It is from this epistemological platform that Kant intends to license his later account of practical reason and moral psychology, especially in the *GR* and the *CPrR*. My intent is to offer an exegesis of what I call Kant's meta-compatibilist theory of freedom and explain how he believes it permits us to have faith that our choices make a difference in the world. At the end of Chapter II, I will evaluate this theory of freedom and explain some of the major hurdles it faces in offering a satisfying account of moral responsibility.

1.1 The Third Antinomy

1.1.1 Outline of the Third Antinomy

As indicated by its title, the Third Antinomy is one among four antinomies found in the *CPR*. An antinomy is a “rhetorical form of presentation in which opposed arguments are presented side-by-side one another.”¹⁶ In the *CPR*, Kant uses this form in order to delineate the proper limits of human reason, showing that when finite reason illegitimately oversteps the bounds of experience, it generates opposed yet equally justifiable inferences.¹⁷ Kant expresses that the antinomies together are indicative of properties intrinsic to reason itself; namely, they arise from reason's situation as an inherently limited faculty in a world more complex than it can comprehend and so they

¹⁶ Howard Caygill, *A Kant Dictionary* (Malden: Blackwell Publishing, 1995), 75.

¹⁷ *Ibid.*, 76.

are an inevitable result of reason making sense of the world. Therefore, in highlighting the theoretical limitations of reason they signal a limit that is intrinsic to reason itself.¹⁸

Each antinomy contains a thesis and an antithesis that together generate a contradiction that, as Kant says, “falls of [reason] itself and even unavoidably.”¹⁹ Kant maintains that the thesis and antithesis of each antinomy tempt reason to surrender either to a “skeptical hopelessness” or to “assume an attitude of dogmatic stubbornness.”²⁰ Either of these alternatives to the exclusion of the other constitutes the “death of a healthy philosophy.”²¹ In the case of the Third Antinomy, Kant intends to show that are permitted to live with two contradictory statements about human freedom as though both of them are simultaneously true. In holding these particular contradictory statements together, a limit is placed on the scope of theoretical reasoning. In this way, the section of the *CPR* on the antinomies of pure reason fits well into the broader project of providing a critique of pure reason and establishing its limits. The antinomies are designed to fit within Kant’s greater project of denying knowledge in order to make logical room for faith in God, freedom, and immortality.²²

In the Third Antinomy of Pure Reason, Kant is concerned primarily with the second of these three articles of faith—the belief in causality through freedom. He aligns the thesis of the Third Antinomy with this postulate; it states that “causality in accordance with the 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

¹⁸ *CPR*, A408/B435.

¹⁹ *CPR*, A407/B434.

²⁰ *Ibid.*

²¹ *Ibid.*

²² *CPR*, Bxxx.

to explain them.”²³ Kant proceeds to explain that this thesis is supported by the contradiction that results if one does not posit this second form of causality according to freedom but maintains only the first form of causality according to the laws of nature. (Here what Kant has in mind in speaking about the “laws of nature” is the principle that “nothing happens without a cause sufficiently determined a priori”²⁴ or, more simply, that there are no events that occur uncaused.) For without the possibility of an act having been caused through freedom, there would exist no terminus for any given series of events in the world. Without the existence of such a terminus, the given series of events and the prior causes of each event within that series would recede indefinitely backward in time without ever meeting a beginning in the form of a first cause. As a result, such a series could not ever attain completeness and for that reason would contradict the laws of nature on which it is founded. For such laws demand completeness and that completeness is defined in terms of an exhaustive list of efficient causes for each existing event in the world; however, given an infinite regress of causes, such a list of efficient causes (or events, for that matter) cannot be obtained and the series remains incomplete.²⁵ Therefore, to attain such completeness, it is necessary to assume this second form of causality through freedom as an “absolute causal spontaneity beginning from itself.”²⁶ In other words, it becomes necessary to posit transcendental freedom.

²³ *CPR*, A444/B472.

²⁴ *CPR*, A446/B474.

²⁵ Note Allison’s objection to Kant’s holding this infinite regress to be problematic [Henry Allison, *Kant’s Theory of Freedom* (Cambridge: Cambridge University Press, 1990), 13].

²⁶ *CPR*, A446/B474.

On the other hand, the antithesis states that “there is no freedom, but everything in the world happens solely in accordance with laws of nature;”²⁷ therefore it denies precisely what the thesis sought to establish—the existence of a free cause. As noted by Henry Allison, the argument of the antithesis is established not on the basis that its opponent argument (the thesis) possesses an internal contradiction but rather on the basis that its opponent theory contradicts the conditions of possible experience.²⁸ The particular condition of possible experience that the thesis contradicts is the law of causality. For Kant maintains that the law of causality must necessarily hold in order for the unity of experience or consciousness (also called apperception) to obtain because Kant believed that without the law of causality there is no possibility of a “thoroughly connected experience.”²⁹ In fact, Kant argues, if causality in accordance with freedom were to be upheld in the realm of experience, experience itself would break down into lawlessness. In this way, Kant suggests that although freedom from the laws of nature is indeed a liberation from coercion, it is (unfortunately for the case of freedom) also a liberation from guidance—a guidance without which the unity of experience cannot be maintained.³⁰

In spite of the necessity of the antithesis for the unity of experience, the Third Antinomy demonstrates that the antithesis too leads to an undesirable incomprehensibility in the form of an “infinite descent” of causes.³¹ However, it points out that to reject the antithesis on account of this incomprehensibility would come at the cost of rejecting

²⁷ *CPR*, A445/B473.

²⁸ Henry Allison, *Kant's Theory of Freedom*, 20.

²⁹ *CPR*, A447/B475.

³⁰ *Ibid.*

³¹ *CPR*, A449/B477.

many other “fundamental powers” which proves to be an equally incomprehensible thing to do.³² It is safe to assume at this point that the other fundamental powers that Kant has in mind include, perhaps most notably, the power of judgment,³³ as well as cognition in general, neither of which, he contends, would be possible without the unity of experience.

This makes it clear that Kant does not hold the antithesis up as an undeniable theory about the way things in themselves are (I will discuss Kant’s idea of things in themselves shortly). Rather, so long as it restricts itself to the realm of experience alone, the antithesis presents a coherent way of thinking about the world. In fact, Kant believed that it presents a necessary way of thinking about the world insofar as without the sort of causal connectedness found in the antithesis, the subject cannot even experience the world. The necessity of this causal connectedness is outlined in the Second Analogy of the *CPR* when Kant says that “it is only because we subject the sequence of the appearances and thus all alteration to the law of causality that experience itself, i.e., empirical cognition of them, is possible; consequently they themselves, as objects of experience, are possible only in accordance with this law.”³⁴ It is not clear that this degree of causal connectedness (i.e., complete causal determinism) must actually hold in order for the unity of experience to obtain. Nevertheless, this is what Kant believed and for this reason, it is safe to assume that he deemed it necessary to uphold in tandem with a theory of freedom.

³² Ibid.

³³ This is explained by Kant as the subsumption of objects of the understanding under rules in *CPR*, A132/B171 or by Allison as the uptake of data by the mind followed by taking something as “such and such” (Henry Allison, *Kant’s Theory of Freedom*, 37).

³⁴ *CPR*, A189/B234.

1.1.2 Resolution of the Third Antinomy

At this point, the thrust of Kant's transcendental idealism becomes clear as a resolution to the Third Antinomy begins to emerge. As previously discussed, the entire project of the *CPR* constitutes precisely a critique of pure reason and a curtailment of the effects of theoretical reason. In particular, it aims to curtail theoretical reason at the point where it threatens the three articles of faith—God, freedom, and immortality, the relevant postulate here being (causation in accordance with) freedom. As I have said, the thesis of the Third Antinomy roughly represents this relevant postulate (causality through freedom) while the antithesis represents theoretical reason (or one could say, pure reason) as of yet unchecked by critique. The antithesis is incomplete because it cannot account for a first cause, so the thesis is necessary to postulate for theoretical reasons—to provide proper closure to any given chain of causes and avoid an infinite regression of causes. Yet Kant holds that the antithesis is built into the nature of cognition as a necessary precondition for experience itself. Both thesis and antithesis are therefore necessary for theoretical reasons despite the fact that they conflict with one another. Kant calls this “the most beneficial error into which human reason could ever have fallen”.³⁵ For it gives him theoretical grounds to affirm the thesis and in doing so, preserve the postulate of causality according to freedom as he originally set out to do.

At this point, it could be asked how exactly Kant can place such a limit on theoretical or pure reason. Part of this is answered in looking at the general aim and

³⁵ *CPrR*, 5:107.

upshot of the antinomies, as they demonstrate that theoretical reason runs into contradiction when it oversteps its proper boundaries. A more satisfying understanding of this picture and therefore a more satisfying answer to the question can be gained from an understanding of the noumena-phenomena distinction that Kant introduces in the *CPR*, as it is perhaps the distinguishing feature of his transcendental idealism.

The distinction between noumena and phenomena is a result of the limitations of possible experience. Phenomena are objects of possible experience and as such, must conform to the structure and activity of the knowing mind, which actively shapes experience.³⁶ In order for the mind to conceive of anything at all, it must posit both space and time. Space and time are necessary preconditions for possible experience—no objects can be experienced or conceived of without previously positing both space and time. Kant calls space and time the two pure forms of sensible intuition and they may be known a priori.³⁷ The twelve categories of the understanding follow from these two a priori conditions and “apply to objects of intuition in general *a priori*... for the understanding is completely exhausted and its capacity entirely measured by these functions.”³⁸

These categories are objective insofar as they provide rules for the conduct of experience that are common to all sensible knowledge.³⁹ Together with the a priori intuitions of space and time, these categories are constitutive of all possible sensible understanding and knowledge; however, they do not penetrate beyond the realm of

³⁶ Lewis White Beck, *A Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960), 22.

³⁷ *CPR*, A22/B36.

³⁸ *CPR*, A80/B105.

³⁹ Lewis White Beck, *A Commentary on Kant's Critique of Practical Reason*, 22.

possible experience and into things as they are independent from experience. Kant summarizes this feature of knowledge in his general remarks on the transcendental aesthetic.

[A]ll our intuition is nothing but the representation of appearance; that the things that we intuit are not in themselves what we intuit them to be, nor are their relations so constituted in themselves as they appear to us; and that if we remove our own subject or even only the subjective constitution of the senses in general, then all the constitution, all relations of objects in space and time, indeed space and time themselves would disappear, and as appearances they cannot exist in themselves, but only in us.⁴⁰

Therefore, according to Kant the only knowledge that is possible is knowledge of the world (and things in it) as it presents itself in and through experience. All knowledge gained through experience is knowledge of the phenomenal realm—the realm that we know through experience, while the realm of things as they are in themselves, independent from our experience of them, is called the noumenal realm. Whereas we can know things about the phenomenal realm, we cannot know things about the noumenal realm.⁴¹ This, in brief, is the defining feature of Kant's transcendental idealism and that being the case, transcendental idealism places a crucial limit on the scope of theoretical reason.

With an understanding of Kant's transcendental idealism in place, we can attain a clearer understanding of how exactly he attempts to limit the scope of theoretical reason and preserve the possibility of causality according to freedom in the Third Antinomy.

⁴⁰ *CPR*, A42/B59.

⁴¹ The knowledge of which I am speaking at present is theoretical knowledge. There is more to be said about the possibility of knowing things in a restricted sense about the noumenal realm, via practical (moral) knowledge, which will be explained in Chapters II and III. In the meantime, for the purposes of this chapter my use of the term 'knowledge' is restricted to mean theoretical knowledge only.

Essentially, he allows theoretical reason, in the form of the complete causal closure of the antithesis, full validity within the realm of experience as it applies to objects of experience. However, every object in the realm of experience has a dual character.

According to transcendental idealism, every object has both a phenomenal character and a noumenal character—a character that can be known through experience and a character that cannot be known through experience. The fact that the complete causal closure that the antithesis (and theoretical reason) seeks is operative within the realm where only the phenomenal character of objects is accessible, means that it can only speak for things as they show themselves in that realm. Importantly, it cannot speak for things as they are outside of or apart from experience. Theoretical reason cannot apply its method of analysis to that other character of things that lies beyond experience—the noumenal realm—because its analysis can never be complete. As the antithesis of the Third Antinomy shows, theoretical reason in this application, leads to an infinite regress of causes, which makes the causal closure that it seeks impossible to achieve. That is, it makes causal closure impossible without the assumption of a first cause, which it itself rules out in principle. As the antinomies are indicative of limits intrinsic to reason, so in this case, the limit of this application of theoretical reason evident in the antithesis is also intrinsic to reason.

Reason itself, in its theoretical aim and application, cannot be complete. It turns out that the antithesis of this as well as the other three antinomies would render the edifice of knowledge impossible.⁴² For it would not be possible to achieve a complete system of scientific knowledge without the conceptual closure afforded by the

⁴² Lewis White Beck, *A Commentary on Kant's Critique of Practical Reason*, 185.

assumption of a free cause. Therefore, it is in the interest of theoretical reason itself to hold onto the thesis of a causality operating in accordance with freedom—in relation to the noumenal. The simultaneous application of the thesis and the antithesis of the Third Antinomy to the object as noumena and phenomena respectively (including the object of the human agent herself), represents Kant's resolution of the Third Antinomy. With this resolution, Kant intends to maintain causal determinism within the realm of experience while insulating the possibility of one of his three practical postulates—causation according to freedom—from the destructive effects of such a form of determinism.

1.1.3 Problems with the Third Antinomy

This solution sounds almost too good to be true—for can it really be the case that human freedom and causal determinism co-exist simultaneously? This seems to be what Kant is saying. However, because freedom may be located in the noumenal realm, which it is beyond the capacity of human reason to examine, he maintains that freedom is neither proven nor provable. This gives rise to a number of very challenging objections to Kant's theory of freedom. The most challenging objection comes in the form of, perhaps, the simplest question—how exactly is it possible that something in the noumenal world can exert causal influence on a thing or series of events in the phenomenal world, whose causal trajectory is already determined? There are several other important objections that more or less follow from this one, for it seems that on a very basic, intuitive level there is something difficult to grasp about Kant's theory of freedom and its claim to hold two

seemingly deeply contradictory theories together at once. In what follows, I will briefly outline what I understand to be the major objections to Kant's theory of freedom.

As just stated, the first and most obvious objection to Kant's theory of freedom is associated with the deep tension that exists between the thesis and antithesis of the Third Antinomy—between causality according to freedom and causality according to the laws of nature (the latter of which implies complete causal determinism). In other words, it is not clear that something in the noumenal world can exert causal influence on something in the phenomenal world when the causal trajectory of all things in the phenomenal world has been determined by prior causes. Terence Irwin has expressed the intuition behind this objection most succinctly in his observation that “if an event is determined, it is true of it under all true descriptions that it is determined.”⁴³ This particular objection suggests that Kant's attempt to escape the effects of the sort of determinism operative in the phenomenal realm by placing freedom in the putative noumenal realm, ultimately fails because every putatively free noumenal action is predictable from preceding empirical conditions.⁴⁴ This is what I will call the hard objection from noumenal determination of phenomena. I believe that this term is most suitable because the objection comes from a general difficulty with causation in the direction of noumena to phenomena. I call it the hard objection because I believe that it presents the most difficult objection to Kant's theory of freedom as it directly challenges the *prima facie* possibility of spontaneous action.

⁴³ Terence Irwin, “Morality and Personality” in *Self and Nature in Kant's Philosophy* ed. Allan Wood (Ithaca: Cornell University Press, 1984), 38.

⁴⁴ Allen Wood, “Kant's Compatibilism,” 85.

A second objection that is closely associated with the hard objection and also finds difficulty with causation in the direction of noumena to phenomena is the objection from overdetermination. Overdetermination occurs when one effect or event is understood to have more than one cause. When Kant appeals to noumenal causation in the resolution of the Third Antinomy he is, in effect, appealing to overdetermination in order to license the idea of human freedom. Given that every effect in the phenomenal world is sufficiently determined by a series of prior causes, adding a second-order noumenal cause would provide two sufficient causes for one effect. This constitutes an overdetermination of that effect. But this is problematic because more than one line of causation is said to determine one outcome and as Irwin points out, if an event is determined, it is true that it is determined under all true descriptions,⁴⁵ so how can there be a description under which it is free? One of these causes must then be unnecessary and the obvious candidate for dismissal is the cause whose existence cannot be known—the noumenal cause. The logic that reacts negatively against the idea of overdetermination is the same logic that underlies the hard objection. Derk Pereboom points out that this objection only applies in the case of a two-world view of noumena and phenomena, where each of the two respective realms remains ontologically distinct from the other.⁴⁶ I will discuss the nature of the two-world as well as the two-aspect views of noumena and phenomena further on. In the meantime, it is helpful to note that according to at least one interpretive stance, the objection from overdetermination makes it difficult to accept Kant's resolution of the Third Antinomy and his consequent theory of freedom.

⁴⁵ Terence Irwin, "Morality and Personality," 38.

⁴⁶ Derk Pereboom, "Kant on Transcendental Freedom" in *Philosophy and Phenomenological Research* LXXIII:3 (November 2006): 551.

The next two objections stem from a difficulty inherent in the concept of a timeless agency and character. In his resolution to the Third Antinomy, Kant introduces the concept of an intelligible character that may be attributed to the causally determined empirical character that is encountered in experience. Together with the empirical character, the intelligible character may also be considered a cause of appearances, though it stands outside the conditions of sensibility.⁴⁷ Most importantly, the intelligible character stands outside the conditions of time, for in such a character “no action could arise or perish.”⁴⁸ Therefore, such a character would possess a form of agency not subject to the law of time-determination and the causal determinism that comes with it. Instead, such a character’s agency would be understood as timeless. But how can timeless agency be properly understood to have time-bound, empirical/sensible effects attributed to it? That is, how can an agent that is timeless be understood to act in time? The difficulties raised by these questions form the grounds for the basic objection from timeless agency.

I call the second objection associated with timeless agency the objection from timeless character and it is associated with the charge that timeless agency delivers counterintuitive results for the concept of character. To illustrate this problem, it can be pointed out that we understand our empirical character from a chronologically linear point of view—we understand ourselves as thinking and acting out of particular character traits that we have formed *over time*. In other words, from the point of view of the empirical world, we appear to have only one causal history that has shaped our character and disposition to act a certain way at any given point in time. However, considering causality in accordance with timeless character appears to abolish this. For if we were

⁴⁷ *CPR*, A539/B567.

⁴⁸ *Ibid.*

suspended from the conditions of time we would be suspended from the conditions of our character as developed in time. Allan Wood points out that this implies that for any given moment in time, our timeless noumenal character would be free to choose its causal history as though from among a certain subset of possible worlds and determine that the actual world will be drawn from this subset of possibilities.⁴⁹ For each choice there is “an almost endless variety of ways in which I might have chosen differently, and endless variety of possible empirical selves and personal moral histories I might have actualized.”⁵⁰ This seems radically counter-intuitive and reinforces the first basic objection from timeless agency.

Furthermore, consider our commonsense intuitions about character formation and the spatio-temporal world in which it takes place. Timeless agency would appear to remove the purpose from punishment (or otherwise forward-looking corrective action/policy) as well as striving to be a better person than one already is.⁵¹ For according to this scheme, punishing the empirical character would not evidently carry any force in guiding the decisions or behavior of the noumenal agent.⁵² For the noumenal agent to whom the decision belongs, would always have available to her a possible causal history in which the punishment did not occur. As a result of her suspension in time (or timelessness), that causal history would act on her with no more or less force than the causal history evident in the empirical world, in which she was punished.

In summary, the major objections to Kant’s resolution of the Third Antinomy and consequent theory of freedom include: 1) the hard objection from noumenal

⁴⁹ The origin of this challenge is found in Allan Wood, “Kant’s Compatibilism,” 91.

