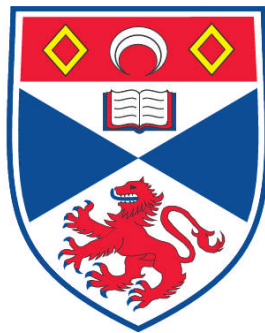


AUTONOMY AND PURITY IN KANT'S MORAL THEORY

Carolyn Jane Benson

**A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews**



2009

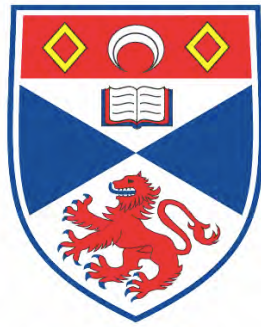
**Full metadata for this item is available in the St Andrews
Digital Research Repository
at:
<https://research-repository.st-andrews.ac.uk/>**

**Please use this identifier to cite or link to this item:
<http://hdl.handle.net/10023/937>**

This item is protected by original copyright

AUTONOMY AND PURITY IN KANT'S MORAL THEORY

CAROLYN JANE BENSON



A THESIS SUBMITTED FOR THE DEGREE OF PH.D.
AT THE
UNIVERSITY OF ST ANDREWS

NOVEMBER 18, 2009.

I, Carolyn Jane Benson, hereby certify that this thesis, which is approximately sixty-nine thousand words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I, Carolyn Jane Benson, was admitted as a candidate for the degree of Doctor of Philosophy in September 2005; the higher study for which this is a record was carried out in the University of St Andrews between 2005 and 2009.

Date

Signature of candidate

I, Dr. Jens Timmermann, hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date

Signature of supervisor

In submitting this thesis to the University of St Andrews, I, Carolyn Jane Benson, understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration.

Date

Signature of candidate

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to Printed copy and electronic publication of thesis through the University of St Andrews.

Date

Signature of candidate

Signature of supervisor

For Kate

ABSTRACT

Kant believed that the moral law is a law that the rational will legislates. This thesis examines this claim and its broader implications for Kant's moral theory.

Many are drawn to Kantian ethics because of its emphasis on the dignity and legislative authority of the rational being. The attractiveness of this emphasis on the special standing and capacities of the self grounds a recent tendency to interpret Kantian autonomy as a doctrine according to which individual agents *create* binding moral norms. Where this line is taken, however, its advocates face deep questions concerning the compatibility of autonomy and the conception of moral requirement to which Kant is also certainly committed – one which conceives of the moral law as a strictly universal and necessary imperative.

This thesis has two main aims. In the first half, I offer an interpretation of Kantian autonomy that both accommodates the universality and necessity of moral constraint and takes seriously the notion that the rational will is a legislator of moral law. As a means of developing and securing my preferred view, I argue that recent popular interpretations of Kantian autonomy fail to resolve the tensions that seem at first glance to plague the concept of self-legislation, where what is at stake is the legislation of a categorical imperative. In the second half of this thesis, I examine the connections between my preferred interpretation of self-legislation and Kant's dichotomisation of reason and our sensuous nature. I argue that some of the more harsh and seemingly unreasonable aspects of Kant's moral philosophy can be defended by bringing to light the ways in which they are connected to his commitment both to the autonomy of the will and to developing a genuinely normative ethics.

ACKNOWLEDGEMENTS

There are many people without whose guidance and support I could not have written this thesis. I would first like to thank my supervisors, Jens Timmermann and Sarah Broadie, whose work continues to be a source of great inspiration to me. I am extremely grateful for their comments, their encouragement and their unfailing patience.

I would also like to extend my great thanks to the staff and students of the Philosophy Departments at the University of St Andrews. In particular, I would like to thank my fellow St Andrews Rhinemaids, Alice Pinheiro Walla and Lucy Richmond, as well as our one and only Rhineman, Ralf Bader, for their insightful and helpful feedback on many of the arguments presented in this thesis.

During the course of my study, I received scholarships from the Social Sciences and Humanities Research Council of Canada, the Overseas Research Students Awards Scheme and the St Andrews Department of Philosophy. I gratefully acknowledge these sponsors, whose financial support made it possible for me to undertake this project.

Finally, I would like to extend my very deepest thanks to my wonderful family and friends for their unfaltering support. I truly could not have balanced the countless changes and challenges of the past year in particular without them. This thesis was completed under quite special circumstances, as its submission followed the birth of my beautiful daughter, Katherine Isla. Thanks to Katie and to her wonderful father, Richard; I love you both with all of my heart.

CONTENTS

| | |
|---|-----|
| Introduction | 1 |
| Tensions within the Notion of Autonomy | 8 |
| 1.1 <i>The Principle of Autonomy</i> | 9 |
| 1.2 <i>Self-Legislation</i> | 16 |
| 1.3 <i>Legislation and Choice</i> | 20 |
| 1.4 <i>Freedom</i> | 29 |
| 1.5 <i>Lingering Objections</i> | 35 |
| A Will Giving Universal Law | 41 |
| 2.1 <i>The Command Thesis and Radical Constructivism</i> | 42 |
| 2.2 <i>Reason and Reasoning</i> | 50 |
| 2.3 <i>The Law and the Good</i> | 60 |
| 2.4 <i>The Two Faces of Legislation</i> | 65 |
| 2.5 <i>The Fate of Command</i> | 74 |
| Pure Rational Constraint | 76 |
| 3.1 <i>A Teleology of Practical Reasoning</i> | 77 |
| 3.2 <i>The Rational Grounds of Morality</i> | 89 |
| 3.3 <i>The Legislation of Reason</i> | 98 |
| 3.4 <i>From Autonomy to the Dualism of Practical Interest</i> | 107 |
| Material Principles and the Form of Law | 109 |
| 4.1 <i>Form and Matter</i> | 111 |
| 4.2 <i>Duty and Desire</i> | 120 |
| 4.3 <i>The Heterogeneity of Moral and Non-Moral Concern</i> | 127 |
| 4.4 <i>Interest, Dialectic and Character</i> | 132 |
| On the Purity of the Moral Disposition | 137 |
| 5.1 <i>The Concept of a Supreme Maxim</i> | 139 |
| 5.2 <i>Empirical and Intelligible Character</i> | 145 |
| 5.3 <i>Conditioned Goodness</i> | 152 |
| 5.4 <i>Assessment and Luck</i> | 163 |
| Moral Identity | 167 |
| 6.1 <i>The Tyranny of Reason</i> | 170 |
| 6.2 <i>Confronting the Moral Law</i> | 174 |
| 6.3 <i>Identity and Moral Education</i> | 180 |
| 6.4 <i>Identity and Normative Ethics</i> | 189 |
| Conclusion | 194 |
| Bibliography | 197 |

INTRODUCTION

Central Problems and Aims

Connected as it is to concepts such as liberty, integrity, independence and self-expression, the idea of autonomy has played a significant role in twentieth and twenty-first century moral and political philosophy. To the degree that philosophers wish to ground its importance on its Kantian ancestry, however, they face formidable difficulties. As some have recently argued, Kant's understanding of the meaning and value of autonomy is quite foreign to the highly personalised concept at work in contemporary debate; because of this, it is associated with a unique set of puzzles and worries, which bear little resemblance to those affecting the popular notion.¹

Because my focus is Kant's particular conception of self-legislation and its place in his ethical thought, this thesis largely disregards many of the interesting problems brought to light in the recent literature on personal and political autonomy. Instead, my aim is to come to grips with the continually troubling problem of understanding what Kant meant when he proclaimed that the moral law is a law that the rational will legislates and to defend the implications of my preferred interpretation against a cluster of important criticisms. As such, the following is primarily an internal project; whilst I hope to provide a picture of Kantian autonomy that can be appreciated by anyone who is interested in normative ethics, my primary goal is to provide answers to questions that find their home within Kantian ethics in particular. As will become

¹ This point is argued convincingly by Onora O'Neill in 'Autonomy: The Emperor's New Clothes,' *Aristotelian Society Supplementary Volume* 77 (2003): 1-21. See also Thomas E. Hill, Jr., 'The Kantian Conception of Autonomy,' *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca, NY: Cornell University Press, 1992) 76-96.

clear, I take many familiar Kantian starting points for granted throughout the arguments to follow. Where my reader disagrees with these, he or she is unlikely to be persuaded by the story that builds on them.²

My aim in the first part of this thesis is to provide an interpretation of Kantian autonomy that does justice both to the distinctness of Kant's ethics and to the legal metaphor at its heart. Recently, interpretations of Kantian autonomy have tended to assume two different postures. On the one hand, scholars have sought to provide relatively literal interpretations of the idea of 'legislating for oneself,' in which the powers, capacities and activities of individual agents play a central role in explaining the authority of the moral law; on the other, there has been an attempt to maintain a more straightforwardly realist interpretation of moral constraint, according to which the normative authority of the moral law is not in any sense 'created' by human beings.³ Problems attend both approaches. Where emphasis is placed on the supposedly norm-generating powers and capacities of rational agents, what we find is that key features of the Kantian moral law (its categoricity, universality and necessity) are no longer accommodated. By contrast, where emphasis is placed on the independent authority of the moral law, what is needed is a detailed and coherent account of the sense in which 'self-legislation' is a term that still finds a place in Kantian ethics. Although preliminary explanations of this second sort are to be found in the literature

² In this regard, my aim differs from the more ambitious project of Kant's *Groundwork of the Metaphysics of Morals*, which sought to establish that if one wants to believe in morality at all, one must believe in autonomy of the will.

³ Examples of the first camp include the views proposed by Christine Korsgaard and Andrews Reath (see chapters two and three). Karl Ameriks's interpretation of Kantian autonomy provides a fine and convincing example of a moderate version of the second approach (see in particular *Interpreting Kant's Critiques* [Oxford: Oxford University Press, 2003] chs. 10 and 11; and *Kant and the Fate of Autonomy* [Cambridge: Cambridge University Press, 2000], esp. Part I). For a more straightforwardly realist interpretation of Kant's metaethical position, see Allen Wood, *Kantian Ethics* (Cambridge: Cambridge University Press, 2008) esp. ch. 6.

(most notably in the work of Karl Ameriks), many of the central claims attached to this approach require much further elaboration and clarification, particularly when it comes to the claim that reason functions as the formal legislative ‘source’ of the moral law. In what follows, I will argue that the first tack sketched above applies only to a very specific (and limited) aspect of our obligation to the moral law. I will be offering an interpretation that follows largely in the spirit of the second tack, though one which deals with the vexing questions left unanswered in previous studies. In this sense, my work aims to supply much-needed bridgework for a realist interpretation of Kantian moral theory which is genuinely able to accommodate Kant’s insistence that the will is a legislator of moral law.⁴

Careful attention to Kant’s arguments helps to clear up some of the interpretative controversies surrounding his claims. At the same time, however, it brings out the significant difficulties that an ethics of autonomy must confront. My aim in the second half of this thesis is to focus on the connections between the interpretation of Kantian autonomy I will be advancing and the strict dichotomisation of reason and desire which lies at the heart of Kant’s conceptions of moral worth and character. Many of Kant’s most sincere supporters have found his motivational and character rigorism to be a highly undesirable feature of his moral theory; finding these elements to be either expendable or amendable, they have sought to temper Kant’s penchant for dichotomies by offering alternative ‘Kantian’ conceptions of motivation and character in which the line between reason and desire is rendered less stern. Yet because of the important relationship between the dichotomisation of these elements of our nature and the specific aspects of autonomy I will be highlighting, they can ad-

⁴ As will become clear, we must use ‘realist’ carefully in the context of Kantian ethics. This is a term that often obscures more than it clarifies, and in this context I merely use it to designate a position according to which the authority of the moral law is not in any sense created by human beings.

vance these alternatives only at the cost of putting great stress on the inner coherence of Kant's moral theory. My hope is thus to defend what at times seems to be the more harsh and unreasonable of Kant's tenets by bringing to light the ways in which they are ultimately tethered to his most prized discovery concerning the source of moral constraint.

Context and Significance

I undertook the following study in the hope that a detailed engagement with the complications affecting Kant's conception of autonomy can bring clarity and precision to contemporary interpretation. My research has been undertaken in the midst of a boom in English-language interpretations of Kant, in which non-realist, constructivist understandings of autonomy have taken centre stage. In this sense, my own interpretation goes against the grain, and I will be spending much energy deflecting and rejecting the dominant English-language approach. One narrow significance of this project for the broader Kantian community, then, is its contribution to criticism of this strand of thinking. In a more positive light, it also carries significance for the Kantian audience in the sense that it provides an improved way of thinking about Kantian autonomy – one which is firmly rooted in orthodox Kantian doctrine, and one which has yet to receive a specific and coherent articulation in the vast literature on autonomy. More broadly still, my additional hope is to bring to light even to non-Kantians the distinctiveness of Kant's approach to morality and the true extent of his firm anti-consequentialism. If my account helps in any way to convince non-Kantians that there are benefits to advancing an ethics of autonomy, so much the better.

The Structure of the Project

This thesis contains six chapters. Chapter one, ‘Tensions within the Notion of Autonomy,’ introduces the key starting points for the chapters to follow and includes a brief sketch of Kant’s introduction of the concept of autonomy (both principle and property) in the *Groundwork*. It also introduces two important interpretative theses, which have historically played an important role in coming to understand Kantian autonomy, and it explores the various complications and tensions attached to each. In chapter two, ‘A Will Giving Universal Law,’ I focus in particular on the first of these theses, which I call the Command Thesis. This is the approach most intimately connected to what has come to be known as Kantian constructivism, and in this chapter I argue that there are inherent difficulties in reconciling this approach with Kant’s conception of the key characteristics of morality. I argue that the language of ‘authorship’ when it comes to the moral law is less straightforward than the Command Thesis makes it out to be and that there is evidence in Kant’s writing to support the existence of an important, and in my mind much more Kantian, alternative. In chapter three, ‘Pure Rational Constraint,’ my focus is the second of the theses outlined in chapter one, which I call the Rationality Thesis. The aim in this chapter is to make sense of the idea that the moral law is a ‘law of reason’ in a way that avoids the quite serious pitfalls afflicting recent constitutivist interpretations of this notion. It is in this chapter that I advance my positive interpretation of Kantian autonomy, one which ultimately incorporates elements from both the Command and the Rationality theses, though in very specific ways.

Chapters four to six focus on the implications of this interpretation for some of the more difficult worries advanced against Kantian ethics. In chapter four, ‘Material Principles and the Form of Law,’ I begin by articulating the connection between the conception of autonomy I have been advancing and Kant’s derivation of the content

of the categorical imperative. One worry that commonly attaches to Kantian ethics concerns the way in which the moral law directs our attention to the form of our maxims and away from the objects of desire, a move which both solidifies the strict dichotomy between reason and desire in Kant's ethics and would seem to render moral interest a very peculiar, almost inhuman matter. In this chapter, I defend Kant against the charges associated with this worry by emphasising the deep connections between his conclusions about moral interest and his conception of the moral law and its source. In chapter five, 'On the Purity of the Moral Disposition,' I turn to the connection between the conception of autonomy outlined in chapter three and Kant's seemingly unpalatable rigorism when it comes to moral character. Again, my argument is defensive; whilst I acknowledge that many are likely to find Kant's conclusions about character and moral worth unappealing, my hope is to make clear their deep connections to the central elements of his ethical thought, including autonomy, and so to make clear their (perhaps unfortunate) indispensability. In my final chapter, 'Moral Identity,' I turn to the question of whether we have any reason to think that Kant was at all concerned about the apparently despotic character of reason in his moral theory, despite the necessity of presenting it as utterly separate, authoritative and pure. As I argue, attention to how Kant deals with this issue brings to light a promising solution to the problem of how to develop a lasting and settled commitment to morality in the face of the obstacles afforded by our sensuous nature – a question which vexes both Kant himself and countless of his followers.

Note on Quotations and Translations

All references to Kant are to the volume, page and line numbers (where appropriate) of the Royal Prussian Academy of Sciences edition of Immanuel Kant, *Gesammelte Schriften* (Berlin: Walter de Gruyter, 1900-). An exception is made for the Critique of

Pure Reason (*KrV*), for which page numbers for the first (A) and second (B) editions are provided. Specific published works are cited by means of the abbreviations listed below, and unless I have indicated otherwise, English quotations have been adapted from the Cambridge Edition of Kant's works, published under the general editorship of Paul Guyer and Allen W. Wood. Unless there is an indication to the contrary, all italicisation reflects Kant's original text.