⁵⁰ Ibid.

⁵¹ Ibid.

⁵² Lewis White Beck, *A Commentary on Kant’s Critique of Practical Reason*, 191.

determination of phenomena, 2) the objection from overdetermination, 3) the basic objection from timeless agency, and 4) the objection from timeless character. Together, these four objections appear to raise some serious difficulties with Kant's theory of freedom. They all follow in one form or other from his division of things and agents into noumenal and phenomenal character. Objections 1) and 2) reveal a difficulty inherent in the idea of noumenal causation itself while objections 3) and 4) reveal a difficulty with how to comprehend the idea of noumenal causation from a phenomenal, empirical, or otherwise sensible point of view. Is transcendental idealism then a hindrance or help in understanding the nature of human freedom? In what follows I suggest that the answer to that question depends on what one is trying to achieve. I will attempt to cast Kant's project here in a favorable light and argue that *if* we accept his framework of transcendental idealism, we see that he offers a complex account of the nature of human freedom in a way that helps us to arrive at satisfying answers to the above four objections. Before I do that, however, I will say more about exactly what Kant's theory of freedom is, or how we should understand it.

1.2 Meta-Compatibilism: An Interpretive Solution

1.2.1 An Outline of Meta-Compatibilism

In this section, I will outline and defend a compatibilist interpretation of Kant's theory of freedom that I will call meta-compatibilism. Such an interpretation, I suggest, not only provides the most satisfying answers available to the objections raised above

against Kant's theory of freedom, it additionally provides the basis for an adequate, self-standing compatibilist account of freedom independently of Kant and his project in the *CPR*.

As indicated by its name, meta-compatibilism preserves the intuition behind any compatibilist account of freedom; namely that practical freedom and responsibility are consistent with causal determinism.⁵³ In this way, its goal at bottom is to preserve the respective integrity of these two domains of practical freedom and causal determinism. It does this by acknowledging a genuine insolubility of the problem of free will, given human cognitive capacities, and subsequently appeals to its conceptual possibility in order to permit ascriptions of freedom in the face of this insolubility. In other words, it says that although we do not know whether freedom is actual, we do know that it is possible; therefore, in light of this possibility, we are permitted to believe in freedom. Because it does not solve the problem of free will at the level of actuality (as it claims that this problem is insoluble) but operates at the level of belief-formation so as to regulate which beliefs are and are not rationally consistent, I have given it the term meta-compatibilism. As a theory of freedom, meta-compatibilism can be said to have three distinct features: 1) it is compatibilist in nature and therefore aims to hold together both freedom and causal determinism simultaneously, 2) it adheres to a strict limit to human cognitive capacities, and 3) it holds that certain beliefs are rationally permitted in light of 1) and 2) above. I will discuss each of these features in turn, showing their source and role in Kant's theory of freedom.

⁵³ Thomas Kapitan, "Free Will Problem" in *Cambridge Dictionary of Philosophy* 2nd edition, ed. Robert Audi (Cambridge: Cambridge University Press, 1999), 327.

1.2.1.1 Practical Freedom and Causal Determinism Together

Regarding the first feature of meta-compatibilism, that it is compatibilist in nature and aims to hold together both practical freedom and causal determinism, it is helpful to note Allan Wood's comment on the application of the term 'compatibilism' to Kant in this regard. He says that the basic question of freedom (in terms of the compatibilism-incompatibilism debate) is whether, regarding the same effect as determined by nature, freedom can be present or not; in saying that it can, Kant allies himself with the aims of compatibilism.⁵⁴ In the resolution to the Third Antinomy, Kant himself says that freedom and determinism, "each in [their] full significance," may be found in the same actions depending on whether one regards them according to their intelligible or sensible cause.⁵⁵

Kant argues that one action may have both an intelligible cause and sensible cause on the basis of the distinction he makes between empirical and intelligible character, which itself depends on transcendental idealism's distinction between noumena and phenomena. Because the human being herself is an appearance, she possesses an empirical character that is subject to the conditions of experience, those conditions being space and time.⁵⁶ Taken alone, these conditions imply the application of total causal determinism, as evident in the antithesis of the Third Antinomy, to human action. In light of this character, all her actions "would have to admit of explanation in accordance with natural laws, and all the requisites for a perfect and necessary determination of them

⁵⁴ Allan Wood, "Kant's Compatibilism," 74.

⁵⁵ *CPR*, A541/B569.

⁵⁶ *CPR*, A552/B580.

would have to be encountered in a possible experience.”⁵⁷ Thus it is evident that Kant maintains one wing of the compatibilist position—complete causal determinism.

Intelligible character, on the other hand, is not subject to the conditions of experience (space and time) because it is grounded in the noumenal reality of things as they are in themselves. Therefore, the particular human being as she is in herself, acts independently of these conditions of experience and most importantly, independently of the conditions of time.⁵⁸ Intelligible character exists as a result of the need for all appearances to be grounded in a transcendental object.⁵⁹ As previously discussed, the antinomies of the *CPR* raise the issue of the scope of reason’s effectiveness; in particular, the Third Antinomy demonstrates that when reason equates appearances with reality (things as they are in themselves) it cannot satisfy even its own demands for completeness. As a result, it is forced to postulate a first cause in order to complete the otherwise infinite causal chain encountered in experience. This shows that to satisfy reason’s demands, reason must go beyond itself in postulating something unavailable in the world of appearances—it must postulate the noumenal in the form of a first cause.

Although the thesis and antithesis of the Third Antinomy deal with the cosmological concept of a first cause and do not discuss individual agent-causation in particular, together they open up a “logical space” in which individual agent-causation may be thought.⁶⁰ Now that the complete authority of the realm of nature and the causal determinism that it implies has been effectively challenged, it is possible to postulate/attribute a second order of causality—intelligible causality—to the effects of

⁵⁷ *CPR*, A540/B568.

⁵⁸ *CPR*, A539-40/B567-8.

⁵⁹ *CPR*, A538/B566.

⁶⁰ Henry Allison, *Kant’s Theory of Freedom*, 25.

every example of agent causality.⁶¹ (Kant has a creative way of looking at the two orders of causality to which I will return in section 1.2.2 when I discuss meta-compatibilism and the problems of the Third Antinomy.) This second order of causality forms the basis for attributions of freedom to the human being who is otherwise known or encountered strictly in her empirical character; anchored in the intelligible character, this order of causality is simultaneously attributable to the same object as the causal determinism of the empirical character. By regarding the same effect as simultaneously free and determined by nature, Kant preserves the intuition behind compatibilism—that causal determinism does not preclude freedom—and allies himself with its aim to preserve each of these two doctrines in their respective domains of application. In this way, Kant maintains the second wing of compatibilism—the belief in practical freedom and responsibility.

1.2.1.2 Limited Cognitive Capacities

The second distinguishing feature of meta-compatibilism is that it adheres to a strict limit to human cognitive capacities.⁶² This limit is evident as a result of the claim that reason cannot of itself attain the sort of completion that it itself requires. In other words, reason requires the concept of an explanatory whole in order to make sense of the world and to prop up the edifices of knowledge and science.⁶³ Yet reason itself cannot provide all of the elements necessary to make this explanatory whole possible. Therefore

⁶¹ *CPR*, A539/B567.

⁶² Though I define the idea of limited cognitive capacities differently than Derk Pereboom, this particular expression of his inspired my language: “[g]iven the sort of cognitive equipment we have, we can only determine by means of experience whether the conditions of real possibility are satisfied. This fact limits us to discerning real possibilities only for objects of experience, and not for noumena” (Derk Pereboom, “Kant on Transcendental Freedom,” 545).

⁶³ Henry Allison, *Kant’s Theory of Freedom*, 24.

it must postulate some of these elements as basic axioms whose existence is necessary yet whose origins remain unexplainable.

Henry Allison describes this tension inherent in reason in terms of the conflict between the completeness requirement and the universalizability requirement.⁶⁴ He means to point out that, regarding the nature of causality, Kant believes that reason requires that any account of a given event's causal history must be both complete *and* operate according to laws that are applicable to all other events and things. This is because on one hand, the infinite regress that results from an exclusive adherence to the antithesis of the Third Antinomy violates the completeness that reason seeks and renders the previously mentioned edifice of knowledge impossible to attain. While on the other hand, an exclusive adherence to the first-mover like argument of the thesis precludes the possibility of applying causal laws in a universally consistent way. Kant's point is that reason cannot solve this apparent contradiction on its own resources, yet it nonetheless requires the concept of an explanatory whole by which to make sense of the world and the acquisition of knowledge.⁶⁵ As a result of this need in combination with its deficiency, reason postulates a form of causality whose nature, while necessary, is beyond its comprehension.

In this there is a double emphasis, showing that reason cannot completely comprehend the world that it encounters and that while it postulates another form of (noumenal) causality to account for this, it emphatically cannot completely comprehend that causality. This latter emphasis on the unknowability of noumenal causality represents a necessary feature within Kant's theory of freedom. For the causal determinism of the

⁶⁴ *Ibid.*, 18.

⁶⁵ *Ibid.*, 24.

laws of nature is, as Kant contends, a law of the understanding insofar as it is a precondition for the possibility of experience.⁶⁶ If this law of the understanding were not always satisfied, appearances could not be unified. Instead they would lie outside of the conditions of possible experience and be capable of being no more than mere phantasms or “figments of the brain.”⁶⁷ Therefore, to preserve the possibility of freedom and the noumenal realm in which freedom possibly inheres, Kant is forced to limit the scope of this understanding. This leads him to make a strict separation between intelligible and empirical character in the resolution to the Third Antinomy. Whereas reason may know that the actions of the empirical character are “linked with inseparable dependence on the natural chain of causes,” the actions of the intelligible character must be “passed over as entirely unknown except insofar as it is indicated through the empirical character as only its sensible sign.”⁶⁸ This confirms Kant’s overall project in the *CPR* to “deny knowledge to make room for faith,” that is, faith in God, freedom, and immortality as stated in the preface to the second edition of the *CPR*.⁶⁹ Specifically, it is meant to deny knowledge in order to permit us to have faith that our choices make a difference in the world—that freedom is compatible with causal determinism. Furthermore, it demonstrates that Kant adheres to a strict limit to human cognitive capacities and in doing so, upholds the second feature of meta-compatibilism outlined at the beginning of this section.

1.2.1.3 Rationally Permitted Faith

The third feature of meta-compatibilism is that the particular faith that it licenses is rationally permitted. The purpose of this section will be to show that although human

⁶⁶ *CPR*, A542-3/B570-1.

⁶⁷ *CPR*, A543/B571.

⁶⁸ *CPR*, A546/B574.

⁶⁹ *CPR*, Bxxx.

cognitive capacities are limited and consequently, the noumenal realm remains cognitively inscrutable, the belief in freedom is not merely wishful thinking but is formed on rational grounds. First of all, the qualification must be made that meta-compatibilism deals precisely with the formation of faith or belief. It is not a theory about the actual metaphysical status of freedom. Rather, it acknowledges a genuine insolubility at the heart of the metaphysical problem of free will and subsequently proceeds to permit faith in freedom on the basis of its possibility. (The grounds for this possibility were discussed in connection with the distinction between empirical and intelligible character in section 1.2.1.2 above.) In light of the unattainability of knowledge of the noumenal realm, the question becomes whether one can regard the same event as both determined and free without contradiction.⁷⁰ Kant's answer to this question is yes. Because the human agent can be seen through the lens of both her empirical and intelligible character, without one reducing to the other, her actions can be simultaneously regarded as both determined and free.⁷¹ With the prior conditions of its metaphysical possibility and the lack of contradiction in its concept having been met, reason is then justified in believing that the human agent is free.

There is another reason that impels us to regard human action as free and in that way demonstrates that belief in freedom is consistent with rationality. Henry Allison points out that it is a feature of reason itself that it seeks completeness.⁷² When considering the case of causality, the completeness that reason is seeking is the "absolute

⁷⁰ *CPR*, A543/B571.

⁷¹ *CPR*, A541/B569.

⁷² Henry Allison, *Kant's Theory of Freedom*, 24.

totality of conditions in causal relations.”⁷³ As it finds no such totality of conditions in the law of nature and finds an infinite regress of causes untenable, “reason creates the idea of a spontaneity.”^{74,75} This spontaneity entails the ability of a cause “to act from itself, without needing to be preceded by any other cause that in turn determines it to action according to the law of causal connection.”⁷⁶ The actual existence of this spontaneous cause cannot be verified because it is a strictly intelligible cause rooted in the noumenal realm, the certainty of which is beyond the reach of reason. Nonetheless, faith in it is rationally permissible because, according to Kant, rationality seeks a totality of conditions in causal relations that can only be attained if an intelligible cause is postulated. Again, this is similar to what Allison points out as reason’s need to form an “explanatory whole” from which it can make sense of the world.⁷⁷ This explanatory whole is only made possible by holding together the seemingly contradictory completeness and universalizability requirements of the thesis and antithesis, respectively, of the Third Antinomy.⁷⁸ Thus the intelligible character provides the basis for the completeness requirement and the empirical character for the universalizability requirement. The point here is that although it is not possible for reason to prove that intelligible character exists, it is permissible for reason to have faith that it does. The sort of faith that meta-compatibilism issues-in is therefore one that is consistent with rationality.

⁷³ CPR, A533/B561.

⁷⁴ Kant is not the first to use this argument. Kant’s is an adaptation of an argument first developed by Aristotle in his analysis of motion and consequent argument for the necessity of a first mover (or first cause) in Book 8, Chapters 4 to 6 of his *Physics*.

⁷⁵ Ibid.

⁷⁶ Ibid.

⁷⁷ Henry Allison, *Kant’s Theory of Freedom*, 25.

⁷⁸ Ibid.

The last significant rational feature of Kant's belief in freedom foreshadows his work to come in the *GR* and the *CPrR*. In the resolution to the Third Antinomy, he briefly introduces the "ought" as a species of necessity unto its own, taking place nowhere in nature except in the human mind.⁷⁹ He says that "[i]n nature, the understanding can cognize only what exists, or has been, or will be. It is impossible that something in it ought to be other than what, in all these time relations, it in fact is."⁸⁰ What he means by this is that the human being is the only thing in nature that has the (frequently painful) sense that the world should (ought to) be a certain way—more kind, more just, or more gracious than it presently is, to use just a few examples. Often this entails a sense of moral obligation that comes along with it—that one should act in a certain way so as to make the world into the sort of place that it ought to be. However, Kant does not take much time to develop this thought here. Instead, he saves it for much greater attention in the *GR* and the *CPrR*.

Kant continues to say that this ought expresses an action whose only possible ground is a concept, whereas a merely natural action will always be grounded in an appearance.⁸¹ In other words, no matter how many natural grounds may compel a person to will something, none of these grounds can produce the sense that a person ought to do anything. Insofar as we believe we ought to do certain things and that the world ought to be any way at all, Kant maintains that we acknowledge the necessity of an intelligible, non-natural causal ground. Therefore, insofar as it is rational to be moral, Kant provides us with an additional reason that belief in freedom is rational. Because it forms such an

⁷⁹ *CPR*, A547/B575.

⁸⁰ *Ibid.*

⁸¹ *CPR*, A547-8/B575-6.

important part of his idea of moral responsibility in the *GR* and *CPrR*, it will be discussed at length in Chapter II. That being the case, it will not be discussed at length here, but it is worth mentioning because it provides an important rational ground on which the appeal to faith in freedom may be made. In combination with the genuine insolubility of the metaphysical problem of free will and reason's need to attain a completeness of causes and render the edifice of knowledge possible to complete, it shows that faith in freedom is rationally permissible. The rational permission granted to faith in freedom forms the third distinctive feature of Kant's theory of freedom—what I have called meta-compatibilism.

1.2.2 Meta-Compatibilism and the Problems of the Third Antinomy

The crucial feature of Kant's theory of freedom (meta-compatibilism, as I have called it) is the fact that it operates on the level of faith and belief formation. This feature is crucial because it prevents the previously mentioned four objections to his theory of freedom from proving fatal to his theory of freedom. In what follows, I will outline how meta-compatibilism, as an interpretive approach to Kant's theory of freedom, responds to these objections and is able to uphold a satisfying compatibilist account of freedom.

At the end of the resolution of the Third Antinomy, Kant makes it clear that the task he set out for himself was not to prove the reality of freedom, not even its possibility.⁸² Rather, the task he set out for himself was to determine whether the idea of freedom is without contradiction, which he claims to have done by showing that “since in

⁸² *CPR*, A558/B586.

freedom a relation is possible to [noumenal] conditions of a kind entirely different from those in natural necessity, the law of the latter does not affect the former.”⁸³ This is a more modest task than proving the reality of freedom because this task only undertakes to determine what beliefs are rationally consistent in light of the insolubility of the problem of free will. This being the case, it does not claim to say anything at all about the nature of the noumenal realm, for to do so would be a transgression of reason’s limits. As Kant says, it surpasses reason’s authority to even inquire into the nature of the noumenal realm.⁸⁴ For this reason Kant believes that his theory of freedom escapes ruin from the hard objection from noumenal determination of phenomena. For Kant is not making any claims about the reality of the noumenal world as such but only about how we can think of ourselves as agents. Unlike other philosophers, Kant has no problem appealing to faith when it is rationally permitted; here he appeals to faith by saying that we are rationally permitted to believe that our actions make a difference in the world. Noumenal character is not then known as an actual cause of events in the empirical/sensible world; rather it provides a way in which events in the empirical/sensible world may be thought.

Allen Wood describes Kant’s strategy here as an exploitation of the burden of proof that lies on those who believe that we are unfree.⁸⁵ His noumenal theory of causality needs only to show that *if* true, it would save the practical idea of freedom despite determinism being true; he does not actually need to prove that it is true and, as we have seen, he claims that he cannot. This is mostly a matter of the irreducibility of the (noumenal) realm of thought to the (empirical) realm of appearances and he contends that

⁸³ *CPR*, A557/B585.

⁸⁴ *Ibid.*

⁸⁵ Allen Wood, “Kant’s Compatibilism,” 84.

neither of those operates on the level of metaphysical actuality. Therefore, Kant is making only the modest claim that we are permitted to think about our actions as free, which says nothing about the real causality of those actions. Again, Kant's position here serves the general aim of compatibilism—to preserve the simultaneous application of action-guiding freedom and causal determinism to the concept of human agency.

In a similar way, once it can be established that Kant's theory of noumenal causation serves to permit thought to regard actions in the empirical world as practically free and not actually (noumenally) free, the objection from overdetermination effectively dissolves. Overdetermination is only an issue where there are two ontologically distinct (therefore sufficient) causes said to be bringing about one effect. Yet Kant did not intend for his theory of noumenal causation to actually give an account of causality as it really is (or as it is for things-in-themselves). Kant intended only to use this theory as a means of showing that due to our inability to cognize things as they are in themselves, we are permitted by reason to think of causally determined actions as free. His idea is that we are permitted to take it on faith that there is room for overdetermination to work—that our choices do make a difference in the world. Once we realize that Kant achieves this and not a full-blown theory of ontological noumenal causality, the objection from overdetermination no longer retains its force.

Regarding the basic objection from timeless agency, meta-compatibilism's ascription of a more ontologically modest claim to Kant's theory of freedom again helps to dissolve objections held against it. For timelessness is then not a claim about the way that human agency actually operates. Rather, it forms the basis of a strategic move meant to inject doubt into belief in the absolute scope of causal determinism. Nothing can be

said or known about the nature of human agency, as it is in itself; Kant holds that to make such a claim would not be possible for human reason. In this way, timeless agency merely permits the idea of freedom and shows it to be irreducible to empirical laws and causal determinism in particular. As such, the idea of freedom is a regulative principle of reason that applies to things as experienced and says nothing of how things actually are in themselves, that is, how they are beyond the confines of experience.