Abbreviations

| | |
|------------|---|
| <i>G</i> | Groundwork of the Metaphysics of Morals (<i>Grundlegung zur Metaphysik der Sitten</i>) (1785) |
| <i>KpV</i> | Critique of Practical Reason (<i>Kritik der praktischen Vernunft</i>) (1788) |
| <i>KrV</i> | Critique of Pure Reason (<i>Kritik der reinen Vernunft</i>) (1781, 1787) |
| <i>KU</i> | Critique of the Power of Judgement (<i>Kritik der Urteilkraft</i>) (1790) |
| <i>MdS</i> | The Metaphysics of Morals (<i>Metaphysik der Sitten</i>) (1797) |
| <i>R</i> | Handwritten Notes: Reflections on Moral Philosophy (<i>Handschriftlicher Nachlaß: Reflexionen zur Moralphilosophie</i>) |
| <i>Rel</i> | Religion Within the Bounds of Mere Reason (<i>Religion innerhalb der Grenzen der bloßen Vernunft</i>) (1793) |
| <i>SF</i> | Conflict of the Faculties (<i>Der Streit der Fakultäten</i>) (1798) |
| <i>TP</i> | 'On the Common Saying: that may be correct in theory, but it is of no use in practice' ('Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis') (1793) |
| <i>UP</i> | Lectures on Pedagogy (<i>Über Pädagogik</i>) (1803) |
| <i>VC</i> | Moral Philosophy: Collins Lecture Notes (<i>Vorlesung zur Moralphilosophie: Moralphilosophie Collins</i>) (c. 1784-5 [perhaps 1774-79]) |
| <i>VL</i> | Moral Philosophy: Vigilantius Lecture Notes (<i>Vorlesung zur Moralphilosophie: Metaphysik der Sitten Vigilantius</i>) (1793-94) |
| <i>VM</i> | Moral Philosophy: Mrongovius Lecture Notes, first and second set (<i>Vorlesung zur Moralphilosophie: Moral Mrongovius; Moral Mrongovius II</i>) (c. 1774-77; 1784-5) |
| <i>WA</i> | 'An Answer to the Question: What is Enlightenment?' ('Beantwortung der Frage: Was ist Aufklärung?') (1784) |

TENSIONS WITHIN THE NOTION OF AUTONOMY

Despite its frequent attachment to a certain sternness and asceticism, Kant's ethical thought has something of a rebellious aspect. Like those who had made previous efforts to determine the source and content of moral principles, Kant did of course emphasise the strict bindingness of morality and its standards; when it comes to morality, he writes, we are not volunteers, but are rather bound to the moral law in a relation of unconditional duty (*KpV* 5:82). Unlike those who came before, however, Kant proposes that the very possibility of a categorically binding imperative is conditional on the rational will's capacity to be a law to itself — on its capacity to lay down its very own commands, and so to possess the property of autonomy. Obedience to the practical rules imposed by church, state and custom must certainly have their place in our lives, and this Kant recognised; yet he also observed that any rule which has its source outside of the rational will can be acted on only in response to some equally 'external' attraction or constraint, some more fundamental interest, in which case the imperative contained within it must be viewed as conditional.¹ In order to explain the possibility of a categorical imperative, Kant denies that the moral law

¹ See *G* 4:432-33 and *KpV* 5:147. For Kant's criticism of heteronomous moral systems, see also *G* 4:432.25-433.11 and *KpV* 5:539-41.

might find its ultimate source in God, the state or nature and rather affirms the quite innovative idea that its source lies in the will of every single rational being. Where the rational will is indeed subject to the moral law, Kant took himself to have discovered that it is ‘subject to it in such a way that it must also be viewed *as self-legislating*’ (G 4:431.21-22).²

In this introductory chapter, I lay down some important exegetical ground for the arguments to follow. Since Kant first introduces the concept of autonomy in the context of articulating a third variant of the categorical imperative, I begin by tracing the principle of autonomy, as it is introduced in Section II of the *Groundwork of the Metaphysics of Morals*. I then go on to explore some of the more difficult questions associated with autonomy as a property of the will, identifying two important interpretive theses and the quite significant problems attached to each. The remainder of this chapter works toward a better understanding of the basic structure of the Kantian will, the relationship between autonomy and freedom, and the shape of our options when it comes to explaining and supporting the notion of self-legislation.

1.1 *The Principle of Autonomy*

At the close of the second section of the *Groundwork*, Kant claims to have shown that autonomy of the will is the supreme foundation of morality:

² *Als selbstgesetzgebend*; alternatively: ‘as itself lawgiving.’ In the *Groundwork*, the concept of autonomy is introduced at 4:431.16-18 and is more fully developed at 4:433.10 and 4:440.14-32, where it is explicitly characterised as ‘the property of the will by which it is a law to itself (independently of any property of the objects of volition).’ In the second *Critique*, autonomy is officially defined at 5:33.8-33.

By explicating the generally received concept of morality we showed only that an autonomy of the will is unavoidably attached to it, or much rather lies at its basis [*zum Grunde liege*]. Thus whoever holds morality to be something and not a chimerical idea without any truth must also admit the principle of morality brought forward. (4:445)

Autonomy first appears in the *Groundwork* at an important transition in Section II, where Kant suggests that a third ‘practical principle of the will’ follows from the previous variants of the categorical imperative — a principle here stated as ‘the idea of the will of every rational being as a will that legislates universal law’ (G 4:431.6-8).³ This rendition of the Formula of Autonomy is curious, for while it seems to describe a particular property or characteristic (*Beschaffenheit*) of the will (the property of legislating universal law), Kant introduces it under the guise of a variant of the categorical imperative, and so as a principle capable of regulating the selection of maxims.⁴ This complexity requires that we make sense of the connections between the various appeals to ‘autonomy’ in Kant’s text, on the assumption that there is a central idea uniting them. In this section, I briefly consider the principle of autonomy and its role in Kant’s argument. The remainder of this chapter is devoted to considering how we might begin to interpret the more fundamental notion of autonomy as a property of the will.

³ An ‘idea’ (*Idee*) for Kant is a necessary concept of reason, to which no empirical object can ever correspond (KrV A 327/B 383).

⁴ While Rüdiger Bittner argues that this ambiguity poses a problem for Kant’s position, both Henry Allison and Andrews Reath convincingly argue that this shift in meaning is not philosophically problematic. See Bittner, *What Reason Demands*, trans. Theodore Talbot (1983; Cambridge: Cambridge University Press, 1989) 76-77, 75-80; Allison, *Kant’s Theory of Freedom* (Cambridge: Cambridge University Press, 1990) 105-6; Reath, ‘Autonomy of the Will as the Foundation of Morality,’ *Agency and Autonomy in Kant’s Moral Theory* (Oxford: Oxford University Press, 2006) 159n.5.

According to Kant, there is only one categorical imperative, the general formulation of which is stated at *G* 4:421.6-8 (the ‘Formula of Universal Law’).⁵ In addition to the general formula, Kant articulates three variants, which collectively constitute a ‘progression’ (*Fortgang*) towards metaphysics.⁶ For the purposes of practical judgement (i.e. determining what it is that we ought to do), we are to follow the ‘strict method’ of the general formula; the purpose of the variants, by contrast, is to help to secure subjective acceptance of the moral law. Because they involve terms that find analogies in familiar moral theories (a law of nature, an end in itself, a kingdom of ends), they bring closer to intuition the merits of acting morally well, strengthening our reverence for the moral law (*G* 4:436.26-437.4).⁷ The first variant tells us something about the form of a morally permissible maxim: it must be such that it can be willed as a *universal law of nature*. The second, by contrast, reveals something about the matter of a moral maxim, or its end: it is the objective end consisting in every *rational being*, and it restricts the choice of all subjective ends. The key ingredients in these variants are, accordingly, the ideas of universal law and rational beings, and

⁵ ‘There is, therefore, only a single categorical imperative and it is this: act only in accordance with that maxim through which you can at the same time will that it become a universal law.’ The first variant directs agents to ‘act as if the maxim of your action were to become by your will a *universal law of nature*’ (4:421.18-20; emphasis reduced); the second directs: ‘[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’ (4:429.10-12).

⁶ The Law of Nature formulation (the first variant of the general formula) introduces the metaphysical notion of a supersensible nature, and thus constitutes the initial pathway into a metaphysics of morals.

⁷ The relevant ‘general formula’ here is the Universal Law formula from 4:421. This contrasts with Allen Wood’s suggestion that Kant’s reference to the ‘general formula’ at 4:436.32-437.1 points to the formula of autonomy (see *Kant’s Ethical Thought* [Cambridge: Cambridge University Press, 1999] 188). Although Kant suggests in the *Groundwork* that the general formulation is sufficient for practical judgement, a more complicated story is provided in the *Critique of Practical Reason*’s ‘On the Typic of Pure Practical Judgement’ (5:67-71).

Kant incorporates both in his identification of the idea of every rational being as a will giving universal law.⁸ As he writes, the ground of practical lawgiving lies *objectively* ‘in the form of universality which makes it fit to be a law (possibly a law of nature)’; *subjectively*, however, it lies in the end, which in this case is ‘every rational being as an end in itself.’ What follows, then, is ‘the third practical principle of the will, as supreme condition of its harmony with universal practical reason, the idea of the will of every rational being as a will giving universal law’ (G 4:431.9-18).⁹

What makes interpretation of the Formula of Autonomy difficult, however, is the fact that Kant presents this variant in a number of different renderings. Worse still, the version picked out by Kant as a canonical formulation does not have the form of an imperative at all.¹⁰ Although it is difficult to make out precisely how the Formula of

⁸ As Kant writes, ‘[t]he above three ways of representing the principle of morality are at bottom only so many formulae of the very same law, and one of them of itself unites the other two in it’ (G 4:436.8-10) (the original Gregor translation, here amended, makes it sound as though *any* one of the variants can be seen as uniting the other two, which is obviously not an accurate description). For a helpful summary of the connections between the variants (and the general principle), see Jens Timmermann, *Kant’s Groundwork of the Metaphysics of Morals: A Commentary* (Cambridge: Cambridge University Press, 2007) 109-112.

⁹ That Kant means to say that it is the Law of Nature variant that is combined with the End in Itself variant is confirmed at 4:431.25, where he makes clear that he is interested in the combination of the idea of ‘conformity of actions with universal law similar to a *natural order*’ and ‘the universal *supremacy as ends* of rational beings in themselves.’ Compare with Paton’s interpretation in his commentary on the *Groundwork*, where he suggests that the Formula of Autonomy ‘is derived from combining the Formula of Universal Law and the Formula of the End in Itself.’ See ‘Analysis of the Argument’ *The Moral Law: Groundwork of the Metaphysics of Morals* (1948; London: Routledge, 1991) 19. See also Stephen Engstrom, *The Form of Practical Knowledge: A Study of the Categorical Imperative* (Cambridge, MA: Harvard University Press, 2009) 150-51. Curiously, Engstrom equates the general formulation with the Law of Nature formula (160).

¹⁰ E.g. ‘... all maxims are repudiated which are not consistent with the will’s own giving of universal law’ (4:431); the moral law commands ‘that everything be done from the maxim of one’s own will as a will that could at the same time have as its object itself as giving univer-

Autonomy is meant to function, what is at least clear is that it brings to light in explicit terms the nature of the *subject* of universal legislation — namely, that the unconditionally binding and universally valid moral law is in some sense imposed by every rational will on itself, such that the will transgresses *its own law* when it adopts an immoral maxim. The principle thus indicates a particular conception of the relation of any rational will to the necessary and universal law that governs it.

The Law of Nature Formula emphasises a formal feature of maxims which are allowed by the categorical imperative: their strict universality, and so their fitness for universal laws of nature. Because we are already familiar with the system of natural laws, the analogy with natural law helps to bring the content of the moral law closer to intuition: to present it in a more concrete manner for our reflection. Likewise, the End in Itself variant emphasises the objective end, the adoption of which is required by the moral law. Again, this exposes morality and its merits more vividly. In just the same way, Kant's third variant of the imperative brings us to a new way of reflecting on what the moral law tells us to do. It may now be conceived as requiring that we act only on maxims which incorporate or embody a universal law that originates in our own will, or, put negatively, that we refrain from seeking the law which is to determine our actions in anything but the will, and so in any object external to it.¹¹

sal law...' (4:432); '...to act on no other maxim than one such that it would be consistent with it to be a universal law, and hence to act only so that the will could regard itself as at the same time giving law through its maxim' (4:434); 'The principle of autonomy is, therefore: to choose only in such a way that the maxims of your choice are also included as universal law in the same volition' (4:440). The canonical formulation is presented at 4:432 and is described by Kant as 'the *principle* of every human will as *a will giving universal law through all its maxims*.'

¹¹ As we will see, there are complications to this idea.

It is probable that the Formula of Autonomy does not of itself constitute a decision procedure.¹² Rather, its usefulness lies in its ability better to secure motivation on the part of the moral agent to do what the moral law commands. Human reason's ability to represent to itself the general formulation makes it such that it 'knows very well how to distinguish in every case that comes up what is good and what is evil, what is in conformity with duty or contrary to duty' (*G* 4:404.2-3); the principle of autonomy, by contrast, reveals to us that moral principles are grounded firmly in the will itself and that, because of this, we follow our very own laws when we do what we ought to do. What is needed in addition to an understanding of the general formula is a means by which to make the ordinary agent more *attentive* and *responsive* to the will's own principle, and it is here that the Formula of Autonomy finds its most useful purpose. As we see in Kant's transition from the first section of the *Groundwork* to the second, the reasons for embarking on a metaphysics of morals are largely practical: what is needed is a way to 'provide access and durability' for morality's precepts in light of the attractions and distractions stemming from inclination (see 4:404-8).¹³ To the degree that it represents moral laws as stemming from our very own will, the third variant connects obedience to moral law with a unique form of *dignity*; as a result it is likely, as Kant suggests, to bring the moral law 'closer to intuition ... and thereby to feeling' (*G* 4:436.12-13).¹⁴

¹² This interpretation is complicated by the fact that Kant seems to suggest otherwise in a footnote at 4:432. Even if the claim in this note can be vindicated (and this would have to be shown), it might still be the case that the *primary* significance of the third variant concerns motivation, as described below.

¹³ The significance of this problem will be the focus of chapter six.

¹⁴ On the relationship between autonomy and dignity, see *G* 4:434-36.

According to Kant, all earlier systems of moral philosophy fall prey to the same mistake: they do not acknowledge the fact that the moral law must be a principle of autonomy:

Now, if we look back to all previous efforts that have ever been made to discover the principle of morality, we need not wonder now why all of them had to fail. It was seen that the human being is bound to laws by his duty, but it never occurred to them 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 universally legislating will. (*G* 4:432.25-32)

Not only types of will but also moral theories are to be classified as either autonomous or heteronomous, depending on which model of the will they assume (*G* 4:441-4, *KpV* 5:39-41).¹⁵ This passage expresses a condition on any moral theory's ability adequately to characterise the moral law: it must render the moral law as a law which the rational will gives to itself. That is to say, it must represent the will as a legislator of sorts, and not merely as subject to moral requirement. Just why this should be so requires that we consider in further detail what it means for a will to give itself its very own law.

¹⁵ As we will consider in further detail in chapter three, Kantian ethics is an ethics of autonomy to the extent that it presupposes that a good will determines itself in accordance with the moral law on the basis of its supreme authority *alone*, without having to put 'underneath' that law 'some incentive or interest as a basis' (*G* 4:444). Chapter three will also give significant attention to Kant's reasons for thinking that all heteronomous ethical theories *had* to fail in their quest to ground a genuinely *moral* imperative.

1.2 Self-Legislation

As Kant's description of the principle of autonomy makes clear, the property of autonomy is to be understood at least in part in terms of the reflexive notion of self-legislation. The moral law is the will's own law, both in the sense that it is given by the rational will and in the sense that it is given *to* the will. Beyond this observation, however, what we need to understand is what this activity of 'giving law' or 'legislation' amounts to.¹⁶

One option on this front is to offer quite a literal reading of self-legislation as involving the will's promulgation of law, where 'promulgation' is to be understood along the lines of the issuing or expression of a *command*. The role served by such a command might be interpreted in a number of ways, though one option would be to hold that it puts the moral law into effect for an agent, generating a practical obligation. This approach is characterised by its commitment to the following, which I shall call the *Command Thesis*:

(CT): The moral law is a law which the rational agent commands to him- or herself, such that the act of command may be described as the source of the obligation to act in accordance with the law.¹⁷

¹⁶ It is important to note that, while there are countless appeals to 'law-giver,' 'lawgiving' and 'legislation' in the *Groundwork*, 'Selbstgesetzgebung' and 'Selbstgesetzgeber' do not make a parallel appearance. We do encounter 'selbstgesetzgebend' at 4:431.22, however, along with 'die Idee des Willens jedes vernünftigen Wesens als eines allgemein gesetzgebenden Willens' (431.16-18) and the idea that all maxims must agree 'mit der eigenen allgemeinen Gesetzgebung des Willens' (431.19-21).