Finally, the objection from timeless character and its counter-intuitive account of character formation and punishment can be similarly dissolved. For it too only applies in the case that Kant's theory of timeless agency is making claims about how things actually are or are in themselves. Problematically, the concept of timeless character violates Kant's own idea that we cannot know anything about the noumenal realm. However, understood only as a philosophical strategy employed to gain practical results by telling thought what is and is not permitted, the objection cannot hold. The practical results that it aims for are those that accompany the retaining of practical freedom. If one agrees with Kant that this philosophical move is successful, then one can retain commonsensical notions of character development and punishment. For the concept of timeless agency and character is not meant to apply directly to how we conceive of human character. As already said, it is a strictly strategic move meant to inject doubt into belief in the universal scope of causal determinism and demonstrate the irreducibility of practical freedom to laws of causal determinism. Kant's idea of timeless agency clears a 'logical space,' as Allison said earlier, in which the human being may take her actions as free in spite of her actions being simultaneously determined by prior causes. Though it may itself be counter-intuitive or, at least, theoretically awkward given the complex epistemic

backdrop of transcendental idealism, the consequences it engenders are nothing but the rules of common sense embodied in a system of freedom and morality. For this reason, although other problems remain with the idea of theorizing about the noumenal, the objection from the counter-intuitive nature of character cannot be held against a meta-compatibilist reading of Kant on freedom.

1.3 A Significant Objection to Meta-Compatibilism

A significant possible objection to my rendering Kant's theory of freedom as meta-compatibilism arises from a basic disagreement with a compatibilist reading of Kant, for it could be argued that Kant is instead an incompatibilist about free will and causal determinism. I will argue that this objection is not successful and as a result, meta-compatibilism's solutions to the above problems remain intact.

This reading of Kant as an incompatibilist is based on the observation that for him, true freedom (transcendental freedom) requires an "independence of determination by all antecedent causes in the phenomenal world."⁸⁶ This is, in turn, rooted in Kant's distinction between empirical and intelligible character and the timelessness that comes with the definition of intelligible character. It maintains that in making this distinction, Kant is implying that in order to obtain, true freedom must lie outside the conditions of time and causal determinism and to define freedom in this way is, at bottom, to define freedom in an incompatibilist way. Hence true freedom can only be transcendental freedom and transcendental freedom is the only ground upon which attributions of

⁸⁶ Henry Allison, *Kant's Theory of Freedom*, 1.

agency can be founded. In this manner, it is understood that Kant's distinction between empirical (time-bound) character and intelligible (timeless) character upholds an incompatibilist conception of freedom.

Kant does say that it is the “transcendental idea of freedom on which the practical concept of freedom is grounded”⁸⁷ and that practical freedom is “the independence of the power of choice from necessitation by impulses of sensibility.”⁸⁸ But here it is important to reiterate that Kant's meta-compatibilist view on freedom is distinct from that of standard compatibilists. As indicated in the quote above, Kant is an incompatibilist about freedom on one level, so to speak, in the phenomenal realm—the causal determinism of the phenomenal realm (that is a necessary condition for experience) precludes the possibility of free choice in the phenomenal realm. However, he argues that it is possible for freedom to obtain in the noumenal realm even though it cannot obtain in the phenomenal realm and he argues that we are permitted to take it on faith that causal determinism is somehow wrong and choices that are not themselves caused by anything other than the agent do make a causal difference. So causal determinism is true but he holds it is possible for us to have faith, without contradicting reason, that we do have freedom. The net effect of this is that, for Kant, freedom is compatible with determinism but only by invoking a separate (noumenal) order of possible causation.

Kant further clarifies his position on freedom toward the end of the resolution of the Third Antinomy, where he says that his aim is not to establish the reality of freedom nor even its possibility; rather his aim is to show that “nature at least does not conflict

⁸⁷ *CPR*, A533/B561.

⁸⁸ *CPR*, A534/B562.

with causality through freedom.”⁸⁹ Once it is obvious that Kant does not intend to establish the reality of transcendental freedom, it is evident that he only intends to permit us to have faith that our choices make a difference in the world and that freedom is in the final analysis compatible with causal determinism. Of course, this is reinforced by his strong agnosticism towards the reality of the noumenal world and transcendental causality. Regarding this, he makes the following important statement:

[The] intelligible ground does not touch the empirical questions at all, but may have to do merely with thinking in the pure understanding; and, although the effects of this thinking and acting of the pure understanding are encountered among appearances, these must nonetheless be able to be explained perfectly from their causes in appearance, in accord with natural laws, by following its merely empirical character as the supreme ground of explanation; and the intelligible character, which is the transcendental cause of the former, is passed over as entirely unknown.⁹⁰

In light of these limits imposed on his aims by Kant himself, it is evident that he is not making the claims that timeless intelligible character actually exists and exists in such a way that it fulfills the requirements necessary for an incompatibilist conception of freedom. Instead, he is making only the ontologically modest claim that due to the limitations inherent in reason (or the limitations inherent in our cognitive faculties) transcendental freedom cannot be ruled out and it is not inconsistent for us to think of ourselves as acting freely (as noumena) while simultaneously causally determined (as phenomena). Because it operates on the basis of certain beliefs about the limitations of human cognitive capacities and issues in faith and not direct knowledge (albeit rationally permitted faith) about the nature of freedom, Kant’s theory of freedom is best described under the rubric of meta-compatibilism.

⁸⁹ *CPR*, A558/B566.

⁹⁰ *CPR*, A545-6/B573-4.

In concluding this chapter, it is important to reiterate that in his solution to the Third Antinomy Kant did not understand himself to be establishing the actual and irrefutable nature of human freedom. Rather, in line with what he said at the outset of the *CPR*, he understood himself to be denying knowledge in order to make room for faith.⁹¹ In the case of the Third Antinomy and its resolution, he intended to deny the absolute scope of causal determinism in order to permit a rational faith in human freedom—faith in the idea that our decisions actually do make a difference in the world. In this way, freedom is licensed by faith, and faith in freedom is made possible by the incompleteness of reason due to the limitations inherent in our cognitive faculties.

⁹¹ *CPR*, Bxxx.

CHAPTER II

[W]e now see that when we think of ourselves as free we transfer ourselves into the world of understanding as members of it and cognize autonomy of the will along with its consequence, morality.⁹²

In the resolution to the Third Antinomy in the *CPR*, Kant briefly introduced the “ought” as its own species of necessity, which connects to a ground found nowhere else in nature but the human mind.⁹³ As discussed in Chapter I, the brief mention of the “ought” at this point in the *CPR* foreshadows Kant’s later project in both the *GR* and the *CPrR*, for the *GR* and the *CPrR* lay out the structure of moral reasoning and the contents of morality. In this chapter, I will argue that Kant understood morality, unlike freedom, to be capable of proof on rational grounds. In other words, as the contents of practical reason and a system of right conduct, morality can be established conclusively although our ability to act freely in accordance with it remains an item of faith. If freedom is posited, it then follows that human beings are capable of ordering their conduct according to the edicts of morality. For reasons that I will argue in Chapter III below, it also follows from this picture that moral reasoning provides a source of motivation and rational evaluation that can be both internal to every rational agent and capable of delivering an objective form of justification for conduct. As I explained in the Introduction, I argue that these two conditions—internal motivation and objective justification—in addition to at least a compatibilist account of freedom, are what Kant holds to be two necessary conditions for moral responsibility and therefore must be present in Kant’s account of moral reasoning

⁹² Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. and ed. Mary Gregor, (Cambridge: Cambridge University Press, 1997), 4:453.

⁹³ *CPR*, A547/B575.

in order to attain a coherent theory of morally responsible agency. Before discussing these two conditions in detail, I will first outline the transition he makes from the *CPR* to the *GR*, where he first introduces the concept of the moral law in the form of the categorical imperative. Following that, I will explain how he derives the moral law and how his derivation of it connects to his earlier theory of meta-compatibilism in the *CPR*.

2.1 The *Groundwork* and the Idea of the Moral Law

The argument of the *GR* follows a three-part structure in an attempt to establish that there is a domain of laws that apply to our conduct and that a system of morality may be established on rational grounds. Christine Korsgaard says more simply that in the *GR*, Kant is attempting to establish that there is such a thing as morality.⁹⁴ This makes sense, given his overall progression from what he assumes are common sense ideas of morality to an evaluation of morality's metaphysical status. This search for a system of morality culminates in the establishment of the "supreme principle of morality"—the famous categorical imperative.⁹⁵ Because Kant's primary aim is to show that there is such a thing as morality, as just alluded to, he uses Section I of the *GR* to analyze common ways of thinking about morality in order to clarify just what morality would look like, were it to exist. What he intends to do here is simply to establish what morality is according to common sense—to establish what human beings typically mean when they talk about

⁹⁴ Christine Korsgaard, "Introduction to the *Groundwork of the Metaphysics of Morals*" in *Groundwork of the Metaphysics of Morals* trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 1997), x.

⁹⁵ Immanuel Kant, *Groundwork of the Metaphysics of Morals* trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 1997), 4:392.

acting morally and appraise someone's actions as either praiseworthy or blameworthy, moral or immoral.⁹⁶ This analysis occurs independently of the concerns of theoretical reason, which is most obviously indicated by the title of this section: "Transition from common rational to philosophic moral cognition."⁹⁷ Here Kant determines that the good will is the highest of all good things and that it is the condition of every other good⁹⁸ because it alone is capable of bringing all actions into conformity with universal ends.⁹⁹ Though he does not make it entirely clear in this section, it is safe to assume that he believes bringing all actions into conformity with universal ends is a good and proper goal because it entails the universality of reason and, though it may not necessarily do so, it is therefore *possible* for it to harmonize with the universality of reason present in and accessible to all other rational beings. He also determines that the concept of the good will is further bound up with the concept of duty because the good will wills what is universally good unconditionally, without regard for what is beneficial to itself or the particular agent who is doing the willing.¹⁰⁰ He explains briefly in a footnote that moral

⁹⁶ Interestingly, there is no proof that these allegedly common ways of thinking about morality are all that common after all. Instead, they represent what Kant believes common understanding of morality to be. It is quite likely that the opinion of the average citizen of Königsberg in 1785 (when the *GR* was first published) would more readily approximate such a single definition as Kant gives than a present citizen of any country (especially in light of globalization and the growing consciousness of a wide variety of definitions of morality), though this still remains unproven and could certainly be reasonably disputed. However, none of this is relevant to Kant's argument here in the *GR*, for in the *GR* he aims to establish whether this concept of morality can be held in place, regardless of how commonly accepted it may be. For all intents and purposes, Kant could have come up with this definition of morality by himself, without any common citizen agreeing with it; though this would seem to undermine the appeal of his moral system, it would not undermine its claim to rational proof.

⁹⁷ *GR*, 4:393.

⁹⁸ *GR*, 4:396.

⁹⁹ *GR*, 4:393.

¹⁰⁰ *GR*, 4:397.

actions are made according to maxims, which are the subjective principles of volition that particular agents choose to act upon.¹⁰¹ The maxim is the subjective principle of volition whereas the objective principle of volition is the practical law, which would ideally serve as the subjective principle of volition were the agent in question to act completely in accordance with reason. This leads to a preliminary articulation of the categorical imperative (which will be explained shortly in this section in fuller detail), according to which conformity to the universal law as such or conformity to the form of universality as such, serves as the ultimate principle of a good will.¹⁰²

With these alleged principles of common sense in hand, Kant proceeds to Section II of the *GR*, where he develops these principles with rigor and complexity, unpacking the concepts he believes to be implicit in them. Among the concepts he further develops from Section I is the previously mentioned idea of subjective and objective principles of volition (maxims and the practical law, respectively). He maintains that because human beings possess a subjectively imperfect will and do not necessarily act in accordance with the practical law (unlike, for example, the divine will whose volition is necessarily in accord with the practical law),¹⁰³ the practical law is therefore presented to us in the form of an ought—an imperative.¹⁰⁴ Whereas everything in nature *already* conforms to immutable and universal natural laws, morality presents a system of laws to which human action does not already conform but *ought* to conform.¹⁰⁵ From here, Kant proceeds to explain that there are only two types of imperatives—hypothetical and categorical

¹⁰¹ *GR*, 4:402.

¹⁰² *GR*, 4:402.

¹⁰³ *GR*, 4:414.

¹⁰⁴ *GR*, 4:413.

¹⁰⁵ *GR*, 4:412.

imperatives. A hypothetical imperative states that some action is good or necessary in virtue of the fact that it brings about some other more primary good or, in other words, it is good as a means to something else.¹⁰⁶ It has the structure: if you want to attain some primary good ‘G’, then some particular action ‘A’ is necessary. For example, if you want to be healthy and live long, you will engage in some form of regular physical activity (exercise). In contrast, a categorical imperative (the second type of imperative after the hypothetical imperative) states that some action is objectively necessary of itself, without reference to some other end, or in other words, such an action is good in itself.¹⁰⁷ Shortly after making this distinction between hypothetical and categorical imperatives, Kant makes the claim that there is only one single categorical imperative (from now on referred to as the CI) and it is: “act only in accordance with that maxim through which you can at the same time will that it become a universal law.”¹⁰⁸ This is Kant’s famous articulation of the supreme principle of morality. For the sake of clarity, it is important to note in advance that the categorical imperative is the same thing as the practical law or the universal form of reason, it entails the universality of reason, and it serves as the rational basis of the moral law.

It is not immediately clear as one reads the text, how exactly he moves from the concept of a categorical imperative in general to this formulation of the one and only categorical imperative as the fundamental principle of morality. I would like to suggest that it is a result of several things. First of all, before stating the formula of the one categorical imperative, Kant says that a categorical imperative must contain only the

¹⁰⁶ *GR*, 4:414.

¹⁰⁷ *GR*, 4:414.

¹⁰⁸ *GR*, 4:421.

universality of a law as such, without reference to any desired end. Unlike the hypothetical imperative whose command is conditional upon a commitment to attaining some prior end, the categorical imperative applies to all actions aiming to attain any and all ends and therefore commands unconditionally. In summary, a categorical imperative has the form of universal law as such and commands unconditionally. Because morality presents us with a picture of how the world ought to be and because hypothetical imperatives are always dependent on some other condition, often in the form of the attainment of some prior end, only categorical imperatives will work for morality because only they can command unconditional obedience, regardless of any particular agent's desired or chosen ends. Kant claims that there is only one categorical imperative essentially because the form of universal law as such can only be stated in one way. Or, because any condition on the application of an imperative would make that imperative hypothetical, *the* categorical imperative is the only categorical imperative. (From here on I will refer to this categorical imperative—*the* categorical imperative—in abbreviated form as the CI.) Any action that is good in itself must therefore take the form of the CI because it prescribes the only course of action that is possibly consonant with a plurality of ends (namely, all other rational ends). This is clarified somewhat when Kant distinguishes between rules of skill, counsels of prudence, and the commands of morality; while hypothetical imperatives yield no more than rules of skill and counsels of prudence, only the CI yields unconditional commands and therefore only the CI is fit to serve as a fundamental principle of morality.

Secondly, earlier in Section II of the *GR* and prior to introducing the CI, Kant claims:

[A]ll moral concepts [must] have their seat and origin completely a priori in reason... and they cannot be abstracted from any empirical and therefore merely contingent cognitions; that just in the 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; that it is not only a requirement of the greatest necessity for theoretical purposes, when it is a matter merely of speculation, but also of the greatest practical importance to draw its concepts and laws from pure reason, to set them forth pure and unmixed, and indeed to determine the extent of this entire practical or pure rational cognition, that is, to determine the entire faculty of pure practical reason.¹⁰⁹

He adds that unless we are in possession of such an a priori rational ground for morality, it would be impossible to sufficiently anchor moral action and instruction because otherwise we are provided with no moral ideal that we can strive towards.¹¹⁰ This is a large part of what motivates Kant to establish the enterprise of morality on a priori rational grounds. For this reason, he believes that morality must be knowable rationally and a priori. Only the CI supplies such an a priori, rational ground for moral action because only the CI works independently from all sensible or empirical conditions. As we observed earlier, any condition whatsoever placed on an imperative makes that imperative hypothetical and hypothetical imperatives command only conditional obedience. Therefore, because only the CI can command unconditional obedience, only the CI is fit for supplying a fundamental principle of morality.

Finally, in the first sentence of Section I of the *GR*, Kant states that “[i]t is impossible to think of anything at all in the world, or indeed even beyond it, that could be

¹⁰⁹ *GR*, 4:411-12.

¹¹⁰ *GR*, 4:412.

considered good without limitation except a good will.”¹¹¹ The unlimited goodness of the good will is one of the core components of what Kant attributes to the common sense concept of morality. As such, he takes it to form a necessary condition of any moral system deserving the name. The will is “the capacity to act in accordance with the representation of laws”¹¹² and because a rule like the CI, to “act only in accordance with that maxim through which you can at the same time will that it become a universal law,”¹¹³ can only be followed by something that has the capacity to act in accordance with the representation of laws, the will is the only thing we know of that can contain the law of universality as such (the CI) within itself. For that reason, as previously mentioned, it contains within it the capacity to recognize the only formula that is consonant with an action that is good in itself. For this reason, only the CI can prescribe action that is unconditionally good or good in itself, unlike hypothetical imperatives that declare the goodness of an act to be conditional on some other good or good “as a means to something else,” any will that is unconditionally good will necessarily follow the CI and nothing else as its formula for action.¹¹⁴ In this way, only the CI (and it alone) is suitable to provide a truly objective principle of action for the will and thereby serve as the fundamental principle of moral action. In summary, because only the CI provides a sufficient ground for the unconditionally good will, an a priori rational ground that Kant deems necessary and sufficient for moral action, and a formula of necessary action regardless of a particular agent’s ends, Kant claims that it is the primary or fundamental principle of morality.

¹¹¹ *GR*, 4:393.

¹¹² *GR*, 4:412.

¹¹³ *GR*, 4:421.

¹¹⁴ *GR*, 4:414.

Shortly after laying out the CI as the fundamental principle of morality that commands action according to the way the world ought to be, Kant focuses his attention more specifically on the constitution of human nature. He says that the will is “thought as a capacity to determine itself to acting [sic] in conformity with the representation of certain laws.”¹¹⁵ From this definition of the will he moves quite quickly via an argument that we will now examine to the conclusion that “the human being and in general every rational being exists as an end in itself, not merely as a means to be used by this or that will at its discretion”¹¹⁶ and from this conclusion to what he calls the practical imperative, to “[s]o 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.”¹¹⁷ Yet, his argument here is problematic because it is not clear why he makes these moves as one reads the text and further explanation is required to show that his conclusion, in this case the contents of the practical imperative, indeed does follow from previous conclusions.

I would like to argue that the command to treat every rational being as an end in itself and never merely as a means (the basic tenet of the practical imperative) is grounded by the CI in such a way that the CI is prior to and more basic than it. First of all, the CI comes prior to this command in the *GR* (on page 4:421, while this command is found on page 4:428) and it would seem reasonable to assume that Kant would outline the more basic principle first before proceeding to explain what can be derived from it. Yet this circumstantial evidence is not by itself conclusive so it is necessary to look at the immediate textual setting in which the idea of rational beings as ends and never mere

¹¹⁵ *GR*, 4:427.

¹¹⁶ *GR*, 4:428.