¹⁷ This is similar to what Andrews Reath has recently called the Sovereignty Thesis, according to which '[a]n agent who is subject to an unconditionally valid principle (i.e., a practical law) must be (regarded as) the legislator from whom it receives its authority' (*Agency and Autonomy in Kant's Moral Theory* 122). The Sovereignty Thesis is a particular way of flesh-

This interpretation has the peculiarity of associating autonomy with a notion sometimes linked to what must be seen as *externally* imposed requirements by Kant's standards. Generally speaking, the language of moral command is most immediately associated with the kinds of theological moralities which Kant criticises, and in an important essay even Kant himself presents the essential activities of various external authorities — in particular the officer, the tax official and the clergyman — precisely in terms of strict command.¹⁸ With this said, however, the Command Thesis has the virtue of accommodating much of Kant's language, which at times quite explicitly appeals to the commanding power of the rational will.¹⁹ What is more, it preserves the traditional association between the idea of law, or legislation, and the act of promulgation. It is natural, after all, to suppose that if there are requirements on us, this fact can be traced to our having *been required*,²⁰ and though in the case of moral self-legislation the relevant act of the will might not take the form of a literal *declaration* (consisting instead, perhaps, in a decision, endorsement, commitment, or choice

ing out Reath's Legislation Thesis, which states that 'the moral law, and the requirements to which it leads, are laws that the rational will legislates' (92). It is important to note that the Command Thesis only accurately describes the autonomy of an imperfectly rational will. Since a perfect will does not confront the law as an imperative, its representation of the moral law to itself (while perhaps in some extended sense a 'command') does not generate an 'ought' or an 'obligation.'

¹⁸ Consider Kant's 'An Answer to the Question: What is Enlightenment?' 8:37: 'The officer says: Do not argue but drill! The tax official: Do not argue but pay! The clergyman: Do not argue but believe!'

¹⁹ E.g. *MdS* 6:227: 'A (morally practical) *law* is a proposition that contains a categorical imperative (a command). The one who commands (*imperans*) through a law is the lawgiver (*legislator*).'

²⁰ In support of this theme, see Robert Merrihew Adams, *Finite and Infinite Goods: A Framework for Ethics* (New York: Oxford University Press, 1999) 262-63. As we will see in chapter two, this common intuition must, in Kant's case, be rendered more complicated.

that x be done), the idea that agents must be bound to morality by some form of law-giving *activity* is likely to agree with many people's intuitions.

Besides the obvious problem of how to interpret the notion of 'command' in this context, there are at least two additional difficulties with this approach. On the one hand, the very notion of an individual *self* that actively and literally legislates may seem to capture the wrong idea when it comes to Kant's conception of moral requirement. Moral laws are not mere personal decisions: they are meant to apply to certain classes or groups of agents and to many different situations. But if individuals could issue genuine moral commands to others, those others would be bound by those laws, and so would be subject to the (moral) authority of another — a problematic conclusion for a supposed ethics of autonomy.²¹ On the other hand, if individuals can command moral laws *only* to themselves, the language of legislation becomes an exaggeration: an overly dramatic way of describing what amounts to individual choice or commitment.

On the face of things, then, there is a problem in supposing that individual legislation constitutes the source of moral obligation. Either we find ourselves saddled with a view according to which individual agents are subject to external moral authorities, or we would seem utterly to collapse the distinction between moral obligation and personal choice. As attractive as this latter move might seem to some philosophical camps, the worry is that individual 'autonomy' of this sort cannot provide an ad-

²¹ This point is quite helpfully articulated in O'Neill, 'Self-legislation, Autonomy and the Form of Law,' *Right, History, Religion: Kant's Present Significance*, International Symposium, Austrian Academy of Sciences, Vienna, 5th March, 2004. Printed in *Recht – Geschichte – Religion: Die Bedeutung Kants für die Gegenwart*, ed. Herta Nagl-Docekal and Rudolf Langthaler, spec. issue of *Deutsche Zeitschrift für Philosophie* 9 (2004): 14.

equate basis for a Kantian account of moral obligation, bound as it is to the concepts of universality and strict necessity.²²

Should we then deny the Command Thesis by stressing that it is the *rationality* of moral laws, and not the mere fact that we have chosen them, which binds us? Consider, for example, the following view, which we might call the *Rationality Thesis*:

(RT): The moral law is a fundamental law of reason, and this fact explains its universal authority.

Since reason's standards would presumably apply to all rational beings on this view, the worry about arbitrariness and individual discretion is resolved. Yet in resorting to the Rationality Thesis, we risk being unable to make sense of the notion that autonomy, or self-legislation, is the key to understanding moral obligation at all. What significance can the concept of self-legislation really have, if the authority of moral principles is located in some conception of reason? Rather than describing moral action as a species of following the will's *own* law (and emphasising the will's legislative power), why don't Kantians simply propose that in acting morally well we are doing what it is right or rational to do?²³

²² As Wood puts this worry, '[t]o make my own will the author of my obligations seems to leave both their content and their bindingness at my discretion, which contradicts the idea that I am obligated by them' (*Kant's Ethical Thought* 156). Onora O'Neill similarly worries that conceptions of individual autonomy, 'whether read voluntaristically as mere, sheer pursuit of preference or more cautiously as pursuit of the right sorts of preferences ... substitute self-expression for moral obligation' (*Autonomy and Trust in Bioethics* [Cambridge: Cambridge University Press, 2002] 90).

²³ As Wood writes, we might very well question whether this view renders the notion of self-legislation 'a deception or at best a euphemism.' See *Kant's Ethical Thought* 156.

The central dilemma facing the notion of self-legislation takes the following form. Kant's description of morality would suggest on the face of things that there is a self that serves as the legislating source of the moral law. Yet to make myself the source of my obligation is, on the face of things, to introduce an unpalatable arbitrariness at the level of moral requirement, rendering questionable the idea that I am actually obliged. The problem is that we run into significant difficulties if we abandon this tack and stress instead the fact that the moral law is a law of reason, binding all agents universally and with necessity. If we choose to pursue this horn, Kant's description of morality as grounded in self-legislation becomes quite curious: for why ought we to understand the relation that the 'self' who is subject to moral requirements bears to the moral law as relevantly different from cases of heteronomy? Why not view the source of the moral law, understood in this way, as lying outwith the individual self to whom it applies, in the already-established 'canons of reason'?²⁴

If we are to resolve these worries, we must consider in much further detail the relationship between self-legislation and practical reason in Kant's ethical thought. As a means of embarking on this discussion, it will be helpful to think about the structure of the Kantian will and the roles assigned to its elements in self-legislation.

1.3 Legislation and Choice

Kantian autonomy is one of a cluster of Elizabeth Anscombe's targets in her seminal paper 'Modern Moral Philosophy,' in which she argues against the use of what she

²⁴ Wood, *Kant's Ethical Thought* 156.

takes to be culturally deracinated concepts such as ‘duty,’ ‘ought’ and ‘obligation’ in ethics. In support of this conclusion, she constructs a genealogy of these term, according to which quite ordinary and indispensable words like ‘should,’ ‘needs’ and ‘ought’ (‘the machinery needs oil, or should or ought to be oiled’) acquired a special sense over time by being equated in the relevant contexts with the idea of being obliged, bound, or required, ‘in the sense in which one can be obliged or bound by law, or something can be required by law.’²⁵ The reason for this shift lies in history: Christianity, with its law conception of ethics, derived its ethical notions from the Torah, and the concepts of being bound, permitted and excused became securely established elements in our language as a consequence of its dominance. Yet despite the secure place of such notions in everyday moral reflection, Anscombe proposes that, with the secularisation of ethics, the context which once gave them their significance has now been abandoned:

To have a *law* conception of ethics is to hold that what is needed for conformity with the virtues failure of which is the mark of being bad *qua* man (and not merely, say, *qua* craftsman or logician) – that what is needed for this, is required by divine law. Naturally it is not possible to have such a conception unless you believe in God as a law-giver; like Jews, Stoics, and Christians. But if such a conception is dominant for many centuries, and then is given up, it is a natural result that the concepts of ‘obligation,’ of being bound or required as by a law, should remain though they had lost their root; and if the word ‘ought’ has become invested in certain contexts with the sense of ‘obligation,’ it too will remain to be spoken with a special emphasis and a special feeling in these contexts.²⁶

²⁵ ‘Modern Moral Philosophy,’ *Philosophy* 33 (1958): 5.

²⁶ ‘Modern Moral Philosophy’ 6. It is worth noting that Kant’s ethics is not ‘secular’ in the sense of excluding an important place for God in the general picture of morality and moral understanding. There is a way of representing the moral law, according to Kant, as a law endorsed or commanded by God; what his ethics does not allow, however, is that God could be the source of the moral law in the usual way – the force that brings it into being and that

The idea of self-legislation arises in this paper as a potential reply to this worry. While the modern moral philosopher might not believe in God as the source of practical requirement, what is to prevent us from appealing to the now-familiar Kantian notion that we in some sense lay down moral requirements for ourselves? Might such a notion allow us to preserve the law-like bindingness of moral principles in the absence of the relevant theological framework? Anscombe's remarkably brief reply is as follows:

Kant introduces the idea of 'legislating for oneself,' which is as absurd as if in these days, when majority votes command great respect, one were to call each reflective decision a man made a *vote* resulting in a majority, which as a matter of proportion is overwhelming, for it is always 1-0. The concept of legislation requires superior power in the legislator.²⁷

On Anscombe's view, moral self-legislation is absurd: where there is the genuine legislation of a normative principle, there must be separate authority on the part of the lawgiving body. What one does 'for oneself' may of course be admirable, but 'it is not legislating.'²⁸

ultimately explains its authority for us. It is not the loss of a belief in God or God's command *per se* that occasions Anscombe's worry, but rather the loss of a belief in God as the sole source of moral obligation.

²⁷ 'Modern Moral Philosophy' 2. There are in fact two separate criticisms embedded in this quotation. First, Anscombe mocks the very idea of an analogy between morality and self-legislation because of the connection, on her view, between legislation and voting. A second and perhaps more serious point concerns the idea that legislation requires 'superior power in the legislator.'

²⁸ 'Modern Moral Philosophy' 13. This worry is anticipated by C. D. Broad, who suggests that 'it is doubtful whether any clear meaning can be attached' to the notion of a self-imposed principle. See *Five Types of Ethical Theory* (London: Kegan Paul, Trench, Trubner & Co. Ltd, 1930) 133.

Let us consider this criticism in further detail. So long as we allow that Kant need not conceive of legislation on the model of a vote according to which a majority wins, what he must do in order to block Anscombe's criticism is to provide positive justification for thinking that the will is divided into two separate offices, one of which possesses 'superior power.' Kant is of course able simply to posit a distinction within the will and to show that self-legislation need not be an incoherent notion, but is there any reason to think that the provision of such a story could be anything more than an *ad hoc* addition?²⁹

In ancient philosophy, a strong motivation for endorsing a conception of the soul as divided into different stations is the relationship between such a picture and the possibility of inner conflict and acratia conduct. Here, the idea is that the explanation of situations where one finds that one is pulling oneself in different directions depends on there being different parts of the soul. The acceptance of the phenomena provides a strong justification for speaking about ourselves in this way.³⁰ In Kantian

²⁹ Since Anscombe is in effect mocking the very idea that a will should be thought of as divided into separate offices, simply positing such a distinction is unlikely to satisfy her criticism. What we need is a positive argument in favour of such a picture.

³⁰ See, e.g. *Republic* IV 439a-441c, where we encounter the examples of the thirsty man who refuses to drink, along with Leontius and the corpses (Plato, *Republic*, trans. G. M. A. Grube, rev. C. D. C. Reeve, *Complete Works*, ed. John M. Cooper, associate ed. D. S. Hutchinson [Indianapolis, IN: Hackett Publishing Company, Inc. 1997]). See also Aristotle's *Nicomachean Ethics*, I.13.1102b14-25: 'For in the continent and incontinent person we praise their reason, that is to say, the [part] of the soul that has reason, because it exhorts them correctly and toward what is best; but they evidently also have in them some other [part] that is by nature something apart from reason, clashing and struggling with reason. For just as paralysed parts of a body, when we decide to move them to the right, do the contrary and move off to the left, the same is true of the soul; for incontinent people have impulses in contrary directions ... [W]e should suppose that the soul also has something apart from reason, countering and opposing reason' (trans., ed. Terence Irwin, 2nd ed. [Indianapolis, IN: Hackett Publishing Company, Inc., 1999]).

ethics, however, the motivation for supposing that the soul must have different parts, one exercising rightful authority over the other, does not stem from *a posteriori* psychological observation. Instead, this picture is taken to follow as a consequence of positions which Kant thinks that moral philosophers must (and do) accept, in the sense that very fundamental tenets cannot be maintained in its absence. One striking example of this occurs in the *Metaphysics of Morals*, where Kant discusses the idea of a duty to the self – i.e. a duty whose object is the agent him- or herself alone.³¹ The existence of such duties is a central theme in Kant's ethics: their importance is noted as early as the Collins moral philosophy lecture notes (27:340-41), and they take up one whole side of the classification table of types of moral obligation in the *Groundwork* (4:421-4, 429-30). In the *Metaphysics of Morals*, Kant suggests that there is something of a puzzle attached to the idea of a duty to oneself: if they exist, then I would seem to play two roles with respect to them – I am both passively constrained (to the extent that I am bound) and actively constraining (as the obligating power), resulting in a situation where I am the very subject to which my duty is owed. In Kant's words,

... the I that imposes obligation [*das verpflichtende Ich*] is taken in the same sense as the I that is put under obligation [*dem verpflichteten*] ... [T]he concept of duty contains the concept of being passively constrained (I am *bound*). But if the duty is a duty to myself, I think of myself as *binding* and so as actively constraining (I, the same subject, am imposing obligation). (*MdS* 6:417.7-12)³²

³¹ As Jens Timmermann points out, the category of duties to the self is quite exclusive. Not any duty that happens to concern the very agent it addresses counts as a duty to the self. See 'Kantian Duties to the Self, Explained and Defended,' *Philosophy* 81 (2006): 505-508.

³² A guiding assumption, here, is that a duty to the self is owed to oneself *qua* obligator. In everyday speech, I might be said to 'owe' a duty to someone insofar as she is the recipient or object of my dutiful behaviour (e.g. I owe it to my neighbour not to steal his car, though it is not strictly the case that *he* imposes that obligation, in the sense of binding me directly). In a

The problem is that there is something incoherent, we might suppose, in the thought that a single agent could occupy both roles. Is there any sense in which I am genuinely bound by a duty if I am the very agent imposing the constraint? If I am not bound by a duty, in what sense is there a duty in play at all? Kant ultimately resolves this puzzle, in characteristic fashion, by appealing to two different ways in which we conceive of ourselves. On the one hand, we can consider ourselves as ruling noumena (purely *intelligible* beings endowed with genuine freedom), on the other, as constrained phenomena (*sensuous* beings subject to natural necessity).³³ On the assumption that such a distinction is possible, the ruler and the ruled need not be identical at

Kantian duty to the self, the idea is not merely that I owe *x* to myself *qua* recipient of *x*; rather, it is also that I owe *x* to myself *qua* the imposer of obligation. As a result, there might well be a sense in which all ethical duties (including duties to others) are *in addition* duties to the self; in the case of any duty to *x*, even where I am not the direct recipient of *x*, I might nevertheless be the imposer of the obligation to *x*. Kant supports the claim that duties to the self are foundational with the following short argument: ‘For I can recognise that I am under obligation to others only insofar as I at the same time put myself under obligation, since the law by virtue of which I regard myself as being under obligation proceeds in every case from my own practical reason; and in being constrained by my own reason, I am also the one constraining myself’ (417.25-18.3). Compare Reath, ‘Self-Legislation and Duties to Oneself,’ *Kant’s Metaphysics of Morals: Interpretative Essays*, ed. Mark Timmons (Oxford: Oxford University Press, 2002) 361, where the claim is that it is clearly wrong to think of all duties as coinciding with duties to the self.

³³ Kant thus follows his usual pattern of tracing the illusion of a paradox to the assumption of transcendental realism (the assumption, that is, that phenomena are identical to things as they are in themselves). Note that Gregor’s translation at 6:418.19 obscures what is actually an ambiguity in the original German. Where her translation suggests that *homo noumenon* is regarded as being able to be ‘put under’ obligation to himself, what it ought to convey is the idea that *homo noumenon* is regarded as being capable of ‘obligation’ (*Verpflichtung*), which can be interpreted either as being obligated or as *obligating*. Put in context, Kant’s thought must have been that, when considered as *homo noumenon*, the agent is considered as one who is capable of obligating himself. The original German reads as follows: ‘*Eben derselbe aber seiner Persönlichkeit nach, d. i. als mit innerer Freiheit begabtes Wesen (homo noumenon) gedacht, ist ein der Verpflichtung fähiges Wesen und zwar gegen sich selbst (die Menschheit in seiner Person) betrachtet...*’ (6:418.17-20). For a similar gloss on this complication, see Reath, ‘Self-Legislation and Duties to Oneself’ 355.n.

all, though in the end they must be conceived as constituting different aspects of the same human being (6:418.5-23).