¹¹⁷ *GR*, 4:429.

means is stated. Immediately after stating this idea (which comes seemingly out of nowhere) and before stating it as the “practical imperative,” Kant explains that all the inclinations that are grounded in nature have only conditional worth and cannot confer absolute worth on the being to which they happen to belong.¹¹⁸ In contrast, beings with reason, *in virtue of possessing that reason/rational nature*, are ends in themselves and so possess unconditional value.¹¹⁹ It is natural to understand him as saying that it is rational nature that confers this unconditional value on such beings. Moreover, as the supreme practical principle that supplies rational form to the will¹²⁰ and, as we have seen, applies to all actions unconditionally, the CI is the distillation of this rational nature. It is therefore in virtue of possessing this rational nature and capacity to legislate action in accordance with it that human beings are to be treated as ends in themselves and never merely as means.

Beings the existence of which rests not on our will but on nature, if they are beings without reason, still have only a relative worth, as means, and are therefore called things, whereas rational beings are called persons because their nature already marks them out as an end in itself, that is, as something that may not be used merely as a means, and hence so far limits all choice (and is an object of respect).¹²¹

Shortly after this, he says that “without it [it referring to a being with rational nature] nothing of absolute worth would be found anywhere; but if all worth were conditional and therefore contingent, then no supreme practical principle for reason could be found anywhere.”¹²² This passage indicates that it is the capacity of human beings qua rational

¹¹⁸ *GR*, 4:428.

¹¹⁹ *GR*, 4:428.

¹²⁰ *GR*, 4:412.

¹²¹ *GR*, 4:428.

¹²² *GR*, 4:428.

beings able to act in accordance with the CI, which alone can prescribe action in accordance with unconditional or absolute worth, that marks them out as ends in themselves because without them, *nothing* could be recognized as being of unconditional or absolute worth. It is their capacity to legislate universal law and thereby confer the status of unconditional worth on actions that makes them worthy of treatment as ends and not merely as means, for Kant. This requires prior establishment of what that universal law is, for it is only in virtue of legislating *it* that these beings are so valuable and this universal law is nothing other than the one and only CI. Put differently, only rational beings are capable of legislating in accordance with the CI that is more basic than them and, according to Kant, this is a capacity that deserves respect.¹²³

The practical imperative must further be considered in light of the earlier distinction made between hypothetical and categorical imperatives; for this distinction reinforces the distinction between things with relative worth and persons with infinite worth because they can legislate in accordance with the CI. There, only categorical imperatives were shown to contain the formula of action that serves as an end in itself because only categorical imperatives prescribe action that is good in itself and possibly consonant with a multiplicity of universal ends, whereas hypothetical imperatives prescribe action only for the purpose of attaining some prior/other end. Therefore, if a being that is aware of the CI also has a will which affords the capacity to act in accordance with the CI, it is also possible that its actions are good in themselves or may be considered as ends in themselves. Furthermore, in virtue of it being possible for other rational agents to act in such a way, it makes sense to ensure that one's own actions

¹²³ *GR*, 4:436.

accord with and do not violate this property of other agents and so ought to treat them “always as an end, never merely as a means.”¹²⁴ In this respect, Kant holds that persons are very different from things; for while things possess no rational nature and have only relative worth, persons possess a rational nature and their actions may take the form of universality or ends in themselves, therefore action towards them ought to be consonant with their being so considered ends in themselves—this marks them out as a putative limit of choice and object of respect.¹²⁵ This status would not be possible without the prior establishment of the CI as a properly basic principle of morality and formula for action according to how the world ought to be.

The capacity of individual persons not only to be aware of but to act on the basis of the CI, Kant calls the principle of autonomy. Once the capacity to think rationally, in accordance with the CI, is coupled with the faculty of the will, which is the capacity to act according to the representation of a given law or set of laws, it follows that the will itself may be considered a rational faculty. This is because once it is coupled with the awareness of the CI as a formula for possible action, if freedom makes a difference in the world, it is then possible for the will to act in accordance with the CI so that its actions take the form of universal law; by this process, the universal law therefore becomes the proper object of every rational will.¹²⁶ In this way, every rational will contains the form of universal law—the CI—within it. In virtue of being a rational being with a will, each human being is able to be aware of the basic principle of morality in the CI. As a result, each human being is also aware of what her moral duty is, namely to “act only in

¹²⁴ *GR*, 4:429.

¹²⁵ *GR*, 4:428.

¹²⁶ *GR*, 4:432.

accordance with that maxim through which you can at the same time will that it become a universal law.”¹²⁷ But the origin of that awareness is internal to her person in the faculty of her own rational will.

It might be said here that this rational form of morality is only present in the ideally rational will which the actual will may not and, in fact, is more than likely not to, line up with. This challenge will be dealt with in more detail in Section 3.3 of Chapter III on external justification but in the meantime, it suffices to say that as long as this moral ideal is cognitively accessible to the individual agent from within herself, it provides a purpose and end for action in such a way that, although her action may not perfectly line up with or reach it, she may act so as to approximate it in greater or lesser degrees and thereby use it as a standard of justification. The crucial point here is, again, that such awareness of the moral standard originates from within the structure of an agent’s will. This is what Kant had in mind when he said that “the human being is bound to laws by his duty, but it never occurred to them [i.e. those who have previously attempted to justify morality] that he is subject only to laws given by himself but still universal and that he is bound only to act in conformity with his own will, which, however, in accordance with nature’s end is a will giving universal law.”¹²⁸ This is perhaps the clearest statement of the formula of autonomy in the *GR* and it points out that the individual’s own will is the primary means by which morality is made known to her. Therefore, for an individual to act in accordance with morality is for her to act autonomously, in accordance with her own will. Henry Allison says that an agent acts autonomously when she takes the categorical imperative as a first-order ethical

¹²⁷ *GR*, 4:421.

¹²⁸ *GR*, 4:432.

principle.¹²⁹ With this in mind, Allison concludes that the formula of autonomy “does not add anything to the other formulas by way of specifying further the criteria of dutiful action, but it does characterize the regulative idea under which rational agents ought to act.”¹³⁰

As it will become important shortly in Section 2.2.1, it is helpful at this point to briefly define what Kant calls heteronomy. In contrast to autonomy, Kant defines heteronomy as anything that is motivated by something other than the agent’s autonomous will (as defined above).¹³¹ It therefore includes anything of empirical origin, for he believes that all moral concepts must originate completely a priori in reason.¹³² As a result, both deterministic laws of nature like those described in the antithesis of the Third Antinomy and desires that initially *seem* to arise from within our own person (for example, such desires might include hunger/desire for food, thirst/desire for water, or sexual desire), are truly external to our person as a whole because both originate in natural forces that are beyond our control. As Kant says, any imperative arising from them is only ever conditional or hypothetical because it places something/some incentive (such as the satisfaction of a desire) prior to its command and in doing so, forfeits the possibility of delivering an unconditional, categorical command in the same way that the CI is able to do.¹³³ Allison restates this helpfully when he says that Kant is thereby claiming that all heteronomous moral theories reduce the categorical imperative to a

¹²⁹ Henry Allison, *Kant’s Theory of Freedom*, 105.

¹³⁰ *Ibid.*, 106.

¹³¹ *GR*, 4:433.

¹³² *GR*, 4:411.

¹³³ *GR*, 4:444.

hypothetical imperative.¹³⁴ Actions in accordance with such theories are not consistent with autonomy, which requires that action conform to universal law; such actions are thus deemed heteronymous.

In summary, in Section II of the *GR*, Kant has established that a system of morality may be conceived of a priori on rational grounds as the way human beings ought to act. Because morality conceives of action in the form of an ought, it means that human action does not necessarily operate this way and in fact, as we are aware, frequently does not operate this way. However, if human action were ordered perfectly in accordance with reason, it would be this way—it would operate so that every action would contain within it the form of universality and so *could* hold as a possible universal law of nature.¹³⁵ Because human beings are not completely rational (or do not act in necessary/perfect accordance with reason as, for example, the divine will does), this system of morality remains something we do not already do but *ought* to do—it presents itself to us as a system of duty and an object of life-long striving.¹³⁶ At least, this is the system of morality that we are obligated to act in accordance with if free moral action is possible. Section II of the *GR* thus presents us with a preliminary version of the moral law in the categorical imperative, as it is not yet evident that we are permitted to think of ourselves as *freely* acting in accordance with it. To determine this, a discussion of the concept of freedom and its relation to the moral law is required. Kant undertakes this discussion in Section III of the *GR*.

¹³⁴ Henry Allison, *Kant's Theory of Freedom*, 100.

¹³⁵ *GR*, 4:421.

¹³⁶ *GR*, 4:439.

2.2 From Meta-Compatibilism to the Moral Law

Section III of the *GR* sets out to define how a system of morality is metaphysically possible; this being the case, it forms a key link between the problematic concept of freedom in the *CPR* and the more robust concept of freedom and system of morality that he advances in the *GR* and *CPrR*. In Chapter I of this thesis, I outlined the theory of meta-compatibilism and argued that it is the most natural and defensible way of interpreting Kant's resolution to the Third Antinomy and his consequent theory of freedom. The Third Antinomy required us to postulate both a free first cause and a causally closed/deterministic universe, for both are required to render a picture of the world complete on rational grounds. As a result of the rational need to assume a free first cause, we were permitted to believe in free agency on rational grounds. The upshot of meta-compatibilism is therefore that human beings are able to treat themselves as free on rational grounds regardless of the indeterminate metaphysical status of actual freedom. In what follows of this section, I will show how Kant defines the relationship between freedom and the moral law and argues that by recognizing the concept of freedom, the moral law may be recognized and applied as a fact of knowledge. This means that I will discuss Kant's idea that the moral law gains objective reality through the concept of freedom in *GR* Section III. Then I will turn to the *CPrR*, where he explains that the concept of freedom may be derived from an awareness of the moral law.

2.2.1 Freedom and Choice

In Section III of the *GR*, Kant directly takes up the question of free moral action and argues that if “freedom of the will is presupposed, knowledge of what we ought to do together with its principle [the CI or moral law] follows from it by mere analysis of its concept.”¹³⁷ What he means by this is that if freedom of the will is presupposed, then the possibility of free action in accordance with the CI and the contents of morality follows. The point is that the only true exercise of freedom (or the only time we have autonomy) occurs when we self-legislate in accordance with the CI. This self-legislating form of freedom is what Kant calls positive freedom and only positive freedom counts for autonomous action. In what follows I will undertake to explain how this is the case.

Kant first states this idea in metaphysically aggressive¹³⁸ fashion, seeming to imply that autonomy and the universal principle of morality can actually be derived from the concept of freedom.

As a rational being, and thus as a being belonging to the intelligible world, the human being can never think of the causality of his own will otherwise than under the idea of freedom; for independence from the determining causes of the world of sense (which reason must always ascribe to itself) is freedom. With the idea of freedom the concept of autonomy is now inseparably combined, and with the concept of autonomy the universal principle of morality, which in idea is the ground of all actions of rational beings, just as the law of nature is the ground of all appearances.¹³⁹

¹³⁷ *GR*, 4:447.

¹³⁸ A phrase used in different context by Allan Wood (Allan Wood, “Kant’s Compatibilism,” 75).

¹³⁹ *GR*, 4:452-3.

On the surface, it appears he is claiming that the moral law can be directly derived from the concept of freedom. However, in fact, he is presupposing the work he has already done in Section II of the *GR* and saying that if we accept the idea of freedom, we must accept that the only genuine freedom that exists is self-legislation in accordance with the CI and so we must think of ourselves as free moral agents. This is evident in his discussion just prior to this passage. He goes into great detail describing the distinction between a world of sense and a world of understanding that his epistemological framework of transcendental idealism results in.¹⁴⁰ This is not much more than a brief review of what he already established in the *CPR* in the resolution to the Third Antinomy. At this point Kant assumes that the arguments he made in the *CPR* are convincing and indeed do support the adoption of both of these worlds as two respective rational standpoints that the human agent is permitted to understand herself to inhabit; Kant believes we are licensed to inhabit these two standpoints on rational grounds. Given the acceptance of these two standpoints, he says here in the *GR* that the rational agent can “cognize laws from the use of his powers and consequently for all his actions; first, insofar as he belongs to the world of sense, under laws of nature (heteronomy); second, as belonging to the intelligible world, under laws which, being independent of nature, are not empirical but grounded merely in reason.”¹⁴¹ As we saw above (in Section 2.1) in our discussion of *GR* Section II, laws grounded exclusively in reason can only be imperatives of categorical form and there is only one categorical imperative—the CI or moral law,

¹⁴⁰ *GR*, 4:451.

¹⁴¹ *GR*, 4:452.

which says to “act only in accordance with that maxim through which you can at the same time will that it become a universal law.”¹⁴²

While speculative reason only postulates a need for the noumenal in the *CPR*, practical reason goes further than this in the *GR* by establishing the CI as the ground for morality and right action. Importantly, freedom is a sort of link between the noumenal (intelligible character) and morality because it enables morality to be possible as something that can apply to human existence and therefore more than an empty intellectual exercise. So important was freedom to Kant that he would later say in the *CPrR* that it is the keystone of his entire theory: “the concept of freedom... constitutes the keystone of the whole structure of a system of pure reason, even of speculative reason.”¹⁴³

A rational being that believes herself to be free will be free in both a negative sense—free *from* being determined by natural laws outside of her own will, as well as free in a positive sense—free *to* effect causes in accordance with a possible system of universal moral laws. This is what Kant has in mind when, at the outset of *GR* Section III, he says:

The preceding definition of freedom is *negative* and therefore unfruitful for insight into its essence; but there flows from it a *positive* concept of freedom, which is so much the richer and more fruitful. Since the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with natural laws, is not for that reason lawless but must instead be a causality in accordance with immutable [or rationally undeniable] laws but of a special kind; for otherwise a free will would be an absurdity.¹⁴⁴

¹⁴² *GR*, 4:421.

¹⁴³ Immanuel Kant, *Critique of Practical Reason* trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 1997), 5:3-4.

¹⁴⁴ *GR*, 4:446.

Both negative and positive freedom count as free action in the sense that they are motivated by an internal choice and not necessarily determined by alien causes. But if an agent *chooses* to act in accordance with her desires, she is choosing to allow herself to be controlled by something alien and external to her own will and is thereby lapsing into a form of heteronomy. Recall that heteronomy is action in accordance with laws that are conditional (cannot be universal) and whose origin is external to the agent in question.¹⁴⁵ Importantly, this lapse into heteronomy was preceded by an act of free choice and so it is something the agent is accountable for because it properly belongs to her.

On the other hand, positive freedom for Kant is action in accordance with the form of universality inherent in the CI and continuous with the moral law. Kant sometimes refers to positive freedom as though only it is truly free and all other action (including action under the rubric of negative freedom) is determined. For example, as quoted above, he says that without the “positive concept of freedom.... A free will would be an absurdity.”¹⁴⁶ But he is not discounting the action of negative freedom as being freely chosen. Instead, he means that because the output of this action is in accordance with a heteronomous principle and the agent *allowed* herself to be motivated by this heteronomous principle, the action would not be free though the choice behind it was. For while both are freely chosen, only the latter contains no admixture of heteronomous principles while the former does.¹⁴⁷ In this way, all free action does not follow the form of immutable (unchanging, necessary, and universal) law as Kant says.

¹⁴⁵ *GR*, 4:433.

¹⁴⁶ *GR*, 4:446.

¹⁴⁷ Henry Allison explains the concept of heteronomy well when he says: “[s]uch an [external/heteronomous] object can be said to ‘give the law to the will’ just in case an

Rather, action as positively free does follow the form of universality inherent in the CI and it is possible for all free action to attain this form though it may often not, in fact, do so. The only rational alternative to heteronomy for Kant is autonomy, where one strives to act in accordance with the moral law that is evident in the principle of the will—the CI. And autonomy requires that an agent’s action not be determined by anything outside her person as a whole, including deterministic laws of nature as well as desires over which she has no control (for example, hunger, thirst, and sexual desire).

2.2.2 Awareness of Freedom Through the Moral Law

This brief sketch of the major points of a critique of pure practical reason are taken up in more detail and afforded more clarity in the *CPrR*. As just mentioned in Section 2.2.1, here Kant emphasizes that the concept of freedom “constitutes the keystone of the whole structure of a system of pure reason, even of speculative reason.”¹⁴⁸ He contends that the faculty of practical reason and the awareness of the moral law that it reveals, now grants objective reality to the formerly problematic concept

agent is moved to act by an inclination or desire for the realization of that state of affairs. This being ‘moved to act’ should not be understood in a causal sense, however, since actions resulting from such motivation are still genuine intentional actions based on maxims rather than mere bits of behavior causally conditioned by stimuli. The point is rather that as based on inclination these maxims reflect the agent’s needs as a sensuous being, needs that are themselves explicable in terms of laws of nature. Accordingly, it must be because these needs provide the only sources of motivation or reasons to act, not because they are the causal determinants of behavior, that the heteronymous agent is, qua agent, subject ‘only to the law of nature—the law of his own needs’ (*GR*, 4:439),” (Henry Allison, *Kant’s Theory of Freedom*, 97).

¹⁴⁸ *CPrR*, 5:3-4.

of freedom in the *CPR*.¹⁴⁹ In what follows in this section, I will outline how Kant derives the concept of freedom from the moral law in the *CPrR* and explain how the moral law can be understood as a basic and properly rational belief, or as Kant says, how it is a fact of reason itself.

In Theorem I and II of the *CPrR*, Kant discusses the possibility of material and empirical principles acting as a determining ground of the will. He resolves here that any action that is motivated by an empirical principle cannot produce a practical law. For example, such a motivating empirical principle might include a desire for pleasure. He maintains that the conditions of satisfaction for any such desire for pleasure can vary significantly from person to person because it is based on the representation of some object and therefore depends on the existence of that object for its satisfaction.¹⁵⁰ Because it belongs in this way to the senses and not to the understanding and because the senses cannot grant the objective necessity of that object, or in other words, cannot guarantee that that object will exist, such an empirical principle cannot in principle deliver a ground from which an objective practical law may be derived. For Kant, all such material principles are equally incapable of producing a practical law with objective necessity.¹⁵¹ This makes sense because a law or principle that cannot apply objectively and unconditionally cannot be applied consistently throughout varying situations and circumstances and it would seem that any law suitable for universal (moral) legislation should be able to do so. This is in sharp contrast to the CI, which prescribes only what is unconditionally necessary and therefore necessary for all agents in all circumstances

¹⁴⁹ *CPrR*, 5:6.

¹⁵⁰ *CPrR*, 5:22.

¹⁵¹ *CPrR*, 5:21.

regardless of what their particular desires are. Although he does not employ the term at this stage in the *CPrR*, these material principles may be understood as heteronomous causes as defined in the *GR* and explained above.

Once one prevents everything material or empirical from acting as a motivational ground, one is left only with the mere form of giving universal law.¹⁵² In other words, once one gives up any particular end that one hopes to achieve in one's actions, one is left with the idea of making all ends and therefore universality itself one's ultimate end and aim in action. This is what it means to possess the mere form of universal law. It consists in giving up what is particular for the sake of what is universal in principle. The idea of the universal law is elucidated by Kant's remark on the definition of the principles of pure practical reason where he says that anything that may count as a practical law must be a product of reason, from which alone can be derived the concept of necessity.¹⁵³ Kant's aim then in Theorem III is to show that either we cannot think of maxims as universal laws because they grant no objective necessity, or we must assume that their form alone, with necessary disregard for material content, makes them "fit for a giving of universal law, of itself, and alone makes them practical laws."¹⁵⁴ This is the derivation of the moral law in the *CPrR*.

It is interesting to note that Kant presents the moral law to the reader in the form of an either/or with the strategy of reducing the opposing position to absurdity (*reductio ad absurdum*). The opposing principle here is anything that puts an empirical or external (that is, external to the agent in question) principle/law where he contends only the moral

¹⁵² *CPrR*, 5:27.