Of course, we might not believe in duties to the self in Kant's specific sense.³⁴ What is more, although an implicit notion of different elements of the self is always at work when we speak of judging ourselves, respecting ourselves, or 'owing it' to ourselves (Kant himself points out the significance of this last expression in a footnote at 6:418), this way of thinking is highly metaphorical. As I aim to show in chapter three, Kant's ultimate argument in support of such a conception will hinge on the notion that unless such a picture of the will is available to us – unless self-legislation is a real possibility for moral theory – there can be no such thing as an unconditionally binding practical principle.³⁵ The existence of a will in which a legislative power rules over and constrains a separate and subordinate executive power is a necessary condition of the possibility of morality, according to Kant, and it is *this* argument that provides his key response to the worry that the concept of self-legislation is meaningless when applied to the human will. This is not to say that we must accept the assumption that there is an unconditionally binding practical law, of course; but it is to suggest that the stakes are high if we wish to share Anscombe's position, for if we do, we can no longer hold out hope for the possibility of a genuine categorical imperative.

³⁴ In any case, we have reason to be suspicious of Kant's quick equation of the object of the duty with the obligator of that duty in the passage outlined above from the *Metaphysics of Morals*.

³⁵ As we shall see, Kant will argue that not even the theological picture described by Anscombe in 'Modern Moral Philosophy' can ground a categorical imperative. Only an autonomous ethics has room for such a principle.

Any claim that the will is divided into ‘parts’ does of course lead to questions concerning the ontological status of the different elements (e.g. could these ‘parts’ exist separately, or are they analogous to elements of something that necessarily entails *both*, like the convex and concave sides of a curve?).³⁶ Yet the simple fact that metaphysical questions such as these may be raised does not imply that the conception is incoherent or philosophically undesirable. Indeed, part of its desirability hinges on what it allows us to say about the nature and possibility of moral requirement, and it is this set of questions with which a foundational ethics is primarily concerned.

To speak of the will as divided is in the first place to identify a difference at the level of functions and capacities. Accordingly, Kant distinguishes two capacities within the rational will and their associated ‘parts’ or ‘powers.’³⁷ The function of actively *giving* law is attributed to *Wille*, the legislative element of the will, while the function of choosing what to do in light of *Wille*’s legislation is attributed to *Willkür*, the executive element of the will (the free power of choice). As Kant writes at *MdS* 6:226:

Laws proceed from the will [*Wille*], *maxims* from choice [*Willkür*]. In man the latter is a free choice; the will, which is directed to nothing beyond the law itself, cannot be called either free or unfree, since it is not directed to actions but immediately to giving laws for the maxims of actions (and is, therefore, practical reason itself). Hence the will directs with absolute necessity and is itself *subject to* no necessitation. Only *choice* can therefore be called *free*.

³⁶ See Aristotle, *Nicomachean Ethics* 1102a30-31. In characterising these metaphysical worries, I am indebted to Sarah Broadie, *Ethics with Aristotle* (New York: Oxford University Press, 1991) 61.

³⁷ This distinction is not explicitly drawn in the *Groundwork* or the second *Critique*, but it is necessary to appeal to something of this sort if we are to begin to understand how a will can at once be the ‘source’ of a practical law (active law-giving) and be bound by that very law.

At *MdS* 6:213-14, Kant distinguishes free (*freie*) *Willkür* as ‘choice which can be determined by pure reason,’ and he describes *Wille* as having the power to determine choice. Thus *Wille* is identical to practical reason. Both elements comprise the broader faculty of *Wille*, understood as volition or will as a whole, though they express significantly distinct functions within this broader power.³⁸

This distinction is welcome not only because it provides a reply (of a limited sort) to Anscombe’s worry, but also because it helps to resolve a troubling ambiguity in the *Groundwork*, where Kant both equates the will with practical reason and describes practical reason as determining (or failing to determine) the will (*G* 4:412). *Wille* in the narrow sense legislates laws, and this explains why the will is identified with practical reason. *Willkür*, on the other hand, is able to choose in accordance with or against the dictates of narrow *Wille*, and this explains why Kant is able to describe reason as either determining or failing to determine the will.

Following Henry Allison, then, we can say that it is the will *as a whole* (*Wille* in the broad sense) which gives the law *to itself*.³⁹ Insofar as narrow *Wille* legislates the moral law to *Willkür*, and insofar as both faculties are ultimately united as different parts of a greater power, this broader faculty, the will proper, is ‘a law to itself (inde-

³⁸ It is somewhat unfortunate that Kant uses the term *Wille* both to describe the broad notion of the will as a whole, which is comprised of both elements, and to describe the sub-faculty of rational legislation, or pure practical reason. With this said, it is possible that Kant does not use the term *Wille* ambiguously in the *Metaphysics of Morals*. When considering the whole of his works, the notion is indeed ambiguous, but there is good reason to think that a clear division was made by 1797. In what follows, I hope to have made my own usage clear by indicating whether a broad or narrow meaning is to be applied.

³⁹ For Allison’s take on the kinds of freedom that can be attributed to these different aspects of the will, see *Kant’s Theory of Freedom* 129-36.

pendently of any property of the objects of volition)’ (*G* 4:440). Kant’s parenthetical addition is crucial: a will with the property or character of autonomy imposes laws on itself, the validity of which is independent of any interests stemming from the agent’s sensuous nature. Autonomy thus refers at least in part to the will’s own ‘laying down’ or ‘legislation’ of laws *of a certain sort*, to which the faculty of choice must then respond.⁴⁰

1.4 Freedom

According to Kant, our everyday notion of free agency involves the (purely negative) thought that genuinely free action is not causally determined by preceding conditions. To use Kant’s language, freedom in this sense involves ‘the independence of our will [*Willkür*] from necessitation through impulses of sensibility’ (*KrV* A 534/B 562).⁴¹

⁴⁰ As I have noted, Kant’s distinction between autonomy and heteronomy is also a distinction between two kinds of moral theory. A theory bases morality on autonomy of the will when it supposes that the basic principle of morality finds its source in pure reason alone, which is able to represent that law to the faculty of choice. By contrast, a theory of heteronomy supposes that morality is grounded on some external good, in which agents have a contingent interest. Kant’s quite novel idea is that any satisfactory moral theory must represent the moral law as legislated by the will to itself: since the moral law applies unconditionally, and since it is conceptually true that agents who are subject to the moral law must be viewed as legislators, a moral theory that does not represent the moral law as self-legislated simply cannot explain its unconditional validity. See, e.g. *G* 4:441-444 and *KpV* 5:33.

⁴¹ Alternatively, independence from determination by ‘alien causes’ (*G* 4:446). See also *KpV* 5:33 and *MdS* 6:213-14. On Kant’s view, the nonhuman animal will (*arbitrium brutum*) is such that it is always necessitated by some impulse afforded by inclination. The human will (*arbitrium liberum*) is not necessitated by inclination but merely ‘affected’ by it: its causality works via representations rather than direct material determination (*KrV* A 534/B 562). Importantly, Kant objects to *any* view according to which free action is grounded in an antecedent natural state – even where it is nonetheless caused ‘from within, by representations produced by our own powers’ (see *KpV* 5:95-97). As he explains, ‘all necessary events in time in accordance with the natural law of causality can be called the *mechanism* of nature,

Attached to this notion is the positive idea of a will that is the cause of its own actions, such that it initiates a novel causal chain. If we are to rule out the possibility that free action might be utterly uncaused, governed by blind chance, there must be *some* causal force at its seat.⁴² We must thus suppose that a free will is not only *independent* of determination by ‘alien’ forces, but also *spontaneous*, in the sense that it constitutes an unconditioned cause, or the absolute beginning of a causal chain.⁴³

In addition, however, Kant proposes that the idea of a negatively free will – a will not subject to the ‘alien’ laws of nature – gives rise to the idea of a will that is governed by its very own *law*. According to Kant, there must always be a law in accordance with which, ‘by something that we call a cause, something else, namely an effect,

although it is not meant by this that the things which are subject to it must be really material *machines*. Here one looks only to the necessity of the connection of events in a time series as it develops in accordance with natural law, whether the subject in which this development takes place is called *automaton materiale*, when the machinery is driven by matter, or with Leibniz *spirituale*, when it is driven by representations; and if the freedom of our will were none other than the latter (say, psychological and comparative but not also transcendental, i.e. absolute), then it would at bottom be nothing better than the freedom of a turnspit, which, when once it is wound up, also accomplishes its movements of itself’ (*KpV* 5:97.7-20). Leibniz describes the human soul ‘as a kind of *spiritual automaton*’ (though it is not subject to what he calls an *absolute* necessity) in the *Theodicy*, trans. E. M. Huggard, ed. Austin Farrer (1710; La Salle, IL: Open Court, 1985) ¶52, p. 151.

⁴² On the idea that a mere absence of natural determinism does not in itself preclude governance by chance (a lawless efficacy), see *KpV* 5:95.14.

⁴³ Absolute spontaneity (*Spontaneität*) is equated to transcendental freedom at *KrV* A 533/B 561. See also *KpV* 5:48.20-3 and 101.11. According to Kant, practical freedom presupposes that the cause of my action ‘in appearance was thus not so determining that there is not a causality in our power of choice such that, independently of those natural causes and even opposed to their power and influence, it might produce something determined in the temporal order in accord with empirical laws, and hence begin a series of occurrences *entirely from itself*’ (*KrV* A 534/B 562). Compare Hume’s description of the ‘liberty of spontaneity’ and the ‘liberty of indifference’ in *A Treatise of Human Nature*, 2.3.2.1 (ed. L. A. Selby-Bigge, rev. P. H. Nidditch [1739-40; Oxford: Clarendon Press, 1978] 407-8).

must be posited' (*G* 4:446.16-18).⁴⁴ Thus causality must operate in accordance with laws of some sort. If the free will were subject to the laws of nature, it would be determined by alien causes, for the 'causes' linked to certain effects by natural laws can only be empirical, and so governed by the determinism that applies to the world of appearances. Given that a free will is not determined by alien causes (this is just the definition of negative freedom), we can conclude that it also cannot be governed by natural law. Yet according to Kant, if the free will is neither lawless nor governed by an alien natural law, it must be governed by its very own law.⁴⁵ Negative freedom thus implies not only spontaneity, but also autonomy. What is more, because the will's own law turns out to be nothing other than the *moral* law, 'a free will and a will under moral laws are one and the same' (*G* 4:447.6-7).

There is thus a clear connection between Kant's conception of free will and morally good action. For Kant, morally good action just is action in which the will conforms

⁴⁴ A lawless cause is, for Kant, an absurdity (*ein Unding*). In the *Groundwork*, this assumption is not given positive support. The connection between negative freedom and autonomy is given a more thorough articulation, however, in Reflection 7220, 19:289: 'Imagine freedom, i.e. a power of choice that is independent of instincts or in general of direction by nature. This freedom in itself would be a rulelessness and the source of all ill and all disorder, where it is not a law to itself. Freedom must accordingly stand under the condition of universal conformity to rules and must be an intelligent freedom, otherwise it is blind or wild.' See also *R* 6961, 19:215.

⁴⁵ As Kant suggests in the *Groundwork*, freedom of the will *is* autonomy, 'i.e. the property of the will to be a law to itself' (4:446.24-447.2). See also *KpV* 5:33.15-21. Kant perhaps illegitimately assumes here that if a law is not a law of nature, it must belong to the will itself in some relevant sense. Given his identification of the will with practical reason (*G* 4:412.28-30), the dichotomy between laws of nature and laws of reason is thus fully in play. Language which associates freedom with a kind of self-imposed law can also be found in Rousseau's *Of the Social Contract* (1762), where the thought is that 'the impulsion of mere appetite is slavery, and obedience to the law one has prescribed to oneself is freedom' (I.8, ¶3) (*The Social Contract and Other Later Political Writings*, trans. ed. Victor Gourevitch (Cambridge: Cambridge University Press, 1997) 54.

to the law of pure reason. As such, it consists in the realisation of freedom. For some of Kant's critics, this spells trouble – for if only morally worthy action *realises* freedom, does this not suggest that only morally worthy action *is* free? Does Kant not seem to be saying that we are responsible only for our morally good actions, but never for those that are immoral? Where does this leave the concept of culpable evil?⁴⁶

In the *Metaphysics of Morals*, Kant clarifies his position by proposing that the freedom of the will in the sense characterised by 'autonomy' must be conceived as a *capacity* to act in accordance with the moral law from the pure motive of duty (6:226.12-227.9).⁴⁷ Put differently, to be 'governed' by the moral law is to possess the power to act from an *a priori* motive utterly unconnected to our sensuous nature.

⁴⁶ For an important articulation of this worry, see Henry Sidgwick, *The Methods of Ethics*, 7th ed. (1874; 7th ed. 1907; Indianapolis, IN: Hackett Publishing Company, 1981) 57-59, 511-16. Here, Sidgwick suggests that we dispose of a conception of freedom as consisting in morally good or rational action whilst maintaining a conception of freedom as the capacity to choose. As Henry Allison points out, this criticism is also commonly attributed to Carl Leonhard Reinhold, who accuses Kant of rendering impossible free yet immoral actions. See Reinhold's 'Erörterung des Begriffs von der Freiheit des Willens,' *Materialien zu Kants Kritik der praktischen Vernunft*, ed. Rüdiger Bittner and Konrad Cramer (Frankfurt am Main: Suhrkamp, 1975) 255-6. See also Allison, *Kant's Theory of Freedom* 133-5. For a discussion of these themes, to which I am greatly indebted here, see Timmermann, *Kant's Groundwork of the Metaphysics of Morals: A Commentary* 164-67.

⁴⁷ See also *VC* 27:267.37-9. This interpretation is developed by Henry Allison in *Kant's Theory of Freedom* (see esp. 97-98). In contrast to this interpretation, Allen Wood proposes that we distinguish between autonomy and practical freedom and that the former is best understood as the full realisation of the capacity in which the latter consists (the freedom to act on moral, as opposed to natural, law) (see, 'Kant's Compatibilism,' *Self and Nature in Kant's Philosophy*, ed. Allen Wood (Ithaca: Cornell University Press, 1984) 73-101. This interpretation allows us to draw a direct connection between autonomous action and morally worthy action, as Kant's language sometimes suggests, and it helps to explain Kant's claim that the categorical imperative '*commands* neither more nor less than just this autonomy' (*G* 4:440; my emphasis). As will become clear, I do not follow Wood's terminology in what follows.

On this interpretation, even the immoral actions of a free will incorporate sensuous incentives *without being necessitated by them* (*KpV* 5:32.26-7). Whilst a free will might in certain cases choose to ‘borrow’ the law which is to determine its actions from nature (and so to act in accordance with it), it is not *subject* to that law. It thus ‘stands under’ moral law in the sense that the latter constitutes an unconditionally authoritative principle by which the will is able to effect its particular form of causality, should it so choose. As such, on this interpretation, there is no intermediate between the full possession of autonomy (*qua* capacity) and the full lack of autonomy (the lack of the capacity). If a will does not possess the property of autonomy, it is heteronomous, and lacks altogether the capacity to be determined by the mere form of law.⁴⁸ A finite yet autonomous will might of course fail to live up to this remarkable capacity when an agent chooses to act immorally. This does not, however, render the will any less autonomous.

⁴⁸ Here I am appealing to a useful adjective, though Kant does not use words like ‘autonomous’ and ‘heteronomous’ in his foundational ethical writings. It is helpful to note, in addition, that in Kant’s writings there is a sense in which a will might be more or less free to the degree that it is influenced by forces suggesting alternatives to the path set out by the moral law. As Kant writes in a note from (around) 1783-4, ‘The freedom of the divine will does not consist in its having been able to choose something other than the best; for not even human freedom consists in that, but in being necessarily determined by the idea of the best, which is lacking in the human being and thereby also limits his freedom’ (*R* 6078, 18:443). If it is to be made consistent with the notion that freedom consists in a capacity, the thought must be that different kinds of will are able to actualise or realise freedom (*qua* capacity) more or less perfectly (though the having of the capacity itself is not a matter of degree). Thus there may be a sense in which I realise my freedom ‘more perfectly’ when I act morally well, since in this case my inclinations do not interfere in my choice (though I am nonetheless affected by them) and reason is practical *of itself alone* (see, e.g. *KpV* 5:24.35-40). C.f. Descartes’s discussion of freedom in Meditation IV, where one important thought is that the more one is inclined by reason in a certain direction (as opposed to experiencing one’s options with complete indifference), the freer is one’s choice (*Meditations on First Philosophy*, trans. ed. John Cottingham, revised ed. (1741; Cambridge: Cambridge University Press, 1996) 40.

This solution, should it be successful, would seem to clear space within Kant's framework for the possibility of freely chosen immoral action. Nevertheless, a serious objection to this conception of freedom emerges from an attempt to explain, on its own terms, what immoral action might involve. As this interpretation asserts, the (free) will is not determined by natural causes. Such a will cannot effect its causality by means of natural law, for it is not 'subject to' or 'governed by' the laws of nature, but is instead governed by the moral law. As such, the moral law is a law in terms of which the free will can effect its causality, should it so choose. Yet if this is true, then the failure to choose the moral law as the law of one's causality is ultimately mysterious. Moral weakness of this sort cannot be the result of determination by natural law (the effect of natural causes), but by definition nor can it be the result of determination by moral law (for then it would not constitute moral failure). Whilst we might speak of the morally weak will as one that allows itself to be affected by sensibility, as one that in some sense chooses to borrow its law from nature, what we cannot say is that it allows itself actually to be *determined* by natural law. But then what does it really *mean* to say that a will 'borrows' its law from nature? By what law *does* it effect its causality, if it does so via neither natural nor moral law?⁴⁹

It is beyond the scope of this project to attempt to resolve this seemingly ineluctable problem. For our purposes, the important point to note is that the will's legislation is intimately connected to a positive conception of freedom on Kant's view. To this point, then, we have come some way in identifying what Kant was up to when he wrote of the will's own giving of universal law. We have considered the elements of the Kantian will and their designated roles in decision-making, identifying in particu-

⁴⁹ In *Religion Within the Limits of Mere Reason*, Kant develops the theme of the inexplicability of moral failure. See, e.g. 6: 43.12-17.

lar the legislative power within every rational being, and we have examined (albeit in brief) the relationship between the will's own legislative power and Kant's conceptions of freedom. Nevertheless, there is much work to be done when it comes to clearing up the puzzles I outlined in §1.2, which concern in particular the relationship between self-legislation and the source of the binding *authority* of the moral law. Did Kant mean to introduce the notion of moral self-legislation as a part of the Command Thesis, tying our subjection to the moral law directly to the commanding *activity* of the legislative will, or did he mean to defend something more along the lines of the Rationality Thesis, linking the moral law's authority for us to its basis in reason? Does either of these theses capture his conception of autonomy?⁵⁰ Even with the distinction between *Wille* and *Willkür* in hand, these questions remain, for they concern the precise meaning and function of the 'legislation' of the former. Setting out the basic structure of the Kantian will and the significance of its parts for freedom is of course important and is a necessary step in coming to terms with the meaning of autonomy, but this does not in itself resolve the tension between legislation and law in Kant's thinking.

1.5 *Lingering Objections*

The tension between the Command Thesis and the Rationality Thesis hinges on the question of whether we want to emphasise the legislative activity of the will or the

⁵⁰ To recap, the Command Thesis holds that the act of moral self-command constitutes the source of the obligation to act in accordance with the moral law (see p. 16). The Rationality Thesis, by contrast, holds that the moral law's binding authority is fully explained by the fact that it is a fundamental 'law of reason' (see p. 19).

universally binding character of moral requirement.⁵¹ The move towards focusing on the ‘self’ in self-legislation may be tempting for some to the degree that one aims to connect morality to individuality, authenticity, self-realisation and to a certain conception of freedom. If my will is the source of moral laws, then there might seem to be something very personal about my morality, something which links it inherently to me.⁵² Yet how can such a self-imposed law be said to *bind* me at all?⁵³ Worse still, how could a law which finds its source in *my* legislative will ever be shown to be valid for every rational being as such?

Some of Kant’s greatest supporters address this problem by attempting to reconcile a commitment both to the universal validity of moral laws and to the thought that their validity is grounded in an act of the will. On a cluster of quite popular antirealist,

⁵¹ As Wood puts it, this tension depends on whether we highlight the ‘*autos*’ (self) or the ‘*nomos*’ (law) in *autonomia*. See *Kantian Ethics* 106.

⁵² As will become clearer, I take this personalised approach to autonomy to be deeply mistaken. The concerns articulated in this section are meant to set the stage for the chapters to come.

⁵³ This worry of course echoes Anscombe’s concerns, as outlined above. In his comments on Christine Korsgaard’s interpretation of autonomy, G. A. Cohen links this worry to an argument articulated by Thomas Hobbes (see Korsgaard et al., *The Sources of Normativity*, ed. Onora O’Neill [Cambridge: Cambridge University Press, 2000], 167–70). As Hobbes writes, ‘the Sovereign of a Common-wealth, be it an Assembly, or one Man, is not Subject to the Civill Lawes. For having power to make, and repeale Lawes, he may when he pleaseth, free himselfe from that subjection, by repealing those Lawes that trouble him, and making of new; and consequently he was free before. For he is free, that can be free when he will: Nor is it possible for any person to be bound to himselfe; because he that can bind, can release; and therefore he that is bound to himselfe only, is not bound’ (*Leviathan* II.26.6.137–8); see *Leviathan*, ed. Richard Tuck, revised student ed. (1651; Cambridge: Cambridge University Press, 1996) 184. As we have seen, this worry is repeated by Kant himself in his discussion of duties to the self in the *Metaphysics of Morals* (6:417). As we will see in chapter three, given that he held a sufficiently rich understanding of self-legislation, this criticism does not apply to Kant. On its own, it actually begs the question against those who hold that the will is the source of its own obligations (see Timmermann, ‘Duties to the Self’ 515–16).

antirealist interpretations of Kant's ethics, for example, a guiding notion is that autonomy implies that the rational will could never be subject to an 'external' or 'independent' order of standards and values, which find their source outwith the volitional activity of rational agents. As John Rawls puts this idea,

... it suffices for heteronomy that first principles are founded on relations among objects the nature of which is not affected or determined by our conception of ourselves as reasonable and rational persons (possessing the powers of practical reason) and by our conception of the public role of moral principles in a possible realm of ends.⁵⁴

According to this general approach (which comes in many more specific varieties), a genuinely binding moral principle must be created or *constructed* by a legislative act of the will, via the use of certain privileged procedures (furnished by the familiar Kantian formulae, including the Formula of Universal Law).⁵⁵ Objective norms are not valid independently of our conceptions and activities, but are instead 'constituted by the activity, actual or ideal, of practical (human) reason itself.'⁵⁶ Because of the

⁵⁴ *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, MA: Harvard University press, 2000) 236.

⁵⁵ Though many constructivist interpreters of Kant focus on the Formula of Universal Law, Thomas E. Hill, Jr. offers a constructivist approach to the formula of the End in Itself. See 'Kantian Constructivism in Ethics,' *Ethics* 99 (1989): 752-770. Interestingly, constructivist views tend not to emphasise Kant's distinction between *Wille* and *Willkür* (see §§2.1 and 2.2).

⁵⁶ Rawls, *Political Liberalism* (New York: Columbia University Press, 1993) 99. As Rawls explains elsewhere, 'Kant's idea of autonomy requires that there exists no moral order prior to and independent of those conceptions that determine the form of the procedure that specifies the content of the duties of justice and of virtue. Heteronomy obtains not only when these first principles are fixed by our special psychological constitution, a psychology which distinguishes us from other persons, as in Hume, but also when first principles are fixed by an order of moral values grasped by rational intuition, as with Clarke's fitnesses of things or in Leibniz's hierarchy of perfections' (*Lectures* 237). Compare Christine Korsgaard: 'values are not discovered by intuition to be "out there" in the world' but are instead 'created' by human beings by 'the procedure of making laws for ourselves' (*The Sources of Normativity* 112; see

vital role of the procedural construction of moral norms and of rational agents in this activity, the significance of the ‘self’ (or ‘selves’) in self-legislation is guaranteed to be more than a euphemism: moral requirements are at base a species of positive law, existing and binding us ‘*because we legislate them.*’⁵⁷ Yet to the degree that the relevant procedures are not arbitrary and indeed are such that they yield *correct* answers to moral questions, we are at the same time permitted – or so the story goes – to hold these norms to be objectively valid for all rational beings.⁵⁸

We will be examining this move toward constructivism in further detail in chapter two. This explicit connection of the moral law and the wills of individual selves has been a very popular move – so much so that it seems that, in denying its plausibility,

also ‘Kant’s Analysis of Obligation: The Arguments of *Foundations* I,’ *The Monist* 72 [1989]: 311-340). Kantian constructivism of various stripes has been chiefly inspired by Rawls’s approach to Kant’s ethical thought: see in particular his ‘Kantian Constructivism in Moral Theory,’ *Journal of Philosophy* 77 (1980): 515-72. A very helpful account of the evolution of Rawls’s own constructivist views (views which he takes to be Kantian, if not exactly Kant’s) is given by Onora O’Neill in her ‘Constructivism in Rawls and Kant,’ *The Cambridge Companion to Rawls*, ed. Samuel Freedman (Cambridge: Cambridge University Press, 2002) 347-353. O’Neill traces Rawls’s account of what he takes to be Kant’s constructivism on pp. 353-57 of this text. Relatively similar constructivist interpretations of Kantian autonomy have been proposed by Andrews Reath (‘Autonomy of the Will as the Foundation of Morality’ and ‘Legislating the Moral Law,’ *Agency and Autonomy in Kant’s Moral Theory* 121-72 and 92-120) and J. B. Schneewind, ‘Natural Law, Scepticism, and Methods of Ethics,’ *Journal of the History of Ideas* 52 (1991): 289-308; and *The Invention of Autonomy: A History of Modern Philosophy* (Cambridge: Cambridge University Press, 1998) esp. 512-525. For a particularly rich and subtle version of Kantian constructivism, which avoids presenting autonomy as something personal and arbitrary, see Onora O’Neill’s ‘Constructivism in Rawls and Kant’ 357-363 and *Constructions of Reason: Explorations in Kant’s Practical Philosophy* (Cambridge: Cambridge University Press, 1989) ch. 11.

⁵⁷ Korsgaard, ‘Kant’s Analysis of Obligation: The Argument of *Foundations* I’ 331; my emphasis.

⁵⁸ See, e.g. Korsgaard: ‘there are answers to moral questions because there are correct procedures for arriving at them’ (*The Sources of Normativity* 38).

we risk robbing Kant's ethics of one of its most appealing features. Yet we might nonetheless worry about the prospect of deriving categorical necessity, the hallmark of Kant's moral law, from the volitions, commitments, or endorsements of individual rational agents. Surely categorical bindingness cannot depend on any act of *my* will, no matter which 'procedure' I follow in my willing, for such a move would seem to leave Kant's ethics open to the exact same charge that he levied against the theological voluntarists: that such a law can only ever bind my will via some mediate interest, but never solely from itself. The worry, then, is that the Command Thesis is little more than a cleverly disguised endorsement of the idea that the explanation of our subjection to moral requirement (and the only available reason for our compliance with it) lies in the fact that *we have created it*.

As we saw in §1.2, the retreat from the Command Thesis to the Rationality Thesis is not without its own problems. Perhaps most importantly, it risks portraying autonomy as an exaggerated metaphor: if the source of the moral law is really the mysterious faculty of pure reason (the same in each of us), then there would seem to be little force to the claim that the ultimate moral legislator is in any sense my *self*. To say that the moral law is a law of reason is, on the face of things, precisely to say that its authority is independent of any individual volitional act. But if this is the case, then we seem to lose the sense in which our acceptance of the moral law differs at all from our acceptance of church dogma or the edicts of the state, at least when it comes to the question of our ability to identify *ourselves* with its source. Indeed, it has seemed to some of Kant's critics that the moral law possesses something of a *dictatorial* character when it comes to the finite rational agent, such that Kant's ethics establishes a 'tyranny of reason' against those aspects of ourselves with which we most identify in our everyday dealings (our deep-seated desires, values, commitments and attach-

ments).⁵⁹ If reason functions as an internal despot, we may well wonder what appeal Kant's ethics of autonomy really has – and indeed, whether it actually differs from the previous rationalist positions on which Kant took himself to have drastically improved.

⁵⁹ This theme is developed in chapter six.

A WILL GIVING UNIVERSAL LAW

As we have seen, Kant proposed that ‘the idea of the will of every rational being as a *will giving universal law*’ (*G* 4:432) is the key to explaining the possibility of a categorical imperative. Although it is subject to the moral law, the rational will must be regarded as also giving that law – as in some sense legislating the command by which it is bound. In the *Groundwork*, Kant associates this idea with the thought that the will may regard itself as the ‘author’ of the law, an idea which has been taken by some of Kant’s interpreters to provide a key to understanding the troubling concept of moral self-legislation. As Kant writes:

‘... all maxims are rejected that are inconsistent with the will’s own giving of universal law. Hence the will is not merely subject to the law but subject to it in such a way that it must also be viewed *as self-legislating*, and just because of this as subject to the law (of which it can regard itself as the author) in the first place’ (*G* 4:431.19-24).¹

For many of Kant’s interpreters, the language of authorship of the moral law leads straightforwardly to the idea of the *creator* or *maker* of law, which would suggest that

¹ I have strayed from Gregor’s translation on a number of fronts in this passage.

Kant subscribes to a form of the Command Thesis.² Though this is a popular conclusion, I argue in this chapter that there are inherent difficulties in reconciling a particular version of this thesis with the universality and categoricity which Kant presumed to belong to morality. What is more, I suggest that Kant's language of authorship is much less straightforward than it first appears. Though it may be tempting to interpret Kant's claim about our status as authors of the law as implying this version of the Command Thesis, careful attention to Kant's use of the authorship metaphor reveals that there is an important alternative to consider. In clearing the space for this interpretative option, I hope to prepare the ground for the more positive arguments of chapter three.

2.1 *The Command Thesis and Radical Constructivism*

The key idea behind the Command Thesis is that we impose moral law onto ourselves, such that moral constraint can be viewed as the result of deliberative activity. Put differently, it suggests that the exercise or *activity* of the will is the source of the obligation to abide by morality's prescriptions. These core ideas are quite general, and as such the Command Thesis must be made more specific if it is to stand a

² Recall that, according to the Command Thesis, the moral law is a law which the rational agent commands to him- or herself, such that the act of command may be described as the source of the obligation to act in accordance with the law. In a recent study on autonomy in ethics, Charles Larmore quite explicitly equates Kant's position with a strong version of the Command Thesis. As he writes, 'In [Kant's] hands, [autonomy] did have a distinct meaning – namely, self-legislation: principles of thought and action acquire their authority by virtue of rational beings imposing them on themselves, instead of supposing them to be endowed with an independent validity to which they must simply assent' (*The Autonomy of Morality* [Cambridge: Cambridge University Press, 2008] 43). See also p. 109: 'Reason, Kant concluded, must be understood as self-legislating or "autonomous": in endorsing principles, it does not so much recognise as establish their right to regulate our conduct.'

chance of helping us to understand Kant's conception of self-legislation. What, for instance, lies behind the idea of creating moral obligation, and how might agents be conceived as engaging in such an activity? Does the Command Thesis imply that we are the source of the authority of the categorical imperative and its three variants, or that we are the source of the substantive (yet still general) practical laws that are determined via the application of the supreme principle of morality?³ Finally, is it part of this view that the power to confer normative authority onto a principle is in any way connected to the power to determine its content?

The first of these questions concerns the *nature* of the will's legislative activity, and in this regard it will be helpful to draw a parallel between the Command Thesis and the version of Kantian constructivism offered by Christine Korsgaard.⁴ Broadly speaking, Korsgaard follows in the tradition of John Rawls, according to whom an essential feature of Kantian constructivism is the claim that particular practical laws – substantive prescriptions and prohibitions – are to be viewed 'as specified by a procedure of construction,' the structure of which is linked fundamentally to our powers of practical reasoning and to the self-conception that accompanies them.⁵ As I men-

³ E.g. the unconditional command not to lie.

⁴ Versions of constructivism are sometimes offered as interpretations of Kant, sometimes as substantive positions in their own right. Given my purposes, I will be focusing on Korsgaard's own view, which she presents as a modified version of the Kantian account of the source of obligation.

⁵ *Lectures on the History of Moral Philosophy* 237. In its widest sense, the construction metaphor might be taken to describe any view that focuses on the moral law as providing a 'constructive' (positive, action-guiding) procedure for solving moral problems. This is the sense in which John Rawls first used the idea of *construction* in *A Theory of Justice*. Even here, Rawls contrasted ethical construction with intuitionism, which he viewed as being unable to settle practical disputes (and so as non-constructive) insofar as it offered a number of unranked principles (See *A Theory of Justice* [1971; Cambridge, MA: Harvard University

tioned in §1.5, this procedure is typically identified with the categorical imperative, and Korsgaard does not depart from this convention. What distinguishes her view from many other versions of constructivism, however, is her insistence on the idea that moral requirements are *generated by the exercise of a process of deliberation*, and so constitute its outcome. In this sense, the procedure of construction is meant to do more than allow us to *identify* the moral status of a proposed maxim.⁶ Rather, it expresses a reflective process by which individual agents *craft* or *fashion* law for themselves.⁷

On Korsgaard's interpretation, the categorical imperative is a 'test of reflection.' As an impulse to action presents itself to me, I test it to see whether it is a reason to act 'by asking whether the maxim of acting on it can be willed as a law.'⁸ This question is asked from the standpoint of a certain self-conception (or cluster of self-conceptions): where a proposed law is shown on reflection to be compatible with the identity I have adopted for myself, it is imparted normative force and functions as

Press, 1999] 34, 39-40, 49, 52; see also O'Neill, 'Constructivism vs. Contractualism,' *Ratio* 16 [2003]: 320).