¹⁵³ *CPrR*, 5:20.

¹⁵⁴ *CPrR*, 5:27.

and universal law should stand as the form of one's maxim. In his remark to Theorem III, he suggests that if all individuals proceeded to act on the basis of empirical principles, for example, each seeking her own pleasure or happiness before all other things (thus elevating that principle into a universal law), chaos and disorder would result.¹⁵⁵ On the other hand, Kant maintains that it is only by acting in accord with the form of the universal law that harmony and order may be established.

After deducing the moral law in Theorem III, Kant proceeds to outline the transition of freedom from a problematic to an objective concept in his discussion of Problem I and II in the *CPrR*. He claims that once we conceive of the moral law, it presents itself to us as a possible defining ground of the will independently of our sensible conditions and once this happens, we necessarily cognize the concept of freedom as a possible determining ground of the will.¹⁵⁶ What he means by this is that once we become aware of the moral law, we become aware of our ability to do otherwise than what nature compels us to do and thereby become aware of our own freedom in the process. For then we can choose whether to act according to what our desires (sensible conditions) urge us to do or we can choose to act according to the rational ideal that the moral law presents us with. Where the resolution of the Third Antinomy only permitted us to posit freedom, Kant claims that awareness of the moral law enables the concept of freedom to attain objectivity, or objective reality, because it makes us directly and necessarily aware of freedom. In saying this, he means that once we are made aware of the moral law, we are capable of understanding ourselves as able to act in accordance with it by allowing it rather than our sensible nature to determine our actions—by acting

¹⁵⁵ *CPrR*, 5:28.

¹⁵⁶ *CPrR*, 5:30.

in accordance with it—thereby making it a determining ground of the will. With this capability comes the choice to act upon it or not—to employ our rational powers as a cause of action or to allow empirical (or heteronomous) conditions to determine our cause of action. In line with the principle of autonomy, we therefore become capable of acting as a form of causality unto ourselves. Kant illustrates this in the example of the man asked by a prince to give false testimony against another honorable man.

He would perhaps not venture to assert whether he would do it or not [give false testimony, that is], but he must admit without hesitation that it would be possible for him [not to give false testimony and therefore to defy the prince's request]. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes [or becomes directly aware of] freedom within him, which, without the moral law, would have remained unknown to him.¹⁵⁷

This illustrates how Kant now believes that if we are aware of the moral law, we are aware of the fact of our own freedom. This is a crucial change from the more skeptical *CPR* as there, freedom was only permitted but now it is allegedly known.

2.2.3 *Freedom and Morality: Transition From the CPR to the GR and CPrR*

How then does this idea of the moral law as a guide for action line up with meta-compatibilism as a way of interpreting the resolution to the Third Antinomy? As argued in Chapter I, the resolution to the Third Antinomy permits us to believe in freedom on rational grounds, in the face of the deterministic laws of nature. This means that we are licensed to take two standpoints on human action corresponding to the phenomenal (empirical) and noumenal (intelligible) aspects of our character, respectively. The

¹⁵⁷ *CPrR*, 5:30.

phenomenal standpoint recognizes that all actions are determined by prior causes—including both the deterministic laws of nature and human desire—while the noumenal standpoint believes that human freedom is compatible with this determinism and that our decisions can make a difference in the world. This two-standpoint view on the noumena-phenomena distinction is indicated by Kant’s expression “zwei Seiten,”¹⁵⁸ which is translated helpfully to “two-aspects” in the resolution to the Third Antinomy of the *CPR* but means more literally “two sides.” To clarify, the meta-compatibilism of the Third Antinomy only grants permissibility, and does not grant objective or necessary status, to the idea of freedom.

However, in Section III of the *GR* Kant argues that if freedom does exist, the moral law obtains objective reality and may be recognized and applied as a fact of knowledge. In other words, if freedom exists the moral law is necessarily activated as a policy of action and forces us to think of ourselves and our actions, as free and moral or immoral. Where the resolution of the Third Antinomy concluded with only the possibility of freedom, the *GR* establishes a standard of morality, in the form of the moral law, and purports to show that once we recognize the moral law, the autonomy of the will follows necessarily. Kant summarizes his view clearly when he says: “we now see that when we think of ourselves as free we transfer ourselves into the world of understanding as members of it and cognize autonomy of the will along with its consequence, morality.”¹⁵⁹ This highlights what I pointed out earlier in Section 2.2.1, that Kant holds that we can deduce the nature of the moral law from freedom and that the only genuine freedom that

¹⁵⁸ Immanuel Kant, *Kritik der Reinen Vernunft* (Hamburg: Felix Meiner Verlag, 1993), A538/B566.

¹⁵⁹ *GR*, 4:453.

exists is self-legislation in accordance with the moral law. Though, as the case of the man asked to give false testimony cited in Section 2.2.2 above shows, it is the moral law that first makes us conscious of our freedom.

Additionally, in the *GR* Kant reaffirms the simultaneous application of the idea of free-will and causal determinism to human action—more simply put, he reaffirms the compatibilist tenet of meta-compatibilism. In Section III he reminds us that because reason must assume the existence of things-in-themselves and therefore possesses a form of spontaneity over and against the understanding, which unites sensible representations under rules in consciousness, the rational being must regard herself under two standpoints that correspond to these respective faculties of the mind.¹⁶⁰ These two standpoints are first, the world of sense that operates under the deterministic laws of nature and desire and second, the world of reason that operates under the laws of reason independently of determining causes.¹⁶¹ It is instructive to note here that Kant’s original use of the German words “zwei Standpunkte”¹⁶² straightforwardly means “two standpoints” as it is translated in this edition of the *GR*. So it is reasonable to say that, given his language here, Kant is directly advocating a two-standpoint view on the noumena-phenomena distinction. This is less ambiguous, though continuous with his use of the expression “zwei Seiten,”¹⁶³ in the *CPR*.

So although his two-aspect view of noumena and phenomena remains consistent between the *CPR* and the *GR*, Kant’s argument for morality in the *GR* goes much beyond

¹⁶⁰ *GR*, 4:451.

¹⁶¹ *GR*, 4:452.

¹⁶² Immanuel Kant, *Grundlegung zur Metaphysik der Sitten* (Leipzig: Felix Meiner Verlag, 1925), 4:452.

¹⁶³ Immanuel Kant, *Kritik der Reinen Vernunft*, A538/B566.

the more skeptical and tentative character of meta-compatibilism in the *CPR*. In the *CPrR* he brings these two lines of argument together. He rearticulates the categorical imperative from the *GR* and calls it the “Fundamental law of pure practical reason,” which is to “so act that the maxim of your will could always hold at the same time as a principle in a giving of universal law.”¹⁶⁴ Like the CI, this law is immediate and unconditional; it does not prescribe itself for the purpose of attaining any end or goal that is of empirical origin or is external to itself or the will of the agent.¹⁶⁵ Kant claims that this law is valid for all rational beings insofar as they have a will (here the will amounts to the ability to determine one’s causality in accordance with representations of rules) and can act in accordance with a priori principles (the CI or moral law).¹⁶⁶ Then he describes both how the moral law can be deduced from freedom and how it is that we become aware of the moral law in our experience.¹⁶⁷ Along these lines, he believes that both freedom and the ability to act morally may be known to exist and each must be true in order to understand human actions as free and moral. These basic beliefs enable us to think of ourselves as moral agents and, as I will argue in Chapter III, two further conditions of moral responsibility can be derived from them—namely, internal motivation and objective justification.

¹⁶⁴ *CPrR*, 5:30.

¹⁶⁵ *CPrR*, 5:31.

¹⁶⁶ *CPrR*, 5:32.

¹⁶⁷ *CPrR*, 5:29-30.

2.3 Evaluating Kant

Before I proceed to define these two further conditions of moral responsibility and explain how they hang together in Kant's thought, I will offer a brief evaluation of what I have argued is Kant's position so far. This evaluation will consist of a look at the following objections to his position as defined in Chapters I and II. The first objection comes from Derk Pereboom and raises the question of whether the concept of freedom is good or useful to have in the first place. I will respond first by explaining some of what I think Kant could say in response to this. Then I will respond with a suggestion for how Kant's position could be modified and adapted within a consequentialist framework. To be clear, this suggestion constitutes a departure from Kant's own position as found in the *CPR*, *GR*, and *CPrR*. Nevertheless, it remains inspired by his position and in some ways compatible with it. Perhaps it constitutes a fruitful line of further inquiry to undertake in future work. The second objection deals with the problem of maxims that nearly everyone who reads Kant's account of moral deliberation comes up against. I believe this problem remains for Kant. The third and final objection is to Kant's extreme hostility to self-interest in morality. I will explain briefly why I believe this tenet of Kant's moral theory remains problematic though it need not be so.

2.3.1 *Objection to Freedom as a Useful Concept*

This objection is formed on the basis of primarily practical, moral reasons and it is directed toward Kant's theory of freedom as meta-compatibilism. It acknowledges the

general picture of meta-compatibilism with its limitation to the realm of beliefs in light of the limitation of human cognitive capacities, but it contests the practical grounds on which freedom is retained. In other words, this objection acknowledges that freedom is possible but holds that we do not need the concept of freedom in the first place. I anticipate this possible objection on the basis of some very forceful arguments presented by Derk Pereboom in his article “Kant on Transcendental Freedom.” Pereboom advances two major counterexamples against the practical grounds on which Kant aims to license faith in freedom.

He raises the first counterexample in response to Kant’s well-known passage on the malicious liar in the *CPR*.

[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. 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 empirical conditions of the deed, is fully free, and this deed is to be attributed entirely to its failure to act.¹⁶⁸

Pereboom believes Kant is suggesting “we have good practical reason to judge the liar blameworthy, and since blameworthiness requires transcendental freedom, we thereby have a good practical reason to believe that he is transcendently free.”¹⁶⁹ However, Pereboom argues that the mere possibility of transcendental freedom—which turns out to be, in fact, the inability to prove its impossibility—coupled with the condition that ascriptions of freedom are only superficially consistent (in other words, problematic because unclear), does not warrant responsibility for the malicious lie. It is not justified, according to Pereboom, to cause someone harm or deprive someone of liberty (as punishment would effectively do) on these grounds.¹⁷⁰

In response, I argue that belief in freedom and belief in the moral law are available to the malicious liar as well as being available to any other human being. Reason may function as a regulative principle regardless of one’s psychological history; no matter how bad one’s upbringing, company, and natural temper, it is always available for one to believe in and strive for freedom and the moral law, which both take reason as their cause.¹⁷¹ The cognitive faculties of the malicious liar may be unable to see into the true nature of his character and freedom, as it exists in its intelligible (noumenal) capacity, and his empirical character may have brought him to a point where morally bad actions flow naturally out of him as though from second nature; yet he remains consistently free to posit the moral law in the form of the rule of reason and his own

¹⁶⁸ *CPR*, A554-5/B582-3.

¹⁶⁹ Derk Pereboom, “Kant on Transcendental Freedom,” 563.

¹⁷⁰ *Ibid.*, 563-4.

¹⁷¹ *CPR*, A554-5/B582-3.

freedom to act in accord with it. For this reason, not only his belief in freedom but the belief in freedom in general may be retained and meta-compatibilism in that way upheld.

Pereboom's second counterexample comes from Kant's own work again but this time finds its source in the later *Metaphysics of Morals*. In this text, Kant gives the example of a last murderer remaining in prison in an imaginary society that is about to dissolve itself. Kant believes that this murderer should be executed for reasons of retributive desert alone and Pereboom maintains that not only is his execution not warranted on grounds of retributive desert, but is also not warranted on any further teleological or otherwise practical grounds either.¹⁷² He holds this because we cannot know whether this man actually acted freely or not and there is no one who will benefit from his death. That is, because he is the last murderer in prison in an island society about to dissolve itself, his death will serve as an example and be instructive to no one and hence will serve not even a teleological purpose. In light of this realization and the idea that we cannot know the nature of his agency, Pereboom argues that we cannot justify causing him harm (least of all in the form of execution) on moral grounds.¹⁷³

Pereboom makes a strong point here and it is important to take into account—if a person was caused to commit a crime irrespective of his own free choice or if the nature of his acting freely cannot be determined and there is absolutely no practical benefit to be gained from his punishment (consequentialist or otherwise), then he should not be punished (least of all executed) for his crime because that crime is not properly his. However, this point can be acknowledged and upheld without threatening the broader enterprise of ascribing freedom to human actions. For this case can be taken to constitute

¹⁷² Derk Pereboom, "Kant on Transcendental Freedom," 564.

¹⁷³ Ibid.

a limit on ascriptions of freedom. In the everyday worlds that we live in, where human beings interact with one another against a greater social background with all of its laws, policies, and institutions, Kant's practical grounds for the imputation of freedom do apply. Practical and sometimes consequentialist reasons for punishment (though perhaps not execution) do apply and more importantly, the laws, policies, and institutions that form a familiar part of our everyday lives require a prior theory of human agency in order to hold together. That is, without believing that there exists a moral law that we may freely adhere to and instantiate into the reality of our everyday lives, these laws, policies, and institutions by which we live peaceably and comfortably, could not be sustained.

Practically speaking, only in the isolated case of this thought experiment is it feasible to relax one's account of freedom to such a degree. Yet if such a scenario were to become a reality, I suggest that we could forego ascriptions of freedom without sustaining damage to the general enterprise of human agency as accountable and rational agency. In other words, because Kant's justificatory grounds for ascribing freedom to human beings are practical and it seems that in everyday life these practical justificatory grounds obtain, in everyday life we are therefore justified in ascribing freedom to human beings. In this way, Pereboom's intuitive reaction to Kant's island murderer thought experiment may be upheld without proving detrimental to the theory of human freedom as we typically make use of it. To do this, however, is to move away from Kant's own position for he remained firmly opposed to the import of any consequentialist thinking into a theory of responsibility, agency, or morality. This is why he built the entire edifice of meta-compatibilism in the first place—to preserve the possibility of human freedom in the face of a mechanistic understanding of human freedom. For Kant, any form of

consequentialist reasoning did not deserve to be implemented in matters of morality, as consequences are not morally relevant. Instead they could be implemented as policy for the purpose of altering social conditions.

2.3.2 *A Consequentialist Adaption of Kant*

Due to Pereboom's concern with Kant's apparently complete lack of concern for consequentialist reasoning, it is important to look closer at the relationship (or a possible relationship) between Kant and consequentialism. For the purpose of building a bridge between Kant and consequentialism, it could be conjectured that Kant would allow for consequentialist considerations to come into play in the case of an agent's diminished capacity to self-legislate (or reason and act in accordance with the moral law). Kant says in the metaphysical lectures that we possess practical freedom except "in tenderest childhood, or insanity, or in great sadness that is only a species of insanity."¹⁷⁴ So Kant does allow that there are circumstances in which individuals have a diminished capacity to self-legislate. Allan Wood expands on the notion of a diminished capacity to self-legislate.

But children too acquire gradually the capacity to resist impulses and to act for reasons. And various circumstances, including mental illness, addiction, brain damage, or psychic malfunction, can partially deprive adults of these capacities. Kantian ethics has no reason not to recognize these facts and admit that imputability is sometimes a matter of degree.¹⁷⁵

¹⁷⁴ *Lectures on Metaphysics* 28:182 as cited by Allan Wood [Allan Wood, *Kantian Ethics* (Cambridge: Cambridge University Press, 2007), 129].

¹⁷⁵ Allan Wood, *Kantian Ethics*, 129.

I have already clarified in the Introduction that I believe Kant is not quite concerned with imputing specific acts to agents but is rather concerned with the more general concept of accountability and what conditions must be in place for an agent to understand herself as morally responsible.¹⁷⁶ Nonetheless, the point that Wood is making remains—that not all people are always responsible and ought not to be held accountable for certain actions (usually stemming from an abnormal or not optimal state of mind). I suggest that this can allow for two points of view on the human agent—one according to which the individual is held accountable for her actions and treated as capable of desert (praise or blame) for them, and another by which she is regarded according to consequentialist considerations of guiding future behavior without necessary regard for desert. This introduces a composite concept of moral assessment that is inspired by though independent from Kant.

This composite concept of moral responsibility relies, first of all, on Kant's admittance of two possible points of view by which to regard human agency. As I pointed out earlier in Chapter II, Section 2.2.3, his use of language in the original German of both the *CPR* and *GR* is consonant with the idea of a two-aspect (dual-aspect) or two-standpoint view of human agency. More importantly though, he directly licenses both a free (rational, intelligible) and what might be best described as a (deterministic) scientific (anthropological, empirical) point of view on the subject in the *CPR*.

¹⁷⁶ This kind of knowledge is of a first-person variety but if the agent in question can understand herself as free and responsible, she can infer the same of other agents in the same way that she would infer the existence of other minds. The problem of other minds goes beyond the scope of this project and Kant does not have much to say about it in his writings but for the purposes of this project it is sufficient to say that the act of holding others accountable or responsible follows a similar pattern of inference. So if we grant that we have knowledge of other minds, we can accept others as morally responsible agents and therefore hold one another accountable.

Thus in regard to this empirical character there is no freedom, and according to this character we can consider the human being solely by observing, and, as happens in anthropology, by trying to investigate the moving causes of his actions physiologically. But if we consider the very same actions in relation to reason, not, to be sure, in relation to speculative reason, in order to explain them as regards their origin, but insofar as reason is the cause of producing them by themselves—in a word, if we compare them with reason in a practical respect—then we find a rule and order that is entirely other than the natural [deterministic] order.¹⁷⁷

What Kant has in mind by the “rule and order that is entirely other than the natural order” mentioned above, is the idea of free and rational agency that he defends as intelligible character in the resolution to the Third Antinomy where this passage is found.¹⁷⁸ Given the veracity of a deterministic scientific point of view on the individual agent, it is reasonable to think that while assuming this scientific point of view, it is permissible to engage in the project of redirecting human behavior toward certain desirable ends (for example, toward the further enhancement of human freedom and dignity). To do this would be to employ a consequentialist form of justification for corrective treatment and/or socialization.

This move can endorse, at least in a limited way, the project of social engineering that consequentialist theorists Owen Flanagan and Daniel Dennett advance. In his

¹⁷⁷ CPR, A550/B578.

¹⁷⁸ This idea is stated similarly by Korsgaard when, in describing her own two-aspect interpretation of Kant, she says that “[w]e view ourselves as phenomena when we take on the theoretical task of describing and explaining our behavior; we view ourselves as noumena when our practical task is one of deciding what to do” in Christine Korsgaard, *Creating the Kingdom of Ends*, 204. She does not cite the CPR at all when she says this so, though she and I are in agreement, I take this as my own idea especially as it issues from my previous work on freedom as meta-compatibilism in Part I.

discussion of William James in *The Science of Mind*, Flanagan explains such a consequentialist justification of moral responsibility.

The soft-determinist views as meaningless the idea that people are responsible for past actions in the sense that they could have done other than they in fact did. But the idea of promoting a sense of responsibility is utterly intelligible, since it is prospective. It involves the acquisition of knowledge, information, and motivation so that we have the ability to respond differently than we now do.¹⁷⁹

Although this position depends upon causal determinism, it is ironically also based on what Flanagan believes is the incredible malleability of human nature in Conscious Mental Life and the contents of the mind.