⁶ Though he goes beyond this thought in stressing that the categorical imperative procedure has the power to generate binding reasons, Andrews Reath also emphasises the role of the categorical imperative as a criterion of moral permissibility: 'A constructivist account takes the FUL [formula of universal law] to be the basis of a procedure of moral deliberation (a 'procedure of construction') that is the final criterion of right in the sense that the outcome of applying the FUL, whatever it is, defines the moral status of a maxim' (*Agency and Autonomy* 169.n40).

⁷ Korsgaard labels her view a form of procedural realism, since she takes values to be constructed by a procedure: 'the procedure of making laws for ourselves' (*The Sources of Normativity* 112). Strictly speaking, on Korsgaard's view, not every maxim that we find we can will as a universal law will count as a 'maxim of duty.' It is only when a maxim fails the test, she writes, that we get a duty – the duty of adopting the opposite of what the failed maxim proposes. See 'Kant's Analysis of Obligation in *Foundations* I' 328-29.

⁸ *Sources* 89, 113.

a source of reasons. Where it is shown to be incompatible with an identity I have adopted for myself, I obligate myself to refrain from acting as it prescribes. Thus obligation derives, according to Korsgaard, i) from my having adopted a certain self-conception, and ii) from my having reflected in accordance with the categorical imperative from the standpoint afforded by this self-conception.⁹ We *confer* the status of law onto proposed maxims of action by engaging in this process of reflection, such that whilst it is true that in some sense the categorical imperative test allows us to discover that a maxim is fit to be a law, ‘the maxim isn’t a law until we will it, and in that sense create the resulting value.’¹⁰

Construction is a metaphor. It is linked with the idea of activity and typically stands in contrast to views which posit that the relevant norms or truths exist independently of human reasoning, so that the latter might be used to discover or intuit them. On less radical constructivist views, the claim is that facts about obligation are grounded in facts about the possible conclusions of procedures hypothetically carried out by rational agents: a view which is consistent with there being genuine obligations which are independent of the *actual* carrying out of a certain form of reasoning or deliberation. I have an obligation not to lie, on the moderate view, not because I have engaged in a certain course of reasoning, but because there are facts about what a rational agent *would* choose if she were to deliberate in accordance with the procedure

⁹ As Korsgaard puts her idea, ‘the principle of choice [maxim] derives its ability to confer normativity on [a given] incentive from the way that the agent identifies herself ... Normativity derives from our self-conceptions’ (*Sources* 248-9).

¹⁰ *Sources* 112. This argument is summarised on pp. 112-113. Note that, on Korsgaard’s view, certain of our maxims (the subjective principles on which we act) will be made objectively normative (actual laws with normative force) by our reflective endorsement. What we actually ‘legislate,’ on this view, are maxims.

in question.¹¹ But a more radical constructivist reading of Kant, such as that proposed by Korsgaard, takes him to be suggesting that human beings must create substantive principles of justice and morality by actually carrying out the categorical imperative procedure.¹² The idea is not merely that there are procedures by which we can demonstrate that certain principles are justified and binding. The principal idea is, rather, that there is a procedure the observance of which confers validity or justification onto certain principles, such that those agents who undertake that chain of reasoning can be viewed as the source of their normative force. It is true that before the use of the relevant procedure, we cannot know any of our moral obligations. But this is precisely because, before the use of the procedure, there simply *are no* moral obligations to be known.¹³

Does our volitional activity have any influence on the content or validity of the ‘procedure of construction’ itself, on Korsgaard’s view? Is this a principle that we can be said equally to ‘construct’? Here, some constructivists maintain that the answer must be ‘no.’ Rawls, for instance, proposes that while the substantive content of moral

¹¹ In casting things in terms of what the fully rational agent would do, this version of Kantianism would contain strong parallels to a virtue-theoretical way of thinking about the content of morality. For an insightful exploration of some parallels between Kantian and Aristotelian ethics, see Robert Louden, *Morality and Moral Theory* (Oxford: Oxford University Press, 1992), esp. ch. 6.

¹² This contrast is nicely articulated by Eric Watkins and William Fitzpatrick, ‘O’Neill and Korsgaard on the Construction of Normativity,’ *The Journal of Value Inquiry* 36 (2002): 358. Watkins and Fitzpatrick also label Korsgaard’s view ‘radical.’

¹³ Because they hold that there are non-realist, justified, objective, action-guiding ethical prescriptions, most constructivist views are highly ambitious. See O’Neill, ‘Constructivism in Rawls and Kant’ 348-9. O’Neill’s own constructivist view avoids the language of creation and conferral of obligation, focusing instead on the notions of justification and vindication. For this reason, I will not be giving very much attention to her view here, where the idea is to work through possible specifications of the Command Thesis.

doctrine is indeed constructed, the same cannot be said for the categorical imperative procedure. The latter is simply ‘laid out’ and can be identified by considering the implicit structures at work in our conception of the reasoning of free and equal, reasonable and rational, persons.¹⁴ In a similar vein, Korsgaard suggests that the categorical imperative ‘does not impose any external constraint on the free will’s activities, but simply arises from the nature of the will.’¹⁵ On her view, the categorical imperative sets out what a free will must do in order to continue being what it is, such that *not* following this principle renders one unable to continue to view oneself as the free

¹⁴ *Lectures on the History of Moral Philosophy* 239-40. Though her constructivism differs in important ways from those offered by Rawls, Reath and Korsgaard, it is useful also to consider O’Neill’s suggestion that human beings can be said to construct a common plan or procedure for the determination of substantive principles of behaviour. On her view, there is no independent order of reason that imposes a common procedure of this sort: ‘As Kant depicts the matter, the very predicament of a plurality of uncoordinated agents is all we can presuppose in trying to identify principles of practical reason: it is because reason’s authority is *not* given that it must be instituted or constituted – constructed – by human agents’ (‘Constructivism in Rawls and Kant’ 357-8). If human beings are to organise their thinking together and to find common principles of behaviour, they need to construct some common authority. How is this to be done? O’Neill answers this query by suggesting that, ‘[s]ince all they have in common is their lack of a given “plan of reason,” all that they can do is to refuse to treat any of the various faiths and beliefs, traditions and norms, claims and propositions they variously adhere to as having unrestricted authority for organising thinking and doing. However, those who do not regard any specific faith or beliefs, tradition or norms, claims or propositions as having an unrestricted authority for organising thinking and doing in effect adopt the overarching principle of thinking and acting only on principles which they regard as open to, and followable by, all’ (‘Constructivism in Rawls and Kant’ 358).

¹⁵ *Sources* 98. A parallel thought is articulated by Reath, who argues that, whereas substantive practical laws are generated via our employment of the universality test, the categorical imperative is ‘given’ by the will insofar as it ‘emerges from the very nature of rational volition’ and that this explains the sense in which we can be said to ‘legislate’ the categorical imperative (along with substantive moral principles) (*Agency and Autonomy* 99). For an important criticism of this view, see Thomas E. Hill, Jr., ‘Legislating the Moral Law and Taking One’s Choices to be Good,’ *Philosophical Books* 49 (2008): 102-3. On Reath’s view, the moral law is not a law that we impose on ourselves via some act of volition. This distinguishes him from Korsgaard (‘Autonomy, Taking One’s Choices to be Good, and Practical Law: Replies to Critics,’ *Philosophical Books* 49 [2008]: 126-29.)

cause of one's actions.¹⁶ What is more, we must, as reflective rational agents, maintain a conception of ourselves as free in order to be able to act at all. Because this is a self-conception which we necessarily adopt in acting at all – because we must identify with this self-image – we are committed to willing in accordance with the categorical imperative.¹⁷ With this last move, however, the story is complicated, for we now see that for Korsgaard there is a sense in which our own volitional activity can be said to ground at least the normative force – though not the content – of the categorical imperative. According to her interpretation, principles derive their ability 'to confer normativity on [a given] incentive from the way the agent identifies herself'; normativity, even at the level of this basic principle, derives from our adoption of a self-conception – in this case, our conception of ourselves as free agents who cause our own actions.¹⁸ Although we are not free arbitrarily to determine the content of the categorical imperative, its bindingness (its force as an obligation) is ultimately grounded i) in the fact that it describes what a free will must do in order to continue being a free will; and ii) in the fact that we *must* adopt a conception of ourselves as free in order to act at all.¹⁹ Because it is ultimately our actual identification with a self-conception that does the normative work, here, there is a fundamental sense in which our own activity can be said to ground the categorical imperative itself.

¹⁶ *Sources* 98, 235-7. Korsgaard presents this argument as an interpretation of the opening argument of *Groundwork* III.

¹⁷ See *Sources* 90-113.

¹⁸ *Sources* 248-9. On Korsgaard's view, '[a]n obligation always takes the form of a reaction against a threat of a loss of identity' (102). Here, Korsgaard is explaining her own view, but she suggests that this particular aspect (or one that is very similar) can be found in Kant's thoughts on autonomy as well (see p. 249). Korsgaard's complete argument is developed in chapters three and four of *Sources*.

¹⁹ See *Sources* 257. Although I will not push this criticism here, both of these claims are suspect. The first prong of this argument (the idea that the moral law describes what a will must do in order to will at all) will be discussed in some detail in §3.1.

So much for the basic contours of Korsgaard's explanation of the source of normativity. On her view, however, we must delve deeper if we are to account for the particular obligations that are to count as *moral*. Departing from Kant's own distinction between the categorical imperative and the moral law, Korsgaard proposes that we think of the former as telling us to act only on maxims that can be willed as laws, the latter as telling us to 'act only on maxims that all rational beings could agree to act on together in a workable cooperative system.'²⁰ Where the categorical imperative directs us to act on *any* law that is consistent with our identity (on any universal principle), it does not on its own, she argues, get us to moral constraint. We enact laws (and reasons) from a first-person standpoint in which we identify ourselves as certain kinds of people, and whether a maxim can serve as a law for us will depend 'upon the way that we think of our identities,' such that different laws will hold for egoists, mothers, and mafiosi.²¹ If an agent is to bind herself to substantive *moral* obligations via reflective endorsement, she must identify herself simply as a human being – or as Korsgaard puts it, as a 'Citizen of the Kingdom of Ends.'

Korsgaard seeks to preserve the universality and necessity of moral obligation by arguing that we *must* adopt a conception of ourselves as citizens of the kingdom of ends. Her argument involves the following five steps: i) human beings must act for reasons; ii) without a conception of her practical identity, an agent cannot have a reason to act; she must therefore have some conception of her practical identity;²² iii) an agent can adopt a practical identity only if she adopts her humanity as a practical

²⁰ *Sources* 99.

²¹ *Sources* 113. See also p. 107: 'Autonomy is commanding yourself to do what you think it would be a good idea to do, but that in turn depends on who you think you are.'

²² *Sources* 120.

identity; iv) to adopt one's humanity as a practical identity is to value one's own humanity; v) valuing one's own humanity involves valuing the humanity in others as well, and this gets us to moral obligations.²³ Here, the general strategy is parallel to that by which Korsgaard argues for the normativity of practical laws in general. Moral obligations derive their binding force from the fact that they are the result of reflection in accordance with the categorical imperative from the standpoint afforded by a certain self-conception. Unlike certain of our identities, however, the particular self-conception in question is one that every agent by necessity must adopt (or so the story goes) if she is to act for reasons at all.

2.2 *Reason and Reasoning*

When it comes to constructivism more generally, the procedural generation of substantive moral norms need not be conceived as an exclusively individualistic matter.²⁴ On Rawls's view, for instance, constructivism is meant to explain how the content of moral doctrine will look once it is revealed after reflection according to the relevant procedure. The reflection here is not best understood as individualistic reflection or as 'perfect reflection at the end of time'; rather, it is identified with 'such increasing critical reflection as might be achieved by a tradition of thought from one generation to the next,' such that there should be a continual progress in formulating the doctrine

²³ *Sources* 121.

²⁴ Of course, there is nothing wrong in principle with a view according to which much of the stuff of morality (e.g. coming to know moral norms, being able to act on them) is an individual matter.

in its entirety.²⁵ With this said, however, other constructivist accounts do tend to focus on the legislative authority of a process of reflection undertaken by individuals. As Korsgaard puts her view, the relevant idea behind Kantian constructivism is that ‘[n]othing except my own will can make a law normative *for me*.’ By employing the universality ‘test,’ my will commits me to certain maxims, and in endorsing certain principles as carrying normative force, I *myself* bring moral laws into existence.²⁶

Yet even if the focus is on individual reflection, radical constructivists such as Korsgaard need not be viewed as allowing for the *arbitrary* or *discretionary* creation of moral obligations, which leads us to a second worry voiced in chapter one. On Korsgaard’s account, individual agents are in the first place to be viewed as the ultimate source of the obligation to comply with particular moral principles, transforming normatively neutral propositions into authoritative laws:

The source of obligation is a legislator. The realist objection – that we need to explain why we must obey that legislator – has been answered, for this is a legislator whose authority is beyond question and does not need to be established. It is the authority of your

²⁵ *Lectures on the History of Moral Philosophy* 274. To be clear, ‘progress’ on Rawls’s view ought not to be understood as becoming better at grasping an independent set of truths. As he explains in his lectures, ‘[r]ational intuitionism says: the procedure is correct because following it correctly usually gives the correct (independently given) result. Constructivism says: the result is correct because it issues from the correct reasonable and rational procedure correctly followed’ (242).

²⁶ ‘Kant’s Analysis of Obligation in *Foundations* I’ 330-331. Crucially, this universality test is not a test of knowledge or truth, for Korsgaard. It is a test of ‘reflective endorsement,’ by which the individual will actively commits to the suitability and validity of certain principles, and by which the normativity of these principles is generated. An individualistic interpretation somewhat along these lines has also been offered by Robert Paul Wolff, who argues that ‘I must be the source of unconditional obligation ... [Kant believes] (I think correctly) that rational agents are bound to substantive policies only insofar as they have freely chosen those policies’ (*The Autonomy of Reason: A Commentary on Kant’s Groundwork of the Metaphysics of Morals* [Gloucester, MA: Peter Smith, 1986] 181). For an even more radically individualistic conception of self-legislation, see Rüdiger Bittner, *What Reason Demands* ch. 5.

own mind and will ... It is not the bare fact that it would be a good idea to perform a certain action that obligates us to perform it. It is the bare fact that we *command ourselves* to do what we find it would be a good idea to do.²⁷

There is a sense in which Korsgaard would be happy to say that the content of *substantive* moral requirements, and not merely their binding force, is generated or constructed by the process of deliberation, to the degree that it is determined completely by the application of the constructive procedure to a practical situation or maxim.²⁸ Since one cannot say in advance whether a proposed maxim will pass the universality test – since, on Korsgaard’s view, there *is* no answer to this question independently of its application – the content of our obligations is wholly settled by this procedure. This is not to say, however, that moral agents have any immediate discretion over the content of their substantive obligations, for the question of whether a maxim is able to pass the universality test is not a matter of choice.²⁹ As a result, radical constructivists need not necessarily equate construction with arbitrary creation at the level of the content of practical laws. The generation of substantive principles can only be effected by the following of a deliberative procedure, and so long as the same maxims have been tested, the content of moral laws simply cannot vary capriciously from

²⁷ *Sources* 104-5.

²⁸ See, e.g. Reath *Agency and Autonomy* 111. This reply assumes, of course, that the age-old concern that the categorical imperative ‘procedure’ is too formal a tool to generate substantive practical implications (as voiced, for example, by Hegel and Mill) is unfounded. For a convincing defence of the practicality of the categorical imperative, see O’Neill, *Constructions of Reason* Part II.

²⁹ *Agency and Autonomy* 147. What I mean, here, is that the content of our obligations (what obligations we have) will depend on the principles to which we are committed (and in turn, on the self-conceptions with which we identify – e.g. free cause, mother, mafioso), and in that sense it is not arbitrarily determined. With this said, Korsgaard is nonetheless vulnerable to the charge that she has not shown that we *must* take up any particular practical identity (from the standpoint of which certain maxims must be endorsed), and to the degree that identities may be chosen or constructed at will, the charge of arbitrariness might well infect her position after all.

agent to agent, from tradition to tradition. Similarly, a more detailed look at constructivism reveals that we need not worry that it allows for the possibility of arbitrarily ‘unbinding’ oneself from the moral laws one has constructed. The categorical imperative procedure is a chain of reasoning that confers necessity onto substantive principles, generating obligation where there was no obligation before. The legislative activity of an individual agent *must* follow this procedure if she is to be able to view herself as the cause of her actions. As such, she cannot release herself from the commitments to which her reasoning leads without abandoning this self-conception. Thus *if* Korsgaard is correct about the necessity of this self-image for rational agency (itself a separate question), the worry that individual agents might be able easily to release themselves from the obligations that they at one point legitimately faced becomes much less threatening.³⁰

With these observations noted, however, fundamental worries are likely to persist. First, if we cannot guarantee that all agents bring the same maxims to the universality test – if we cannot guarantee that they engage in this constructive procedure at all – then we cannot guarantee the universal and unconditional validity of moral requirements for all (imperfect, morally responsible) rational beings. In effect, the radical constructivist owes us a story about why we should suppose that the individual moral agent or ‘tradition’ of moral thought actually undertakes the relevant procedure, applied to the same proposed maxims. If she cannot provide such an account, then little sense can be made of the very central Kantian notion that all imperfect rational agents are universally bound to observe the same moral duties. Now given a plausible story about the uniformity of human nature, some kind of response to this worry might in

³⁰ As Korsgaard puts this idea, we cannot change the law without changing our minds; because ‘we cannot change our minds about just anything,’ escape from our own laws becomes more difficult (*Sources* 234).

the end be crafted by those who wish to defend the above approach. If it could be shown empirically that we have grounds for thinking that human beings tend to bring the same *kinds* of general maxims to the table, we would have reason to suppose that Korsgaard escapes this charge.³¹ Even if this could be shown, however, a second objection hinges on the thought that it is unlikely that constructivists could ever give us reason to think that every rational agent as such actually undertakes the relevant process of reasoning, and so generates practical laws at all. Korsgaard, though optimistic on this front, admits that since obligation is constructed from the first-person perspective, ‘the obligatory is like the visible: it depends on how much of the light of reflection is on.’³² To the degree that an agent does not in fact identify as a member of the kingdom of ends or does not in fact reason in accordance with the categorical imperative, she is not bound by the conclusions that would have followed from that perspective. Yet where there is reason to doubt that every agent actually does reason in accordance with the categorical imperative, there is reason to doubt the adequacy of this view when it comes to accommodating the universal obligatoriness of practical laws.³³

Third, it seems difficult on this view to explain how agents can have duties that they fail to recognize. If substantive obligations are generated via my actively legislating maxims as law, the possibility of violating our legislated maxims becomes difficult to explain. According to Korsgaard, normativity is created by the reflective endorsement

³¹ At various points in Kant’s writings, the uniformity of (empirical) human nature is a recurring theme.

³² *Sources* 257.

³³ As it happens, Kant himself seemed confident that the voice of conscience functions as a powerful background prompt, urging us diligently to use the categorical imperative in the selection of maxims. See, e.g. *Rel* 6:185-87.

of maxims. As a result, if room is to be made for moral wrongdoing, there must be a distinction between the reflective endorsement of a maxim as morally appropriate and its actual adoption as a subjective principle of action (i.e. *acting* on it). But then how are we to explain why an agent would reflectively endorse a principle or reason without actually acting on it?

A final objection concerns what this view forces us to say about moral obligation in general. If the normative pull of obligation derives from our having adopted a certain self-conception and from our having reasoned in accordance with the categorical imperative from that perspective, then moral obligations become *contingent* rather than necessary. Korsgaard hopes to preserve the necessity of duty by arguing that we ‘must’ adopt a conception of ourselves as members of the kingdom of ends if we are to value anything at all. Yet even if it could be shown that every rational agent must adopt this self-conception in order to be able to act for reasons – and this, of course, raises difficult questions of its own – what kind of ‘must’ would this be? Korsgaard’s presentation of this idea suggests that this would be an empirical, psychological fact (a fact about the reflective structure of human consciousness). To the extent that this is so, however, what this ‘must’ actually describes is empirical generality rather than genuine necessity. Our psychology could have been otherwise, which means that the validity of the moral law for us is a contingent fact on her view. What is more, even *if* an *a priori* ‘must’ could be established, Korsgaard’s position renders the authority of the moral law conditional on our (supposedly ‘necessary’) need to preserve our practical identity. Since it is central to her view that the normativity of principles derives ultimately from the normativity of self-conceptions and the importance of preserving them,³⁴ Korsgaard’s moral law can only ever be a conditionally valid principle (con-

³⁴ See, e.g. *Sources* 249.

ditional, that is, on the value or the normative pull of the relevant self-conceptions). It might turn out to be true that the condition she picks out is always met; yet even where this is so neither Korsgaard's moral law nor the imperative she calls 'categorical' genuinely possesses the core features of necessity and categoricity.³⁵

Perhaps in response to these worries, Andrews Reath departs from Korsgaard's radical constructivism and suggests that Kant is best read as having proposed that the authority of a (substantive) practical law 'comes from its having the form of law, or, what amounts to the same thing, from the reasoning that determines that the relevant maxim can or cannot be willed as a universal law without inconsistency.'³⁶ In some instances, I will *actually* engage in reasoning fully in accord with the categorical imperative, in which case it is my very own reasoning which both confers validity on my conclusions and binds me to follow them. Even when I do not actually engage in this process of reasoning, however, Reath's rather striking claim is that so long as this chain of reasoning is available to me, I retain the very same *capacity* to carry out the reasoning that makes a principle a law, such that there is no relevant difference that justifies denying me the status of legislator.³⁷ On Reath's view, the formula of universal law captures the reasoning that gives a principle its moral status; yet given the lack of relevant difference between carrying out this reasoning and possessing the capacity to do so, the moral requirements that would be generated from its actual application apply to all rational beings for whom its performance is a possibility. As

³⁵ This type of criticism is developed in John Skorupski, 'Rescuing Obligation,' *European Journal of Philosophy* 6 (1998): 335-55.

³⁶ *Agency and Autonomy* 140. As Reath goes on to say, 'an agent subject to a practical law is bound to the law, and given sufficient reason to comply, by the reasoning that makes it a law.'

³⁷ *Agency and Autonomy* 142.

such, Reath's particular brand of constructivism does not suppose that one is subject to a moral requirement only if one has actually carried out the relevant form of reasoning:

It holds rather that a chain of reasoning that establishes the requirement must be available to the agent, and accordingly that the agent have the capacity to understand and be moved by this reasoning, whether or not the agent actually carries this reasoning out.³⁸

There is thus a crucial difference between moral law and civil law: where the latter is in force as obligatory only once it has actually been enacted by a legislature, the former is binding upon an agent so long as there is a process of reasoning *available* to her that establishes its validity.

Reath's modification is desirable, to the extent that constructivists wish to rescue the possibility of having duties that we do not recognise. Perhaps even more significantly, though, its adoption would allow constructivists to preserve the unconditional and universal character of duty – elements which would be lost if we were to maintain that moral requirements apply only to those who actually perform the relevant process of reasoning.³⁹ With this said, however, the accommodation of these important features of morality comes at a hefty price, from the constructivist point of view. Few Kantians would object to the notion that, in order for a moral principle to apply to me with the force of law, I must be able to reach a reasoned understanding of its validity and to be motivated to act in accordance with its prescriptions (even if this capacity is never actualised). Yet in divorcing the idea of legislation in Kant's central metaphor from the idea of the actual activity of the will – in associating it, in most

³⁸ *Agency and Autonomy* 143.

³⁹ Reath, *Agency and Autonomy* 144.

cases, with the notion of a merely *available* chain of reasoning – Reath renders obscure the extent to which actual *agents* can be viewed as generating the moral laws to which they are subject.⁴⁰ In effect, Reath’s modification shifts law-conferring power from agents who are reasoning in a certain way to a universally available process of deliberation, in which particular agents may or may not engage. This available process of reasoning, even in the absence of actual *reasoners*, is sufficient to establish a principle as universally binding.⁴¹ Yet to say this is just to say that certain principles are valid for all rational agents (agents, that is, for whom the chain of reasoning specified by the categorical imperative is available) *independently* of their actual volitions and exercises of thought. Reath does leave room for the idea that those who do actually engage in this reasoning can be viewed as giving moral law in an extended sense. But presumably even before rational agents engage in this activity, they possess the capacity to follow the relevant chain of thought, and so are under obligation to comply with that conclusion prior to any actual reasoning they might carry

⁴⁰ To be clear, that agents generate the laws to which they are subject is something that Reath thinks Kant is committed to, not something that I particularly wish to see defended.

⁴¹ ‘The FUL is a kind of legislative process that confers the status of law on a principle ... [it] is not a way of determining whether a maxim satisfies some further substantive principle, but rather is the procedure that makes a principle a law; it lays out the reasoning that gives a principle its moral status’ (*Agency and Autonomy* 143). See also Reath’s synopsis of his view in ‘Autonomy, Taking One’s Choices to be Good, and Practical Law: Replies to Critics’: ‘I believe that Kant’s idea is that since a practical law applies unconditionally, its authority cannot be based on appeal to contingent interests, but instead must come from the reasoning (deliberative procedure) that makes it a law. In other words, sufficient reasons to accept the principle are given by whether it can be willed as universal law through the FUL. (It is important here that universalisability *confers* normative status on proposed principles of action.) That means that an agent is bound to the law by the reasoning that makes it a law. In holding that an agent is bound in this way, we presuppose that the agent has the capacity to carry out the deliberative procedure that makes the principle a law. But an agent who can employ that deliberative procedure has the same capacities as would be required of its legislator. I want to say that agents with this rational capacity can be regarded as a kind of legislator and that Kant’s references to giving law through one’s will are best understood in this light’ (128).

out. Given this shift, we might well ask what separates this view from traditional forms of realism, according to which the moral law is to be understood as a rational constraint which is valid independently of any particular exercise of the will and which individual agents may or may not recognise in deliberation.⁴²

Moreover, and more significantly for our purposes, it is crucial to note that this modified constructivist view no longer captures the principal idea behind the Command Thesis, which, if we recall, is the idea that the validity of the moral law finds its source in an *act* of command on the part of the rational will (or rational agents). This is not in itself an objection to the approach, since constructivism need not entail anything like the Command Thesis as I have formulated it. Nevertheless, what Reath's consideration of self-legislation would seem to illustrate is that there are inherent difficulties in reconciling the central features of moral requirement on any recognizably Kantian view with the idea that the volitional or rational activity of individual agents constitutes the ultimate source of moral obligation. Reath moves towards the idea of available chains of reasoning – and so of *possible* rational activity – in a move to rescue these principal features. Insofar as his version of constructivism succeeds on this front, it constitutes an improvement on Korsgaard's more straightforward interpretation. Yet Reath's approach is able to secure what it does precisely at the cost of divorcing itself from the Command Thesis, with the result that the shortcomings of the latter are thrown into stark relief. In searching for a plausible version of construc-

⁴² This type of worry is raised by Patrick Kain ('Self-Legislation in Kant's Moral Philosophy,' *Archiv für Geschichte der Philosophie* 86 [2004]: 298).

tivism – one which secures the universality and unconditionality of moral requirement – we seem to have left the thesis we were hoping to specify far behind us.⁴³

2.3 *The Law and the Good*

Whether or not such problems plague the Command Thesis, we might nevertheless ask whether there is compelling evidence for this reading in Kant's ethical writings. On this front, two passages in particular stand out as promising.⁴⁴ The first can be found in the second *Critique* and concerns the failure of past moral theories to account for the categoricity of moral requirements. Traditionally, constructivist views in general have tended to rely on Kant's descriptions of heteronomous ethical theories as a means of supporting the notion that the authority of moral laws must be a human creation. As we have already seen, Kant criticised previous moral theories for supposing that the human being was merely *subject* to the moral law. According to Kant's criticism, all such views entailed that the law

... had to carry with it some interest by way of attraction or constraint, since it did not as the law arise from *his will*; in order to conform with the law, his will had instead to be constrained by *something else* to act in a certain way.⁴⁵ By this quite necessary consequence, however, all the labour to find a supreme ground of duty was irretrievably lost. For, one never arrived at duty but instead at necessity of an action from a certain interest. This might be one's own or another's interest. But then the imperative had to

⁴³ This is not to say, of course, that there are no plausible versions of constructivism. It is to say, rather, that if there are any then they are likely not to represent the Command Thesis.

⁴⁴ As will become clear, I do not actually think that these elements of Kant's doctrine do in fact support a constructivist reading of moral self-legislation.

⁴⁵ '... sondern dieser gesetzmäßig von etwas andern genötigt wurde, auf gewisse Weise zu handeln.'

turn out always conditional and could not be fit for a moral command. (*G* 4:433)

Because such theories presuppose some ‘external’ attraction or constraint as the determining ground of the will, none is able to account for the categoricity of moral requirement. Any law which depends on *something else* for its normative force and for its motivational pull is merely conditionally valid.

In the relevant passage from the second *Critique*, Kant develops this theme by casting his criticism of previous ethical theories in terms of their prioritisation of the *good* over the *law*. For Kant, good and evil are best conceived as *objects* of practical reason: as possible (real) effects of our free volition (*KpV* 5:57-58). They are thus concepts that represent a state of affairs which might be brought about as a result of our agency. Now as far as things go, there is nothing wrong in speaking about an object of practical reason *per se*, and Kant is certainly not denying that the concepts of good and evil ought to play a part in moral theory. What he wants to make clear, however, is that the categorical nature of morality makes it such that the concept of the good (object) must be *derived from* the moral law, and not vice versa. As he writes, ‘*the concept of good and evil must not be determined before the moral law (for which, as it would seem, this concept would have to be made the basis) but only ... after it and by means of it*’ (*KpV* 5:62.37-63.4).⁴⁶

⁴⁶ This passage is cited in Schneewind (‘Natural Law, Scepticism, and Methods of Ethics’ 305) and in Rawls (*Lectures on the History of Moral Philosophy* 227). C.f. *G* 4:441.3-11: ‘If the will seeks the law that is to determine it *anywhere else* than in the fitness of its maxims for its own giving of universal law – consequently if, in going beyond itself, it seeks this law in a property of any of its objects – heteronomy always results. The will in that case does not give itself the law; instead the object, by means of its relation to the will, gives the law to it. This relation, whether it rests on inclination or on representations of reason, lets only hypo-

On Kant's view, if the concept of the good were not derived from the moral law but instead made to serve as its basis, then pleasure would become the ultimate practical criterion (the criterion of good or evil would consist only in 'the agreement of the object with our feeling of pleasure or displeasure' [*KpV* 5:63.16-18]). There is a question as to why this should be so. An answer is provided at *KpV* 5:58-9, where Kant suggests that a state of affairs can only determine the will by means of a feeling of gratification connected to the thought of the realisation of the object. The objects we encounter in the world (cups of coffee, another's neediness) are *empirical* objects, and, according to Kant, the only way to bind an empirical object to my will is by means of an equally empirical incentive provided by inclination: in particular, a feeling of pleasure or displeasure. This passage spells out a central assumption on Kant's part concerning human nature – one with which we might disagree. Once this perhaps dubious assumption is taken on board, however, Kant's claim concerning the necessity of pleasure or displeasure when it comes to choice based on a conception of a good or bad object is made more intelligible.⁴⁷

Since it is impossible to determine *a priori* which representations of states of affairs will bring pleasure, putting the good before the law would thus allow for empirical conditions for a moral law and, as such, would preclude the universality and necessity of moral requirements.⁴⁸ The moral law would constitute a mere *means* to the reali-

thetical imperatives becomes possible: "I ought to do something *because I will something else*".'

⁴⁷ In addition to spelling out this central assumption about human nature, Kant can be read in this passage as making a second point, employing what is close to Moore's open question argument. As he says, those people who want to put the object first think that goodness is pleasure and displeasure, but we usually make a distinction between the pleasurable/painful and the good/bad. There are different criteria for the application of these terms.

⁴⁸ There are a few controversial steps, here, which for our purposes I will ignore.

sation of some good and, ultimately, would be in the service of realising pleasure. Because of this, it would not possess the kind of necessity that Kant takes it to have, for the fact that x is a good means to y can only ever be contingent.⁴⁹ In addition, were it empirically grounded, the moral law might not apply to all rational agents, for the receptivity to feeling pleasure at the representation of the relevant good cannot be assumed *a priori* to apply to all (indeed, it could not by definition apply to infinite rational beings such as God).⁵⁰ Precisely as a result of this, Kant proposes that if we are to maintain the hypothesis that there are categorically binding practical laws, we must also presume that ‘instead of the concept of the good as an object determining and making possible the moral law, it is on the contrary the moral law that first determines and makes possible the concept of the good, insofar as it deserves this name absolutely’ (*KpV* 5:64.2-5). All previous moral theories sought an object of the will in order to ground the moral law, which rendered the latter only a mediate determining ground of the will (i.e. determining the will only by means of its object). Whether they placed this object in happiness, perfection, moral feeling, or harmony with the will of God, ‘their principle was in every case heteronomy and they had to come unavoidably upon empirical conditions for a moral law...’ (*KpV* 5:64.18-19).

Kant’s lesson is thus that we must begin with a conception of the moral law, and only then may we specify a conception of the good. Kant has thus excluded one possible option when it comes to the source of the authority of the moral law: the option of grounding it in some independently identified good state of affairs. This is an observation which some constructivists find highly important: for it would seem to some to

⁴⁹ Assuming that we can agree with Kant on the point that pleasure is not unconditionally good, this would also entail that morality (its value reduced, on this picture, to its contribution to the realisation of pleasure) is not unconditionally good.