CML [Conscious Mental Life] is an incredibly plastic, informationally sensitive representational system. Any individual CML is capable therefore of responding to enormously varied features of the environment, in enormously complex ways. Our modes of social interaction, our ways of passing information about our attitudes toward the actions of others—from hugs and frowns to moral philosophizing, imprisonment, and canonization—are thus ways of taking advantage of the plasticity and information-processing capacity of CML, and enabling others to become able to respond, in moral and nonmoral situations, in ways which they cannot and do not, at present.¹⁸⁰

Kant would certainly recoil against a position like Flanagan's soft-determinism as defined here, because it views his kind of freedom and responsibility as meaningless. Kant may also not agree with the degree of malleability that Flanagan attributes to the structure of the mind and he would certainly not reject the role of desert in attributing moral responsibility, but nothing that Flanagan actually says here is inconsistent with a dual-aspect view of agency and responsibility. The difference is that a dual-aspect view of agency would prioritize the responsible aspect of the agent. Before taking the

¹⁷⁹ Ibid.

¹⁸⁰ Owen Flanagan, *The Science of Mind* (Cambridge: The MIT Press, 1991), 51.

consequentialist view/standpoint on a particular agent, according to the dual-aspect view, one would have to regard the agent as rational and responsible, thereby holding her accountable. Only if one were to subsequently discover that that particular agent had a diminished capacity to self-legislate for some reason or other (for example, mental illness, addiction, brain damage, or psychic malfunction, as Wood previously noted) would the dual-aspect view permit the assumption of the consequentialist standpoint. Importantly, one ought to do this carefully and cautiously so as not to wrongly exclude the possibility of free, rational action and therefore the possibility of desert.

Daniel Dennett advances a similar justification to Flanagan's for the idea of moral responsibility in his book *Freedom Evolves*. He describes the evolutionary enterprise of memetic engineering in which human beings essentially construct and deploy mechanisms of social change in their existing environment.¹⁸¹ He notes that this includes "the attempt to design and spread whole systems of human culture, ethical theories, political ideologies, systems of justice and government, a cornucopia of competing designs for living in social groups."¹⁸² Dennett goes on to say: "[o]ur communal process of memetic engineering over thousands of years continues today, and this book [his book, *Freedom Evolves*] is just part of that process" and that memetic engineering "can contribute, perhaps, to the refinement of our understanding of ourselves and our circumstances."¹⁸³ This leads him to conclude that those memes or social mechanisms that we construct now will make their effect known on the generations that follow us in a way that either aids or inhibits that generation's realization of certain goods (the primary

¹⁸¹ Daniel Dennett, *Freedom Evolves* (New York: Viking Penguin, 2003), 266.

¹⁸² *Ibid.*

¹⁸³ *Ibid.*, 303.

good that he has in mind here is human freedom).¹⁸⁴ This project of social engineering through memes is also permissible from a dual-aspect point of view because it coincides with the empirical or scientific point of view that he endorses in the *CPR*. Namely, that “we can consider the human beings solely by observing, and, as happens in anthropology, by trying to investigate the moving causes of his actions physiologically.”¹⁸⁵ And like the idea of the plasticity of CML in Flanagan, Dennett’s idea of memetic engineering does not contradict anything in the dual-aspect view of moral responsibility. Therefore we see that a consequentialist point of view is possible and permissible for a theory not far from Kant’s. The remaining question is when we should ascribe responsibility or apply corrective measures on grounds of consequentialist concern rather than desert.

Besides the cases cited by Wood (mental illness, addiction, brain damage, or psychic malfunction), consequentialist considerations should perhaps take over in marginal cases where, for example, an individual has somehow demonstrated that, while capable of doing so, she habitually has failed to and in the future will likely not act as the rational sort of agent that she can be and where her behavior threatens significant harm to others. To be more specific, a serial killer may be determined by her circumstances (bad upbringing/history of abuse, bad company, and bad temper) to do her heinous deeds so that it may be questionable to incarcerate her on the grounds of her having acting autonomously and therefore deserving punishment (though remember that according to Kant, this cannot be ruled out). Therefore, the state may incarcerate her in order to avoid harm to others (others would be safe at least while she is in prison) and attempt to rehabilitate her so that she will not repeat her crime. This perhaps, would be an

¹⁸⁴ *Ibid.*, 304-5.

¹⁸⁵ *CPR*, A550/B578.

appropriate circumstance in which to employ consequentialist justification for incarceration or the assignment of what would otherwise be referred to as punishment (though perhaps not so much for moral responsibility). Either way, it treats her as an agent who is accountable to others and whose capacity for accountability ought to be enabled and enhanced. This accords well with the dignity that Kant ascribes to all human beings when in the *GR* he states that “the human being and in general every rational being exists as an end in itself”¹⁸⁶ and therefore one should “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.”¹⁸⁷

The process that such a hardened criminal would hopefully undergo of maturation and eventual gaining of responsibility is an instance of Dennett’s notion of taking responsibility. Dennett contends that “people want to be held accountable” but that in order to be held accountable (one could just as easily use the word responsible here), it is necessary to allow oneself to be punished when one has “been caught out in some transgression.”¹⁸⁸ Therefore, he says that if one wants to be free one “must take responsibility.”¹⁸⁹ In other words, if an individual wants to be treated as a free and rational agent, capable of acting responsibly one must “take responsibility” for one’s actions.¹⁹⁰ If an individual does not take responsibility by regarding herself as an autonomous, internally motivated agent but chooses instead to “externalize the causes of

¹⁸⁶ *GR*, 4:428.

¹⁸⁷ *GR*, 4:429.

¹⁸⁸ Daniel Dennett, *Freedom Evolves*, 292.

¹⁸⁹ *Ibid.*

¹⁹⁰ *Ibid.*

[her] actions and deny responsibility”¹⁹¹ then this will likely bear out in her moral habits and actions by breaking down her rational and reliable expression of agency. As a result of her not taking responsibility/regarding herself as accountable, she would license others to regard her from the consequentialist standpoint. In this case, were she “caught out in some transgression,”¹⁹² the state, for example, would be justified in regarding her from the consequentialist standpoint as someone whose behavior must be controlled rather than as an autonomous and accountable moral agent.

In regarding an agent from the consequentialist point of view, the state may not be regarding her immediately as meriting dignity in virtue of her autonomous, rational agency (or more appropriately in this case, her lack thereof); however, the state may not ultimately act against it. Rather, because this freedom and rationality is always in reach of even the most depraved individual (an idea of Kant’s), the possibility always remains open that she will attain a sufficient degree of autonomy again (or for the first time). This is born out in Kant’s treatment of the malicious liar discussed in Section 2.3.1 above as well as the idea that the moral law is knowable to the individual in virtue of the most basic cognitive capacities, as also discussed earlier in Chapter II. From this idea we ought to conclude that no individual is beyond hope and as a result, it is not permissible for the state to engage in anything that would ultimately extinguish the hope of recovery. In particular, I have capital punishment in mind here. For the state to engage in capital punishment would be to violate that individual’s dignity and extinguish the possibility of her ever reaching the capacity that she possesses to be a well functioning (rational) moral agent, for this is a possible state that any rational human being may attain. (This is an

¹⁹¹ Ibid.

¹⁹² Ibid.

obvious departure from Kant as he had no qualms about capital punishment as illustrated in the desert island scenario from the *Metaphysics of Morals*, discussed above in Section 2.3.1.) Though the state may occupy a consequentialist and in a sense, object-oriented scientific point of view on an individual in order to redirect her future behavior and to protect others, it may not do so at the expense of violating the dignity of that individual.

Thus I argue that throughout the particular situations and conditions of human life there is room to regard individuals as both free and rational agents accountable for their actions and objects of a consequentialist rubric of justification for corrective measures. It is important to emphasize again that I have not attempted to establish anything near an exhaustive set of conditions for when to inhabit one or the other point of view. However, I do hope to have convincingly argued that there are grounds for developing a composite concept of moral assessment inspired by though not directly attributable to Kant. Most importantly, this composite concept has the result of preserving a robust notion of human dignity. I suggest that there is future work that could be done here in the way of a separate but related project.

2.3.3 *The Problem of the Unknowability of Maxims*

The third major objection to Kant's account of freedom and morality outlined so far in Chapters I and II is what has appeared to be the enduring problem of the unknowability of maxims in light of Kant's project of transcendental idealism. Recall in the *CPR* that he maintains that it is not possible to know ourselves as we truly are.¹⁹³ This

¹⁹³ *CPR*, B153.

clearly becomes an issue for morality when he says in the resolution to the Third Antinomy that because we have no insight into our intelligible character, “the real morality of actions (their merit and guilt), even that of our own conduct, therefore remains entirely hidden from us.”¹⁹⁴ A similar lack of self-knowledge is reaffirmed at one point in the *GR* where Kant says that we cannot be sure of our own maxims and adds that it is possible, in fact, that none of our maxims is morally good (i.e., line up with the moral law).¹⁹⁵ Though in the *CPrR* Kant revises the earlier skepticism of the *CPR* and says that we do have knowledge of the moral law, he does not afford much more clarity on the matter of our insight into our own selves. This is not problematic for his theory of moral responsibility because in it he is concerned only with the more general idea of accountability and he is not concerned to establish the particular facts about our actions in specific circumstances. This is adequate on a theoretical level, insofar as it establishes what it sets out to do, but I am concerned that it could have some negative effects on a practical level.

If our maxims are not clearly knowable to us it would seem that, combined with the rigorous standards inherent in the moral law, an agent who genuinely desires and is trying to be moral could never be sure of her own moral status, nor could any court of law. This might at first sound like a trivial point but it has the potential to abet a rather pernicious cycle of rumination or psychological self-flagellation if a particular moral agent is sensitive enough to the voice of the moral law. For if this were the case, it would seem quite possible that if she looked hard enough she could see her own self-interest in any or all of her maxims. And as it would never be possible to rule it out, she could even

¹⁹⁴ *CPR*, A551/B579.

¹⁹⁵ *GR*, 4:407.

be guilty of acting on immoral maxims that she is not conscious of. Kant certainly allows this when he says that “if we look more closely at the intentions and aspirations in [our maxims], we everywhere come upon the dear self, which is always turning up.”¹⁹⁶ I do not think this is a desirable outcome for a moral theory as it throws a wrench into the place where the gears of deliberation and action meet, making it impossible to get at the bottom of our own reasons for action. It seems instead that we ought to be able to know the moral status of our actions in order to act rightly and effectively undertake the project of character development and not least of all, to be happy with ourselves. It is uncertain whether Kant can get around this problem due to his commitment to transcendental idealism; for it depends on how we define transcendental idealism as to what we do with it and whether it is preferable to change our definition of it.

2.3.4 *Self-Interest and Morality*

Closely connected with the problem of maxims is the problem to do with Kant’s complete rejection of self-interest in morality. But is this necessary? If self-interest were compatible with moral action, should this count against the moral worth of an agent’s actions? This is a significant point made by Susan Wolf in her article “Moral Saints” and here it is relevant to an assessment of Kant. Among other things, Wolf says: “I believe that moral perfection, in the sense of moral saintliness, does not constitute a model of personal well-being toward which it would be particularly rational or good or desirable

¹⁹⁶ Ibid.

for a human being to strive.”¹⁹⁷ I agree with Wolf that it should not count against an agent if she were partly self-interested in her pursuit of morality—after all, we are not gods or angels but human and this seems like common sense. Perhaps in this case Kant ought to allow that an agent have some admixture of self-interest in her maxim and still qualify as having a morally worthy motive.

If this were allowed, it might be able to mitigate the pernicious state of paralysis of the self-reflective agent above. If she were allowed to be simultaneously moral and self-interested (or just interested in her own health, well-being, or happiness in addition to the good of others) she would not have to worry about her self-interest affecting the moral status of her actions. A certain amount of self-interest would not undermine the moral worth of all of her action. However, to allow this would be to go beyond Kant and his theory of moral agency in the *CPR*, *GR*, and *CPrR*. Perhaps this negative effect will not be an issue for some in their appreciation of Kant’s theory of morally responsible agency, perhaps for others it will; my suspicion is that it will be a problem for more rather than less. If this is the case, the problem of maxims might signify a limit for anyone who wants to understand Kant’s concept of moral responsibility as consonant with common sense. But even if this is the case, it does not take away from the fact that on the whole, Kant’s concept of freedom and moral responsibility is consonant with common sense. More importantly, if his argument in the *GR* and *CPrR* is successful, his theory provides us with a stable, agent-independent system of morality that mediates our social interaction and fixes an objective standard for morally responsible agents to attain. More than that, it allows us to access the tools of social change without ever

¹⁹⁷ Susan Wolf, “Moral Saints” in *The Journal of Philosophy* LXXIX: 8 (August 1982): 419.

compromising the belief that we are agents of infinite dignity and worth, accountable to one another for what we do and who we become.

CHAPTER III

*With the idea of freedom the concept of autonomy is now inseparably combined, and with the concept of autonomy the universal principle of morality, which in idea is the ground of all actions of rational beings, just as the law of nature is the ground of all appearances.*¹⁹⁸

Given the picture of freedom and morality present in the *GR* and *CPrR* that I have outlined in Chapter II, I will argue in Chapter III that Kant holds that an understanding of ourselves as morally responsible agents, accountable to one another for what we do, is possible in virtue of our possessing certain features as intrinsic to our agency. I will make it the goal of Chapter III to elucidate what these conditions are for Kant and what they enable us to do. I will argue that for Kant, two further conditions in addition to freedom (as meta-compatibilism) are present in human beings that enable us to understand ourselves as morally responsible agents. The first of these is internal motivation and the second is objective justification. Following that I will conclude this project by offering some brief suggestions for future work and further lines of inquiry that might be pursued in connection with this project.

3.1 Introducing the Conditions of Accountable Agency

Before proceeding to define these two further conditions of moral responsibility, I will begin by, in this section, summarizing what I have established so far. The unknowability of our actual freedom and responsibility was first discussed in Chapter I,

¹⁹⁸ *GR*, 4:452.

Section 1.2, where I pointed out that according to Kant, the only knowledge that is possible is knowledge of the world as it presents itself in experience (the empirical or phenomenal realm) while the world of things as they exist in themselves and independently from our experience (the transcendent or noumenal realm) is off limits for human knowledge. This idea is stated most directly in the *CPR* when Kant says that as a result of what he sought to demonstrate in the Transcendental Aesthetic, we present “even ourselves to ourselves only as we appear to ourselves, not as we are in ourselves.”¹⁹⁹ But, as I argued in the rest of Chapter I, this works to the advantage of Kant, who wanted to preserve the idea of human freedom from a mechanistic kind of knowledge that would otherwise eliminate it. This is what my interpretive theory of meta-compatibilism argued—that because of this limitation in our cognitive capacities (in other words, our ability to know the world only as it presents itself to us in our experience and not as it is in itself) we are licensed to regard ourselves as possessing free agency.

This is what Kant established in his resolution to the Third Antinomy, where in order to resolve the conflict between a need to posit a free first cause (to avoid an infinite regression of causes) and what he understood as the need for causal determinism to be in place for the unity of consciousness to obtain, he drew the distinction between empirical and intelligible, phenomenal and noumenal, character.²⁰⁰ The net result of this distinction is that we are permitted to view ourselves and our actions as practically free, even though the actual status of our freedom cannot be determined. Interestingly, it is specifically because the status of our freedom cannot be determined that we are permitted to believe that it exists. I indicated in Chapter I, Section 1.2.1.2 that this is an expression of one of

¹⁹⁹ *CPR*, B153.

²⁰⁰ *CPR*, A538-541/B566-569.

the practical aims of the *CPR*, to chasten reason in order to make room for the three items of practical faith—God, freedom, and immortality.²⁰¹ Because of this, I argued that Kant should be understood to be supporting a common sense or practical concept of human freedom—he claims that this is his aim in the Preface to the second edition of the *CPR* and it is, in fact, the net result of his resolution of the Third Antinomy. As I argued, he is successful in preserving this common sense concept of freedom in this practical way.

Given this goal of the *CPR*, his most significant work of theoretical philosophy, I suggest that it is reasonable to conclude that Kant's aim is to preserve the practical concept of moral responsibility in a similar fashion to the way that he preserves the concept of freedom in the *CPR*. For he is not trying to establish a highly specific, fact-based idea of precisely when certain acts are attributable to certain agents as Eshlemen's concept of attributability would aim to establish. Rather, Kant aims to license our belief in moral responsibility more generally; in other words, to permit us to understand ourselves as morally responsible agents that are accountable to one another. In this way, he aims to license the concept of moral responsibility in order to preserve the more general common sense and specifically practical concept of accountability, to use Eshlemen's language. Korsgaard points out that Kant reveals that it is his aim to establish the practical concept of moral responsibility in this way when, in the *CPrR*, he identifies it with the three items of practical faith in the *CPR*.²⁰²

As I will argue shortly, the notion of internal motivation is established in Kant by the principle of autonomy, the mechanism of incorporation, and the idea that there are

²⁰¹ *CPR*, Bxxx.

²⁰² Christine Korsgaard says this in reference to *CPrR*, 5:132-4 (Christine Korsgaard, *Creating the Kingdom of Ends*, 208).

rational incentives for moral action. The principle of autonomy is found in the *GR* and establishes first of all, that human beings are capable of becoming aware of the moral law (an objective standard of morality) in virtue of the most basic rational capacity. The idea that the human being is “subject only to laws given by himself but still universal and that he is bound only to act in conformity with his own will, which, however, in accordance with nature’s end is a will giving universal law”²⁰³ means that human beings may be aware of and therefore accountable to the same moral law in the form of universality.

Before I elucidate the concept of internal motivation, it is helpful to note that in Section III of the *GR* Kant intends to show that the moral law is a real standard that human beings can strive for. He does this by first of all establishing the structure and contents of morality/the moral law in *GR* Sections I and II and then pointing out that if freedom is presupposed, we *must* think of ourselves as moral agents. It was with this in mind that Kant said: “the concept of freedom... constitutes the keystone of the whole structure of a system of pure reason, even of speculative reason.”²⁰⁴ By this he meant that once we have recognized ourselves as free, which we are licensed to do as argued in Chapter I, the belief in our freedom conjoins the two respective domains of speculative and practical reason, thereby granting stability and rational integrity to the entire system of reason as well as human action because it enables us to cognize something of the noumenal realm, namely, the moral law. This is because, first of all, the appeal to freedom resolves the contradiction inherent in the Third Antinomy and secondly, it permits us to have faith that moral deliberation and decision-making is not a vain exercise but may make a difference in our actions and the world.

²⁰³ *GR*, 4:432.

²⁰⁴ *CPrR*, 5:3-4.

In fact, if we regard ourselves as free, it is possible for us to regard ourselves as capable of acting in accordance with morality. So if we do then recognize ourselves as free, we will also recognize ourselves as morally responsible agents who are accountable to one another in virtue of our mutual awareness of the moral law. To refer back to the distinction that Eshlemen makes between moral responsibility in terms of attributability and moral responsibility in terms of accountability, this license to regard ourselves as morally responsible agents does not mean that we will necessarily know when we are responsible or have acted morally in particular cases. (In the case of internal motivation, this would mean following the moral law evident in the principle of autonomy by incorporating it as our maxim of action.) Instead, it means that we regard ourselves as having the capability to do so and therefore regard ourselves as being accountable for what we do, qualifying as morally responsible in the latter sense.

Within this context, certainly not the least important contribution that Kant makes to a theory of moral responsibility is to fix an objective system of morality that can guide our behavior, giving us an ideal to strive toward while providing a morally legitimate (because objective and agent-independent) standard by which we can hold others accountable. This system of morality is not threatened because the fact of our having acted freely or not (as well as our maxims in specific circumstances) cannot be known. For unlike the fact of our action and the real status of our maxims, the system of morality is known a priori; as Kant says, “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.”²⁰⁵ Once an individual chooses to recognize herself as a

²⁰⁵ *GR*, 4:411.

free agent, the entire system of morality bound up with/derivable from the moral law comes into full effect just as though it were a natural law.

With the idea of freedom the concept of autonomy is now inseparably combined, and with the concept of autonomy the universal principle of morality, which in idea is the ground of all actions of rational beings, just as the law of nature is the ground of all appearances.²⁰⁶

The only relevant question that remains is whether the individual desires to regard herself as free and so take on the moral responsibility that comes with that recognition. If Kant's claims can be rigorously defended, then a good case could be made that provides an attractive picture of human action and morality as it holds that we are able to engage in rational, responsible, and ultimately meaningful interaction with one another as beings with dignity and worth.

3.2 Internal Motivation

Given the picture of moral reasoning that Kant presents us with in the *GR* and the *CPrR*, I would now like to give explicit attention to the second of the three necessary and sufficient conditions for moral responsibility outlined in the introduction to this thesis—internal motivation. In what follows I will show how Kant's moral theory in the *GR* and *CPrR* produces a concept of action as internally motivated. By saying that action is internally motivated, I mean that it is capable of being supported by a process of reasoning occurring within and guided by the agent, prior to a particular act that she commits. It will become obvious as I proceed, that the concept of internal motivation is

²⁰⁶ *GR*, 4:452.

inextricably tied to a supporting theory of moral deliberation as a rational procedure. To show that this is the case in Kant's theory, I will take a second look at the principle of autonomy and the moral law to explain the process by which the agent is internally capable of being aware of the moral law, arguing that this supports the concept of internal motivation. In addition, I will also take a brief look at what Henry Allison calls the "Incorporation Thesis" as well as what Kant calls the "moral feeling" in the *CPPrR* and show how these also support the concept of internal motivation in Kant.

3.2.1 Autonomy and the Moral Law

The principle of autonomy is the most obvious starting point for an argument that seeks to establish that Kant holds that a theory of action as internally motivated. This is because it contains the idea that every rational agent is capable of becoming aware of the same moral law as every other agent through the use of her own internal faculties.

Therefore such an agent can also be subject to and held accountable to the same moral law as other agents—a moral law that has the requisite scope and objective contents to mediate and possibly govern the actions of all agents. This is reflected in the idea that "the human being is bound to laws by his duty, but... subject only to laws given by himself but still universal and that he is bound only to act in conformity with his own will, which, however, in accordance with nature's end is a will giving universal law."²⁰⁷

In order to understand how an agent can become aware of the moral law through the use of her internal rational faculties, it is helpful to point out just what those faculties consist

²⁰⁷ *GR*, 4:432.

in. I suggest that the requisite understanding an individual needs in order to become aware of the moral law is an understanding of what it takes to maintain important social practices without undermining them and an understanding of the concept of universality.

The derivation of the moral law in the *CPrR* shows that when we draw up maxims for ourselves, we need only think of them as having universal scope in order to become aware of the moral law as a consistent form of action as well as the only consistent form of action. This demonstrates that proper moral evaluation, that is, a process of evaluation that delivers universalizable content in accord with the moral law, easily occurs in virtue of a capacity internal to the individual agent by processes that she directs. Kant does not explicitly say this but it is implied in Theorem III of the *CPrR* when he says that “[i]f a rational being is to think of his maxims as practical universal laws, he can think of them only as principles that contain the determining ground of the will not by their matter but only by their form.”²⁰⁸ What he is saying explicitly here is that an agent is to think of her maxims as practical universal laws, which she can do only in virtue of their form and never their contents. Or in the language of the *GR*, if she is to think of her maxims as practical universal laws, they must have categorical form and not merely hypothetical form; therefore they must have the form of the CI because any condition (matter) to their application would render them hypothetical imperatives and strip them of their universal form. In order to conceive of one’s maxim as a practical universal law, the agent has to have a grasp of the concept of universality. Once the agent has an understanding of the concept of universality, she can conceive of her actions as having universal scope; in other words, she can conceive of her actions as though they were universal laws of nature

²⁰⁸ *CPrR*, 5:27.

that herself as well as all other agents must act in accordance with. This is what Kant has in mind when he says in the *GR* that “the universal imperative of duty [the CI] can also go as follows: act as if the maxim of your action were to become by your will a *universal law of nature*.”²⁰⁹ Once the agent has done this, she is confronted by the fact that her actions, if based on conditions of self-love, for example, cannot attain universal form because they cannot conceivably sustain important social practices and operate as so-called universal laws of nature.

Consider the example that Kant uses in the *GR* of the man who borrows money knowing that he cannot repay it. If he universalizes the maxim: “when 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,” which is a principle of self-love because it serves his needs alone—to the exclusion of the lender. He realizes that this action would result in contradiction because it would undermine the very practice of promise-keeping on which it depends. For if other individuals were to frequently borrow money without repaying it, no one would be willing to lend money to others and no lender would want to fulfill this particular man’s request to borrow money. From this example, it is evident that an agent needs to think of her actions as having universal scope. Once she does this, the moral law as the only form of action consistent with the form of universality follows from the idea of her actions as having universal scope in combination with the principle of non-contradiction. Once she universalizes certain actions (as the man borrowing money in this example), she will see that action on the basis of some maxims undermines certain important social practices while action on the basis of other maxims does not; for

²⁰⁹ *GR*, 4:421.

example, if the man borrowing money were to promise to repay the money knowing that he could repay the money and intending to do so, acting on the basis of this maxim would be permissible because this maxim can be universalized while upholding the trust necessary to sustain the practice of borrowing money. This is the sort of process that Kant has in mind when he says that “[i]t is 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.”²¹⁰ This process of reasoning is readily available to every individual rational agent in virtue of possessing the most basic rational faculties and understanding and it shows that the contents of morality (the form of universality derived from the CI) are available internally to all such individuals.

The upshot of all this is summarized in the principle of autonomy, which states that “the human being is bound to laws by his duty, but... subject only to laws given by himself but still universal and that he is bound only to act in conformity with his own will, which, however, in accordance with nature’s end is a will giving universal law.”²¹¹ The principle of autonomy then provides the basis for action that is internally motivated and has moral value, because every agent who can become aware of the moral law can also strive to make her actions conform to it.

Kant holds that this internally motivated process of deliberation can be understood as making a difference in the world because, as I have argued in Chapter I, the resolution of the Third Antinomy allows us to regard our choices as making a

²¹⁰ *CPrR*, 5:29.

²¹¹ *GR*, 4:432.

difference in the world, on the grounds of faith. Henry Allison also appears to hold that our choices can make a difference in the world in his idea of the “Incorporation Thesis,” by which incentives for action are incorporated into maxims by the individual agent. Allison claims it is Kant’s most important published conception of practical spontaneity and Allison borrows a phrase from Kant’s *Religion Within the Limits of Reason Alone* (REL) to describe it: “freedom of the will is of a wholly unique nature in that an incentive can determine the will to an action only insofar as the individual has incorporated it into his maxim.”²¹² In Chapter II Section 2.1 on the idea of the moral law, we saw that Kant holds maxims to be the subjective principles of volition that particular agents choose to act upon. He contrasts maxims with the practical law (which he develops into the concept of the CI later in the GR) that would serve as the only fundamental maxim for the ideally rational agent—the agent whose maxims would accord perfectly with reason.²¹³ So although every agent is aware of the moral law within herself as the form her action would take if her action were to be completely rational, the fact that the moral law must first be incorporated by the agent as a reason for action indicates that she is free to act in accordance with the moral law or not. She is not physically or otherwise compelled (forced) to act one way or another but possesses the ability to incorporate a variety of incentives as possible reasons for action. Because the individual agent possesses this capacity she thereby possesses the power of choice or the ability to act spontaneously (of her own accord) on the basis of reasons that she selects for herself. In this way, the

²¹² Immanuel Kant, *Religion Within the Limits of Reason Alone*, 6:24 as quoted by Henry Allison (Henry Allison, *Kant’s Theory of Freedom*, 40).

²¹³ GR, 4:402.

actions of the individual agent can be internally motivated (because the agent possesses this capacity within herself). So Henry Allison argues.

This basic apparatus of action that occurs by the incorporation of incentives as maxims corresponds well with Kant's concept of the will as the capacity of the individual to act in accordance with the representation of laws.²¹⁴ Where the will defines the general capacity of the human being for a certain kind of action, the act of incorporation is that action itself. Prior to the act of incorporation, the agent can be understood to have a variety of incentives for action and one of these must be taken up by the agent as a reason for action (maxim) prior to her acting. These maxims each imply a unique formula or law for action because when we act we have the ability to conceive of our action as having a place in a system of rules and different maxims will imply different systems of rules.²¹⁵ It could be said somewhat differently that reasons for action are ultimately determined according to general principles that the agent perceives to offer good reason for acting a certain way and the moral value of an action is determined by the more general principle embedded in the particular maxim.²¹⁶ The structure through which moral value is conferred on an action therefore moves from the general concept of a rule or imperative to the specific maxim and it is in the act of incorporation that the individual agent takes up an incentive as *her* maxim or reason for action at a particular time in a particular circumstance. Because it is a precondition for the uptake of any moral reason by the agent, the act of incorporation could be said to be the most basic and necessary feature of

²¹⁴ *GR*, 4:412.

²¹⁵ This is implied by Christine Korsgaard when she says that our reasons for action are cast as maxims and maxims can be derived from imperatives (Christine Korsgaard, *Creating the Kingdom of Ends*, 163-4).

²¹⁶ Henry Allison says something like this (Henry Allison, *Kant's Theory of Freedom*, 93).

human (moral) agency—it is where the rubber of moral rules and reasoning hits the road of interpersonal moral life. Moreover, it is a feature possessed by all rational human agents, granting them the basic capacity for internally motivated action. But it only clarifies the structure of deliberation; it does not in essence go beyond what Kant already hopes to have established in his meta-compatibilism of the *CPR*.

3.2.2 *Internal and Rational Incentives*

So far, if both Kant and myself are correct, the principle of autonomy shows that the individual agent can be aware of the moral law while the act of incorporation shows that there is an internal apparatus of action that grants the individual agent the ability to employ incentives as reasons for action and the framework of meta-compatibilism enables the agent to regard her actions as free. It would seem that in order to grant complete integrity to action as being internally motivated, it is also necessary to show that the incentives selected to guide action can be evaluated and selected by a rational process that also occurs internal to the agent herself. I will argue that Kant's discussion of incentives and in particular, the moral incentive that he calls the "moral feeling" in the *CPrR*, confirm the rational nature of moral deliberation in Kant's moral theory, which I have already argued is grounded in the moral law.

In the definition of the act of incorporation from *REL*, Kant says that the act of incorporation happens when an incentive is incorporated by an agent as that agent's maxim of (or reason for) action. At this point he uses the term 'incentive' quite generally so that incentives can be moral or immoral, rational or empirical. However, when he uses

the term in Chapter III of the *CPrR*, “On the Incentives of Pure Practical Reason” he only ever discusses the incentive of the moral law (i.e. the moral, rational incentive) though he defines an incentive more generally as the “subjective determining ground of a being whose reason does not by its nature necessarily conform with the objective law.”²¹⁷ On the basis of this definition in the *CPrR* in addition to his using the more general term elsewhere (in the *GR* and *REL*) to apply to both moral and immoral incentives that can be taken up by the individual in the act of incorporation, it is reasonable to conclude that Kant generally means that incentives can be moral *or* immoral. However, it is still somewhat puzzling that he doesn’t discuss immoral incentives in detail in his section on incentives in the *CPrR*. I suggest that a detailed discussion of immoral incentives is lacking because first, it is not controversial to say that there exist incentives for an agent to disobey the moral law and thereby lapse into heteronomy. For example, it would seem easier to “incorporate” the incentive of physical pleasure than virtually anything else, while describing/defending the existence of any moral incentive would appear more challenging.

Second, heteronomy would appear the default position of human agents insofar as heteronomy precisely is a *lapse* into a form of natural causality because it is a relinquishment of rational control to natural inclinations. Allison points out that maxims based on inclination (or you could say, heteronomous maxims) are explicable in terms of laws of nature and provide the sole source of motivation although the agent could have chosen otherwise.²¹⁸ This lapse on the part of the agent is unfree in the sense that the agent is allowing laws of nature to drive her behavior, but free in the sense that the agent

²¹⁷ *CPrR*, 5:72.

²¹⁸ Henry Allison, *Kant’s Theory of Freedom*, 97.

is still *allowing* this because she is choosing not to exercise her rational and volitional faculties in order to make her actions accord with the moral law. This is not a new point in the course of what I have said so far as it corresponds with Kant's complex notions of negative and positive freedom already discussed in Chapter II, Section 2.2.1, where negative freedom is freedom *from* being *necessarily* determined by natural causes, while positive freedom is that plus being free *for* action in accordance with the rational moral law.²¹⁹ The point of bringing it up again is to help explain the absence of immoral incentives in Kant's definition of incentives.

What all this means is that it is not really important for Kant (or us at this point) to think about heteronomous "incentives" in the *CPrR* because we already know that the act of incorporation makes all action free (including heteronomous action, which, as discussed above, is simultaneously unfree according to Kant). So instead of discussing immoral incentives in detail, we can assume that Kant intends to subsume them under the two categories under which he says all natural inclinations fall—self-love (love for oneself) and self-conceit (pride or satisfaction with oneself).²²⁰ However, if there is to be such a thing as free and moral action or free action in accordance with the moral law, there needs to be some reason or additional rational component internal to the moral agent that enables her to follow the moral law. As Kant says, it is "essential in every determination of the will by the moral law that, as a free-will... it is determined solely by the law"; therefore we have to show what the moral law effects in the mind in order to answer how the moral law can be a sole and immediate determining ground.²²¹ I argue

²¹⁹ See also *GR*, 4:446

²²⁰ *CPrR*, 5:74.

²²¹ *CPrR*, 5:72.

that what Kant presents as the incentive to obey the moral law in the *CPrR* reinforces the notion that moral action is motivated by a source internal to the agent.

In an essay on the moral incentive, Stephen Engstrom highlights the internal and rational nature of the moral incentive by looking afresh at the original language Kant used. He points out that “[i]ncentives are carrots and sticks, not the desires and aversions they may trigger. But when Kant speaks of a *Triebfeder* [the original German word that he used, later to be translated as ‘incentive’], he almost always has in mind something in the subject that generates the action, rather than an object or circumstance that prompts it.”²²² Engstrom also says that in its “original literal meaning, *Triebfeder* refers to the mainspring of a clock” and that “we should [therefore] think of a *Triebfeder* as an inner spring or source of choice and action.”²²³ What this inner spring or source of action is, is what Kant calls the “moral feeling.”²²⁴ He uses the term feeling here because the heteronomous incentives of the agent are inclinations, which are based on feeling, and awareness of the moral law thwarts these inclinations, making the agent aware of the immoral status of her inclinations and producing a negative effect on the agent in the form of a feeling of pain and humiliation.²²⁵ But the moral law also has a positive effect on the agent that Kant calls respect. Essentially, by being impressed by the pure universality of the scope of the moral law, the agent is inspired to respect the moral law

²²² Stephen Engstrom, “The *Triebfeder* of Pure Practical Reason” in *Kant’s Critique of Practical Reason: A Critical Guide* eds. Andrews Reath and Jens Timmermann (New York: Cambridge University Press, 2010), 92.

²²³ *Ibid.*

²²⁴ *CPrR*, 5:75.

²²⁵ *CPrR*, 5:73.

and desire to remove anything that would hinder its being realized in the world.²²⁶

Together these negative and positive effects form the moral feeling.

To some this may sound as though Kant is sneaking some form of emotivism in the back door and calling it rational. Admittedly, it is not in all ways transparent as to what Kant is getting at here. He takes up the idea of incentives once again in the final chapter of the *CPrR* on “The Doctrine of the Method of Pure Practical Reason” and here seems to be contrasting properly moral incentives (as those that compose the moral feeling) with mere feelings. He says:

All feelings, especially those that are to produce unusual exertions, must accomplish their effect at the moment they are at their height and before they calm down; otherwise they accomplish nothing.... Principles must be built on concepts; on any other foundation there can only be seizures, which can give a person no moral worth and not even confidence in himself, without which the consciousness of one’s moral disposition and of a character of this kind, the highest good in human beings, cannot come to exist.²²⁷

Moral incentives and the moral feeling are not to be identified with this sort of mere feeling because they rest on a rational principle whereas these feelings do not and so accomplish nothing.

Furthermore, prior to saying the above Kant takes care to distinguish properly moral or pure incentives from impure incentives that contain an interest for one’s own happiness and welfare.

All the admiration, and even the endeavor to resemble this [truly moral] character, here rests wholly on the purity of the moral principle, which can be clearly represented only if one removes from the incentive to action everything that people may reckon only to happiness. Thus morality must have more power over the human heart the more purely it is presented. From this it follows that if the law of morals and the image of holiness

²²⁶ *CPrR*, 5:75.

²²⁷ *CPrR*, 5:157.

and virtue are to exercise any influence at all on our soul, they can do so only insofar as they are laid to heart in their purity as incentives, unmixed with any view to one's welfare.... Consequently every admixture of incentives taken from one's own happiness is a hindrance to providing the moral law with influence on the human heart.²²⁸

This account of incentives is rather puzzling because in it Kant seems to want to invoke moral emotion though he has already excluded emotion in virtue of the strictly rational nature of the moral law. Nevertheless, Kant clearly intends that the moral law (law of morals, and the image of holiness and virtue) is itself to be the incentive and exclusively so.

So the moral feeling and moral incentives remain deeply rational because their source is the moral law itself, which is based on a rational principle (the CI) and which the agent becomes aware of by exercising the most basic of her rational faculties. In this way, the moral law is to be prior to this effect that it has on the agent (the moral feeling), which is crucial because it places the rational moral law at the base of the moral feeling or incentive. In showing that not only awareness of the moral law but the rationally based incentive associated with it is internal to the individual agent, Kant's account of the incentive to obey the moral law does reinforce the argument that moral action is internally motivated.

3.2.3 Some Objections to Kant's Account of Internal Motivation

I will now briefly consider some possible objections to the idea of internal motivation that I have argued is inherent in Kant's moral theory. I will treat successively,

²²⁸ *CPrR*, 5:156.

what I anticipate to be the strongest objections to each principle of internal motivation laid out above.

The first principle that I discuss in connection with internal motivation—the principle of autonomy—appears to be claiming that an agent is exercising freedom/autonomy only insofar as she truly acts in accordance with the moral law (accessible to her cognitive faculties), that is, insofar as her actions have the form of universality. First of all, the objection could be raised that this seems to imply that if an agent's action does not bear the form of universality inherent in the moral law, she is not autonomous or, in other words, not free, and therefore could not be responsible for her actions. If true, this would have the double consequence of rendering Kant's account of autonomy counter-intuitive and undermining the agent's responsibility for her actions. This is because if not in accord with the moral law, her actions would be unfree and I have held that for an agent to be responsible for her actions she must have freely committed them according to at least a compatibilist definition of freedom. This objection can be effectively answered by invoking Kant's distinction between negative and positive freedom. As I mentioned earlier in Chapter II, Section 2.2.1, Kant held to a composite idea of freedom with a negative and positive dimension. There I aimed to show how for Kant freedom as choice is present in acts that are not autonomous (according to his strict definition of the word) and therefore make the agent count as responsible. This means that Kant's concept of internal motivation is not vulnerable to objection on the above grounds.

The second principle that I discuss in connection with internal motivation is the moral feeling as a rational incentive to act on the basis of the moral law. As I have

mentioned, it is possible that some readers might have the suspicion that the idea of a moral feeling as an incentive to act morally does not quite fit with Kant's rational account of moral deliberation and motivation. Though Kant is not totally clear on this issue, I would like to explore this suspicion in more detail with the purpose of defending it from the claim that because of the emotional quality of the moral feeling, Kant's theory of moral motivation lapses into some form of emotivism rather than remaining a consistently rational theory. This is important to do because if it could be shown to lapse into a form of emotivism, even in the limited case of the moral feeling, Kant would forfeit the possibility of action being internally motivated because autonomous (as opposed to heteronomous) action in accordance with the moral law is necessary to establish the idea of internal motivation.

The *Cambridge Dictionary of Philosophy* defines the emotivist as a “noncognitivist who claims that moral judgments, in their primary sense, express the appraiser's attitudes—approval or disapproval—toward the object of evaluation, rather than make assertions about the properties of that object [or action].”²²⁹ By this definition, it is clear that Kant's idea of the moral feeling does not merit the name ‘emotivism.’ In the case of emotivism, the nonrational (or noncognitive, emotional) features of the appraiser are exclusively what do the work in evaluating (or justifying, if you are looking at it from the agent's perspective) the moral status of actions. This is because for the emotivist, although some actions may be superficially rational or appear to have rational justification on a superficial level, that rationality is not possible but emotive meaning is primary instead. In Kant's case, it is the moral law alone that does the work in evaluating

²²⁹ David Brink, “Emotivism” in *Cambridge Dictionary of Philosophy* 2nd edition, ed. Robert Audi (Cambridge: Cambridge University Press, 1999), 260.

the moral status of actions and the moral law is a practical faculty whose rationality can be known to the individual (both agent and appraiser alike) in virtue of their possessing the concepts of universality and non-contradiction. The combination of the emotional features of pain and respect that together form the moral feeling may be the means by which the individual first becomes aware of the moral law but this is not necessarily the case. This is because the individual could also become aware of the moral law by exercising her basic rational faculties without first experiencing the moral feeling; both rationality itself as well as the moral feeling, which is fundamentally a rational phenomena, can be motivating for Kant. Regardless of whether or not the moral feeling is how the individual first becomes aware of the moral law, the moral law can be reasoned out with the basic concepts of universality and non-contradiction, so that it can be justified by the rational/cognitive concepts and faculties of the individual, rather than her emotive/non-cognitive features. Furthermore, it is the rational/cognitive content of the moral law that fundamentally and exclusively evaluates the moral status of actions and the emotive components of the moral feeling play no role in this process. For these reasons, it is clear that Kant's moral theory, though not completely clear on this point, is not vulnerable to a conflation with emotivism and should be understood as a consistently and coherently rational theory. Kant's use of the moral feeling can be understood as an attempt to come to terms with how the emotional side of human nature assumes its proper place in the rational process of moral deliberation and action.

3.3 Objective Justification

Now that I have argued that the internal capacity for individual action can be established in Kant's theory, it is important that the process of moral reasoning can also be shown capable of delivering an objective form of justification for action. By saying that a process of moral reasoning can deliver objective justification, I mean that it can show a standard of conduct to exist consistently for all individual agents regardless of setting or circumstance.²³⁰ Such a theory can be expected to deliver a standard of conduct that is universally applicable despite the varying circumstances of particular agents. As stated in the Introduction, I hold such an objective standard of justification to be the final of three necessary and sufficient conditions that must be in place for a theory of moral responsibility to obtain. It could be objected that such a condition is not necessary for a theory of moral responsibility and I deal with this objection briefly in Section 3.3.3 below. In the rest of this section prior to that, I will argue that such an objective standard of justification is attributable to Kant.

²³⁰ This is a variation of what the *Cambridge Dictionary of Philosophy* defines as ethical objectivism, "the view that the objects of the most basic concepts of ethics (which may be supposed to be values, obligations, duties, oughts, rights, or what not) exist, or that facts about them hold, objectively and that similarly worded ethical statements by different persons make the same factual claims (and thus do not merely concern the speaker's feelings)," (T.L.S. Sprigge, "Ethical Objectivism," *Cambridge Dictionary of Philosophy*, 284).

3.3.1 *The Form of Universality*

The moral law is undoubtedly the first place to look for establishing that objective justification is a feature of Kant's moral theory because it presents to the agent nothing less than the form of universality as the standard of right moral conduct, in commanding the agent to "[s]o act that the maxim of your will could always hold at the same time as a principle in a giving of universal law."²³¹ In commanding this, the moral law commands that *all* agents act so that their maxim or reason for action (the two terms can be used interchangeably) has universal form. Reath states this interestingly when he refers to Kant's idea in the *GR* that rational nature consists in the ability to act according to laws (*GR*, 4:412) and says that "at the core of Kant's conception of rational agency is the idea that rational action is guided by considerations that the agent takes to provide justifications for acting in a certain way."²³² Rational agency is therefore the capacity to guide one's actions by normative standards that are generally applicable.²³³ It helps to point out here that the content of the moral law is available to all rational agents so that all rational agency bears the same capacity to act morally. In virtue of being rational, an agent can also be moral or at least strive to be moral by choosing a maxim as a reason for action whose justification can be understood to have force for all other rational beings. Reath suggests that the point to be grasped here is that when properly applied, the moral law "translates facts about one's situation and one's ends and desires into conclusions about how to act whose force can be understood by anyone"—that is, by anyone who is a

²³¹ *CPrR*, 5:30.

²³² Andrews Reath, *Agency and Autonomy in Kant's Moral Theory* (Oxford: Oxford University Press, 2006), 76.

²³³ *Ibid.*

rational agent.²³⁴ This is correct because this is something that the moral law does. However, it is too weak because Kant believes that the moral law generates reasons for action that can not only be understood but also accepted by everyone. This is what it means for one's action to have the form of universality—for the justification of their action will have force for any other agent in that situation and in that way, will be agent-independent and objective.

Furthermore, as I argued previously in Section 3.2 on internal motivation, the moral law is made known to the agent by way of the most basic act of deliberation—by universalizing a possible maxim or reason for action. It is safe to assume that the requisite faculties for such a process of reasoning are available to any properly functioning agent (to the exclusion of those suffering from forms of cognitive malfunction). I suggest that this capacity to reason is in part constitutive of agency as most would understand it, as self-directed action would require at least such a minimum reflective capacity. So insofar as anyone is an agent, they could be understood to possess these basic rational capacities and because all agents would possess these capacities, the moral law that follows from them can be known independently of the particularities or idiosyncrasies of any individual agent. Moreover, if the moral law is possessed by an agent (and Kant certainly believes it to be possessed by most agents), it could not generate justification for action that is anything other than objective. It could not do this because it is known to the agent and consists precisely in virtue of a rule/maxim that can be universalized without contradiction and it would be a contradiction in terms if these principles generated

²³⁴ *Ibid.*, 77.

maxims as reasons for action that made agent-dependent (particular) rather than agent-independent (universal) statements.

It is possible that an agent could mistakenly believe a maxim to be universalizable when it is not. However, this does not jeopardize the moral law's ability to consistently generate an objective standard of justification. Rather, it reveals that human beings are not perfect in their perception of the moral law—something uncontroversial that can be accepted without threatening the integrity of the moral law itself.

3.3.2 *Externally Evaluable Maxims*

An important outcome of the form of universality inherent in the moral law is that any maxim that an agent employs as a reason for action can be externally evaluated by any other rational agent. This is also implied in the idea that rational action is guided by reasons or justification that the agent takes to have force for all other rational agents. Insofar as an agent's actions are moral (i.e. accord with the moral law), they will also be justified in a way that is rational and therefore have force with other rational agents because all other rational agents will possess an awareness of the moral law in virtue of their own rationality. By contrast, all maxims or reasons for action based on principles of heteronomy or self-love will only have necessary force for the particular agent who selects them and thereby preclude the possibility of an action based on them attaining universal form. As Kant argues in Theorem I of the *CPrR*, all principles based on pleasure or otherwise material/sensible conditions cannot attain the form of necessity (one could just as well substitute universality here for necessity) that is required of a

principle to accord with the universal moral law. This is because all such material or pleasure-based principles presuppose pleasure as a condition of the possibility of choice, while it cannot be supposed that all agents could or would receive pleasure as a result of that action.²³⁵ In fact, many agents would not receive pleasure from that action and some agents may even suffer pain or displeasure from that action in order for the agent carrying out that action to receive pleasure from it. Moreover, what gives a particular agent pleasure could be anything and there is no guarantee that it would give other agents pleasure as well. Kant subsumes all such material principles under the principle of self-love because they are not capable of serving any purpose other than the happiness of the agent employing them—they make that agent's happiness the ultimate reason for action rather than the moral law and consequently rather than reasons that could be embraced by other agents and lead to the happiness of those other agents as well.²³⁶ In this way, a justification for action based on pleasure received by one individual and not others would not have force with other rational agents and therefore could not be universalized or bear the form of universality. Such a justification could be rejected as not a good or moral justification for action.

This gives rise to the idea of justifying actions on the basis of their universalizability; the actions that carry force for all rational agents will be deemed morally acceptable while those that cannot be universalized will be rejected as idiosyncratic and heteronymous, carrying force only for the individual agent and no one else. The more readily a reason for action approximates the form of universalizability, the more likely it will be seen to have positive moral status upon reflection whereas the less

²³⁵ *CPrR*, 5:21.

²³⁶ *CPrR*, 5:22.

readily a reason for action approximates the form of universalizability, the less likely it will be seen to have positive moral status. This ensures that insofar as justification can be persuasive for other agents, it can be objectively determined as moral or not because it cannot be persuasive if it does not bear the form of universality in accordance with the moral law. Insofar as justification cannot be persuasive for other agents, it will be rejected for not having sufficiently moral value. Kant can be understood to be arguing for the goodness of the moral life (defined as action in accordance with the moral law) by demonstrating that only it can bear the form of universality and elevate the individual's reasons for action to the status of a universal law. On the contrary, the immoral life or life defined by self-love cannot attain this sort of greatness in the form of having reasons for action that accord with universal law.

3.3.3 Some Objections to Objective Justification in Kant

In response to the idea of objective justification outlined above, I anticipate two significant objections. The first is directed to myself and the second is directed to Kant. In what follows, I will explain these objections and offer a brief response to them, showing how objective justification as it is defined and put to use above, may be defended from them.

The first objection is to my use of objective justification as a necessary condition for moral responsibility in the first place. It could be argued that we do not need an objective or agent-independent standard of morality in order to be responsible for actions but can be regarded as responsible or accountable agents without any notion of moral

objectivity. To explore this claim in detail goes beyond the scope of this project as it would require a sustained treatment of the notion of objectivity, relativism, and meaning in morality in order to get a firm handle on it. In the meantime, it is instructive to first reflect on something that Andrews Reath says about legislating the moral law: when a rational agent legislates (or acts) according to the moral law, the fact that she is legislating the same moral law as other rational agents means that she and other agents can hold one another accountable to a standard that is not externally imposed but naturally emerges from the nature of each agent.²³⁷ This highlights how the moral law presents a standard against which potentially oppressive or otherwise arbitrary external legislation can be identified and evaluated as such. This is essentially what any objective, agent-independent standard of justification is able to do and I argue that this is a good thing. If we are to hold one another accountable, we ought to be able to understand each other's actions and to do this requires an objective standard. For this reason, I maintain that objective justification is a necessary condition for moral responsibility.

The second objection is to Kant's definition of objective justification and challenges the idea that Kant's arguments for objectivity in the form of the moral law can generate real and objective moral content. In response to this, I point to Section II of the *GR* where Kant gives four examples of real objective moral content that follows from the moral law/CI. The first example purports to show that the CI forbids suicide in a particular case (*GR*, 4:422), the second, that one ought not to borrow money without being capable of and intending to repay it (*GR*, 4:422), the third, that one ought not devote one's life to idleness and enjoyment without undertaking to develop one's natural

²³⁷ Andrews Reath, *Agency and Autonomy in Kant's Moral Theory*, 104.

talent (*GR*, 4:423), and the fourth, that one ought to give to those in need (*GR*, 4:423). The second of these examples, which involves borrowing money without repaying it, I discussed earlier in Chapter III, Section 3.2 on internal motivation and as an example of the moral law generating positive moral content I believe it is sound. The other three examples from the *GR* strike me as similarly sound and while some of the details of the conclusions they draw could be contested, the moral law provides a useful objective backdrop against which to do so. Therefore, I argue that Kant's arguments for objectivity in the moral law may be defended from this objection.

3.4 Conclusion

In Chapter III, I have focused on and developed two features of moral reasoning that I believe to be attributable to Kant and necessary for a coherent theory of moral responsibility—the first feature being internal motivation and the second being objective justification. This is in addition to what I argued in Chapter I—that at least a compatibilist kind of freedom is necessary for moral responsibility and attributable to Kant. At this point, I hope to have convinced the reader that these three conditions are necessary and that it is intuitively plausible to think they are sufficient for moral responsibility: free choice, internal motivation, and objective justification. I hope also to have shown that they are in fact present in Kant's philosophy.

I would now like to conclude by offering several suggestions for future work in connection with this project. What appears to be the most fascinating line of inquiry lies in the connection between empirical and intelligible character and the idea of a

diminished capacity for self-legislation. In Chapter II, Section 2.3.2, I spoke about the possibility of a dual-aspect approach to agency and responsibility by which it is possible to view an agent as both causally/psychologically determined and free. It would be interesting to develop criteria for the purpose of establishing when it is justified to hold an agent accountable and when it is not justified to hold an agent accountable and therefore justified to seek to guide her behavior employing primarily consequentialist means. This could provide an opportunity to delve further into the topic of character development and the acquisition of autonomy in pursuit of such questions as what constitutes a sufficient capacity to self-legislate and whether there are conditions/situations that can effectively override an otherwise mature and rational agent's capacity to self-legislate. Or one could ask how Kant's account of character and autonomy might interact with empirically oriented ethical theories like those of Owen Flanagan and John Doris.²³⁸ I believe these are fascinating questions and my intuition is that a Kantian theory of character, character development, and autonomy has something to contribute to discussions surrounding these questions.

Also interesting is the connection between self-legislation and legal objectivity, in particular, the extent to which positive law must be comprehensible to practical reason and thereby accord with an objective standard of justification. The seeds for further work in this area are contained throughout Chapters II and III of this project but are especially evident in Chapter II, Section 2.1 where autonomy is discussed in detail as well as in Chapter III, Section 3.3, where the notion of objective justification is discussed. On the

²³⁸ See Owen Flanagan, *Varieties of Moral Personality: Ethics and Psychological Realism* (Boston: Harvard University Press, 1991) and John Doris, *Lack of Character* (Cambridge: Cambridge University Press, 2002).

topic of objectivity, it would also be fascinating to explore the correspondence of the moral law with ethical theories from non-Western philosophical traditions like, such as, the Confucian tradition in China.²³⁹ It would be fruitful to evaluate the usefulness of the moral law in contributing to questions of justice that have an international bearing and to see if in such practical matters, the moral law can live up to the objectivity and universality it purports to attain.

Finally, it would be interesting to further explore and define the connection between rationality as the ability to self-legislate in accordance with universal ends and the idea of the inherent dignity of the rational individual. In Chapter II, Section 2.1, I offered a limited assessment of why and how Kant makes this connection but further investigation could be conducted and perhaps a sharper assessment attained. An application could also potentially be made to how we think of the intelligence and value of lower forms of animal life in contrast to the highly developed form of animal life that we see in (and experience as) human beings. It is my hope that further avenues of related work could be opened up by this project; here I have only sought to identify several that appear most promising in my estimation.

Regarding Kant's concept of moral responsibility that I have explored in this project, I hope to have effectively argued that it is fruitful and compelling. I say it is fruitful because Kant's idea of the dignity of human beings and the accountability we have to one another resonates with what I believe is the intuition of many of us that human beings indeed are to be treated this way—as beings of great dignity and worth,

²³⁹ For a treatment of concepts of autonomy, character formation, and normativity in Confucian ethics using the language of contemporary Western (Analytic) philosophy, see Kwong-loi Shun and David Wong, eds., *Confucian Ethics*, (Cambridge: Cambridge University Press, 2004).

capable of doing good and beautiful things for one another. I say it is compelling because it is not only rationally coherent, but robust and versatile because it is able to come to terms with causal and psychological determinism while preserving the common sense idea of a human being as a free, rational, and accountable moral agent.

WORKS CITED

- Allison, Henry. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press, 1990.
- Aquinas, Thomas. "Summa Theologica, First Part of the Second Part, Question VI" in *Free Will*. Edited by Derk Pereboom. Indianapolis: Hackett Publishing Company, 2009.
- Aristotle. "Nicomachean Ethics, Book III, Chapters 1-5" in "Moral Responsibility" by Andrew Eschlemen. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/moral-responsibility/>.
- Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason*. Chicago: University of Chicago Press, 1960.
- Brink, David. "Emotivism" in *Cambridge Dictionary of Philosophy* 2nd Edition. Edited by Robert Audi. Cambridge: Cambridge University Press, 1999.
- Caygill, Howard. *A Kant Dictionary*. Malden: Blackwell Publishing, 1995.
- Dennett, Daniel. *Freedom Evolves*. New York: Viking Penguin, 2003.
- Doris, John. *Lack of Character*. Cambridge: Cambridge University Press, 2002.
- Engstrom, Stephen. *The Form of Practical Knowledge*. Cambridge, MA: Harvard University Press, 2009.
- Engstrom, Stephen. "The *Triebfeder* of Pure Practical Reason" in *Kant's Critique of Practical Reason: A Critical Guide*. Edited by Andrews Reath and Jens Timmermann. New York: Cambridge University Press, 2010.
- Eshlemen, Andrew. "Moral Responsibility" in *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/moral-responsibility/>.
- Flanagan, Owen. *The Science of Mind*. Cambridge: The MIT Press, 1991.
- Flanagan, Owen. *Varieties of Moral Personality: Ethics and Psychological Realism*. Cambridge, MA: Harvard University Press, 1991.
- Irwin, Terence. "Morality and Personality" in *Self and Nature in Kant's Philosophy*. Edited by Allan Wood. Ithaca: Cornell University Press, 1984.
- Kant, Immanuel. *Critique of Practical Reason*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1997.

- Kant, Immanuel. *Critique of Pure Reason*. Translated and edited by Paul Guyer and Allan Wood. Cambridge: Cambridge University Press, 2008.
- Kant, Immanuel. *Groundwork of the Metaphysics of Morals* Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1997.
- Kant, Immanuel. *Grundlegung zur Metaphysik der Sitten*. Leipzig: Felix Meiner Verlag, 1925.
- Kant, Immanuel. *Kritik der Reinen Vernunft*. Hamburg: Felix Meiner Verlag, 1993.
- Kapitan, Thomas. "Free Will Problem" in *Cambridge Dictionary of Philosophy* 2nd Edition. Edited by Robert Audi. Cambridge: Cambridge University Press, 1999.
- Korsgaard, Christine. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.
- Korsgaard, Christine. "Introduction to the *Groundwork of the Metaphysics of Morals*" in *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press, 1997.
- O'Neill, Onora. *Acting on Principle: An Essay on Kantian Ethics*. New York: Columbia University Press, 1975.
- Pereboom, Derk. "Kant on Transcendental Freedom" in *Philosophy and Phenomenological Research*. LXXIII:3 (2006): 537-567.
- Reath, Andrews. *Agency and Autonomy in Kant's Moral Theory*. Oxford: Oxford University Press, 2006.
- Shun, Kwong-loi and David Wong, eds. *Confucian Ethics*. Cambridge: Cambridge University Press, 2004.
- Sprigge, T.L.S. "Ethical Objectivism" in *Cambridge Dictionary of Philosophy* 2nd Edition. Edited by Robert Audi. Cambridge: Cambridge University Press, 1999.
- Wolf, Susan. "Moral Saints" in *The Journal of Philosophy* LXXIX: 8 (1982): 419-439.
- Wood, Allan. "Kant's Compatibilism" in *Self and Nature in Kant's Philosophy*. Edited by Allan Wood. Ithaca: Cornell University Press, 1984.