⁵⁰ See again *KpV* 5:58-9.

provide evidence for the notion that the will itself must create moral laws. What other alternative to the good as a grounding for morality could Kant have conceived? If a law does not derive its status from its relation to something outside of the will, it is more than tempting to conclude that the will *itself* creates this necessity.⁵¹

If we examine this passage from the second *Critique* carefully, however, it becomes clear that Kant rejects the alternatives he lists precisely because they are too entangled with empirical facts to allow for the necessity and universality that Kant presumes to be characteristic of morality.⁵² The problem with grounding the moral law in an independent conception of a good *object* lies in the fact that this renders the ultimate criterion of right and wrong the agreement of compliance with our feeling of pleasure or displeasure. This is precisely because the agreement of an object (or state of affairs) with the feeling of pleasure is presented here as the ultimate criterion of non-moral goodness, which on the rejected views is to serve as the basis for the moral law. But this does not imply that the same fate must belong to *any* view according to which the moral law is valid for us without having been created by individual agents. Kant's discussion on the primacy of the law does indeed foreclose one class of options, but it does this for very specific reasons. So long as it avoids grounding morality in empirical (and so contingent) fact, a view according to which the moral law

⁵¹ Rawls diagnoses Kant's worry as follows: 'Kant thinks that once we start from the good as a prior and independently given object, the moral conception must be heteronomous. This is because in this case pure practical reason is not, as it should be, its own sovereign authority as the supreme maker of law ... Heteronomy means precisely this lack of sovereign authority' (*Lectures on the History of Moral Philosophy* 267). For more on this theme, see Stephen Darwall, *The Second Person Standpoint: Morality, Respect, and Accountability* (Cambridge, MA: Harvard University Press, 2006) chs. 9 and 11.

⁵² See Ameriks, *Interpreting Kant's Critiques* 270-71.

is valid independently of the creative activity of individual agents would seem to present a genuine alternative.

2.4 *The Two Faces of Legislation*

Though this first supposed bit of textual evidence for the Command Thesis is in fact inconclusive, a perhaps more promising passage from the *Metaphysics of Morals* has likewise been highlighted in the recent literature as having something important to say on this front. Here, Kant again appeals to the notion that the rational will is in some sense the ‘author’ of the moral law,⁵³ only unlike the passage from the *Groundwork*, in which this notion most famously appears, this passage distinguishes two senses in which one might stand in such a relation to a law:

A (morally practical) *law* is a proposition that contains a categorical imperative (a command). The one who commands (*imperans*) through a law is a lawgiver (*legislator*). He is the author (*auctor*) of the obligation in accordance with the law, but not always the author of the law. In the latter case the law would be a positive (contingent) and chosen [*willkürlich*] law. A law that binds us a priori and unconditionally by our own reason can also be expressed as proceeding from the will of a supreme lawgiver, that is, one who has only rights and no duties (hence from the divine will); but this signifies only the idea of a moral being whose will is law for everyone, without his being thought of as the author of the law. (6:227.10-20)

⁵³ I first mention this passage on p. 41. If we recall, this is where Kant suggests that ‘the will is not merely subject to the law but subject to it in such a way that it must also be viewed *as self-legislating*, and just because of this as subject to the law (of which it can regard itself as the author) in the first place’’ (*G* 4:431.21-24). See also *G* 4:448.17-18: ‘Reason must regard herself as the author of her principles independently of alien influences.’

Kant here suggests that a lawgiver can be the author of the obligation to act in accordance with a law without also being the author of the law. As Kant describes things, not even God himself can be viewed as the author of the moral law, for this would render the law contingent.⁵⁴ When we look at this passage in isolation, however, Kant leaves us rather in the dark concerning just what this notion – and the parallel notion of commanding through a law – amounts to. In what sense can the moral law be said to ‘proceed’ from someone’s will without his being its author, and how does this relate to ‘authoring’ the obligation to act in accordance with the law?

In his thoughts on this passage, Andrews Reath suggests that, by ‘author of the law,’ Kant presumably means ‘the agent who “writes” the law, or determines its content at his discretion.’⁵⁵ By contrast, the author of the obligation should be understood as the agent who confers the status of law onto the relevant principle and who makes it *binding* for those whom he addresses. As he puts it:

The author of the obligation in some sense addresses the law to some group of agents, and this volitional act is an authoritative reason to comply with the law that is independent of its content (or additional to any reason for complying that may be found in the content of the law) ... By addressing a law whose content is given by reason to a group of agents, a supreme lawgiver with authority over those agents can make the law binding for them; his volition is a sufficient reason for them to comply with the law (i.e., a reason for compliance over and above any reasons based on the content of the law).⁵⁶

⁵⁴ On the one hand, the fact of its being commanded would seem to be contingent rather than necessary. On the other, such a move would need to postulate that the agent takes an interest in obeying this command – which again leaves room for contingency.

⁵⁵ *Agency and Autonomy* 145.

⁵⁶ *Agency and Autonomy* 145-6.

One immediate question is whether there is any sense in which a purely formal moral law can be said to have a *content* that could give us a sufficient reason to act. With this worry bracketed, however, we still might ask whether Reath's characterisation of the two senses of authorship genuinely captures Kant's meaning. On Reath's reading, this passage proposes that the author of the obligation *makes* the law binding for those whom he or she addresses. Since there can be no one who determines the content of the moral law, Kant's description of rational agents as authors of the law in the *Groundwork* must appeal to the idea that we are authors in the sense that our command makes the law – whose content is already given – *binding* for ourselves and perhaps for others. In this way, the above passage might well be taken in conjunction with the passage from the *Groundwork* directly to support the Command Thesis. With careful attention to relevant passages elsewhere in Kant's works, however, it becomes clear that this segment from the *Metaphysics of Morals* provides inconclusive evidence for this version of the Command Thesis, leaving ample space for the kind of alternative I mentioned above. That is to say, the notion of authorship of obligation is less straightforward than we might first suppose.

In the *Metaphysics of Morals* passage quoted above, Kant is primarily concerned to describe the relationship that God might have to the moral law. Insofar as he uses the language of authorship also to describe our relation to morality, the temptation is to view what he says as applying in exactly the same way to human agents. Given his starting point, however, it will be useful to trace Kant's thoughts on God's authorship when it comes to morality, as this notion appears elsewhere in his works.

To begin, it is important to emphasise that Kant absolutely wished to avoid what he called 'theological morality' [*theologische Moral*]: '[t]hat religion, in which I must

first know that something is a divine command in order that I recognise it as my duty' (*Rel* 6:153.29-154.1).⁵⁷ That is, he rejects as heteronomous those views which ground the moral law in the authority of God's will. All the same, however, Kant is happy to allow for what he calls 'moral theology' [*Moraltheologie*], which he describes as 'a conviction of the existence of a highest being which grounds itself on moral laws' (*KrV* A 632/B 660.n).⁵⁸ On this alternative to theological morality, belief in divine command is wholly legitimate, so long as my acknowledgement of something as my duty does not in the least depend on an idea of its having sprung from the will or existence of God. This, according to Kant, would render duty conditional. Our ability to understand moral laws as categorically binding thus depends on denying that God's relation to the moral law is one in which he straightforwardly confers onto duties the status of law.

When it comes to the positive conception of this relation, Kant has a number of interesting things to say. In the Collins lecture notes of 1784-85, for instance, Kant rejects theological voluntarism in favour of a view according to which the nature of certain actions in itself is viewed as the source of obligation. As he argues, obligation can be divided into two exhaustive kinds: positive [*positiva*] and natural [*naturalis*]. The former, he explains, arises by a 'positive and voluntary choice,' whereas the latter arises 'from the nature of the action itself' (27:261.38-262.2).⁵⁹ All positive obliga-

⁵⁷ In the first *Critique*, theological morality is described as containing moral laws that 'presuppose the existence of a highest governor in the world' (*KrV* A 632/B 660.n).

⁵⁸ See also *Rel* 6:153-4.

⁵⁹ As Kant points out, 'Crusius believes that all obligation is related to the will of another. So in his view all obligation would be a necessitation *per arbitrium alterius*; but in fact I am necessitated by an *arbitrium internum*, not *externum*, and thus by the necessary condition of universal will; hence there is also a universal obligation' (*VC* 27:261.6-13). When it falls on post-*Groundwork* ears, this talk of being necessitated by an *arbitrium internum* and of uni-

tion is such that we are obliged to an action which is, in itself, indifferent. That is to say, I am bound to do or omit some action not because it is good or bad in itself, but because it stems from the choice of some putative authority and constitutes the means to fulfilling a command. Had it pleased that authority to choose something else, then *that other action* would be my duty. But since the moral worth of an action depends on the fact that the action is not performed because of fear of sanction or promise of reward (even in this pre-autonomy text, Kant takes this to be true), the possibility of moral goodness depends on there being an alternative to positive obligation: on the possibility of actions being good in themselves, and thus of there being a sufficient incentive to do an act for its own inner nature. God might very well demand various things of us, but if he does, then he does so *because* he recognises (or has made the world such) that an obligation arises from the nature of the action itself, independently of divine command (*VC* 27:261-2).⁶⁰

versal will points naturally to the idea of autonomy. We must be careful, however, not to impose such a reading too hastily, for the language of an action's being good or bad in itself, due to its inner nature, is just as naturally read as pointing to a pre-autonomy conception of obligation (I am not necessitated by anything external to the *action*, such as the will of another, but rather by the nature of the action itself) (see e.g. 27:280.19-24: 'In ethical obligations the motivating ground must be internal; one must do the action because it is proper, I must pay my debt, not because the other can compel me, but because it is proper to do so.' See also the quotation in footnote 60, p. 69). Later in the Collins lectures, Kant's discussion of internal and external obligation would seem to hinge on the difference between having a duty extracted from us by another and having a duty where no one else compels me (e.g. 27:269-70). In the Mrongovius lectures, Kant discusses internal moral compulsion as occurring 'if duty makes the action necessary, against all the other's inclinations, not by the will of another, but through his own will,' which carries strong hints of autonomy (29:617).

⁶⁰ 'All morality, however, rests on the fact that the action is performed because of the inner nature of the act itself; so it is not the action that makes for morality, but the disposition from which I do it. If I do a thing because it is commanded [*geboten*] or brings advantage, and omit a thing because it is forbidden or brings harm, that is not a moral disposition. But if I do it because it is absolutely good in itself, that is a moral disposition. So action must be done,

If it does not ground obligation as such, then what role does God's will play, on this account?⁶¹ In this regard, an important clue is provided at 27:282-3:

Anyone who declares that a law in conformity with his will obliges others to obey it, is giving a law. The lawgiver is not always simultaneously an author [*Urheber*] of the law; he is only that if the laws are contingent. But if the laws are practically necessary, and he merely declares that they conform to his will, then he is a lawgiver. So nobody, not even the deity, is an author of moral laws, since they have not arisen from choice, but are practically necessary; if they were not so, it might even be the case that lying was a virtue. But moral laws can still be subject to a lawgiver; there may be a being who is omnipotent and has power to execute these laws, and to declare that this moral law is at the same time a law of His will and obliges everyone to act accordingly. Such a being is then a lawgiver, though not an author, just as God is no originator of the fact that a triangle has three corners.

Here we find the same lesson as above: if moral laws had been chosen laws, then they would not be necessary; but it *is* a necessary truth that lying is wrong, just as it is necessarily true that a triangle has three angles.⁶² In addition, we encounter a further spelling out of the role that God's command might play in this story: moral requirements are indeed binding without his prior sanction, but God may declare that these

not because God wills it, but because it is righteous or good in itself; and it is because of this that God wills it and demands it of us' (*VC* 27:261-2). C.f. *Rel* 6:99.5-10: 'But neither can ethical laws be thought of as proceeding from the will of this superior (as statutes that would not be binding without his prior order [*Befehl*]), for then they would not be ethical laws, and the duty commensurate to them would not be a free virtue but an externally enforceable legal duty.'

⁶¹ Again, for the moment, I am restricting us to the account given in the Collins lectures.

⁶² At least, according to Kant. As Kant makes clear at the start of the lectures, '[a]ll necessary rules must hold good *a priori*, and hence the principles are intellectual' (27:254.30-32). Where a system of ethics is based on intellectual grounds, 'all morality is derived from the conformity of the action with the laws of reason' (27:252.36-253.1).

necessary requirements are in any case in conformity with his will, thus providing a posterior incentive to observe them.⁶³

The Collins lectures express positions that resemble Kant's mature views in a number of ways, and this is particularly so with regards to Kant's rejection of voluntarism, ethical egoism, and Wolffian perfectionism. Nevertheless, these texts are best read as pre-dating Kant's discovery of autonomy. Indeed, many passages give the distinct impression that Kant accepted a certain form of sentimentalism at the level of moral *motivation* at this stage in his thinking.⁶⁴ Although Kant makes clear that the understanding is that which apprehends the inner nature of an action (*VC* 27:254) – in particular, its agreement or disagreement with the laws of reason – he nonetheless holds that the moral motive is an added *feeling* of abhorrence (stemming from the faculty of sensibility) for what the understanding picks out as morally wrong: a feeling which is in addition to and separate from the understanding, and which is needed in order to make action possible.⁶⁵ As such, the usefulness of these lectures when it comes to understanding self-legislation may be limited.

⁶³ Of course, this is not to say that we ought to follow the law from this posterior incentive. Kant's pre-autonomy view of moral motivation involves the thought that the moral sense, separate from reason, somehow miraculously latches on to reason's apprehension of the law and provides a spring to action. As Kant explains at *VM* 27:1428, to be able to explain how the understanding on its own should have motivating power would be to possess the 'philosopher's stone' (here the term is likely used pejoratively). In any case, even at this early stage in Kant's thinking on moral motivation we see that the moral motive must latch on to the action's inherent goodness or obligatoriness, as ascertained by reason, and not to the fact that it has been endorsed by the deity or to the fact that we ourselves have legislated it.

⁶⁴ This is despite the fact that the *ground* of moral principle must lie in reason, on this view.

⁶⁵ See e.g. *VM* 27:1428-9, *VC* 27:275-76.

On Reath's account, there is no 'originator' (or author) of the law, in the sense of someone who *writes* the law, or formally determines its content. What the legislator of the law achieves, however, is the conferral of the status of law onto principles, such that an obligation is generated where there wasn't one before. Presumably, what this means is that not even God can be the author of the law, insofar as he does not determine the content of the candidate principles that in the end achieve the status of law. But in a sense, Reath's view *is* one according to which God determines the content of the law, to the degree that his legislative activity determines what *counts* as belonging to the class of principles that carry the status of 'law.' In the above passage from *VC* 27:282-3, however, we would seem to encounter a different picture entirely. Here, the lawgiver is described as someone who makes a certain kind of declaration ('The law is in conformity with my will, and I oblige others to follow it!'), though he declares this of a law that is in any case practically necessary, such that it cannot be the case that the opposite of what it commands could be made a virtue by any contingent decision. This lawgiver might indeed possess the power to execute a practically necessary law, in the sense of imposing punishment or reward, of offering posterior incentives to ensure that it is observed, but this is altogether distinct from *making* a principle a law in the first place.⁶⁶ To the degree that this is so, however, Kant's text

⁶⁶ C.f. *VV* 27:529, where Kant suggests that God's command may perhaps be significant as a feature that emphasises the 'force' of natural obligation. C.f. also *VC* 27:277: 'Because moral laws run, Thou shalt not, it is supposed that there must be a third being who has forbidden it. It is true that any moral law is an order, and that they may be commands of the divine will, but they do not flow from such a command. God has commanded it because it is a moral law, and His will coincides with the moral law. It also seems that all obligation has a relation to one who obliges, and thus God appears to be the obligator of human laws. In performance, to be sure, there must indeed be a third being, who constrains us to do what is morally good. But for the making of moral judgements we have no need for any third being. All moral laws can be correct without such a being. But in execution they would be empty if no third being could constrain us to them. It has therefore been rightly perceived that without a supreme judge all moral laws would be without effect, since in that case there would be no inner motive, no reward and punishment.'

would seem to make room for the possibility of some sort of legislative activity that involves *neither* the determination of content *nor* the generation of obligation, presuming these elements can actually be distinguished.

Lest we think that such an option is precluded by Kant's mature moral views, and in particular by the introduction of autonomy in 1784-5, evidence for this alternative can equally be found in the Vigilantius lecture notes, which are based on Kant's lectures from 1793-94.⁶⁷ As Kant is recorded to have proposed:

...an *auctor legis* can be supposed only of a law that has no binding power of its own, but possesses it merely *ex voluntate vel arbitrio alterius* [from the will or choice of another]. Since an *auctor* is *causa per arbitrium liberum*, and therefore everything depends on his choice, it [the concept of an *auctor*] can only be applied to a *lex statutaria*. [...] [I]f we ascribe an *auctor* to laws that are known, through reason, from the nature of the case, he can only be author of the obligation that is contained in the law. In this way too is God, through the declared divine will, *auctor legis*, and precisely because natural laws were already in existence, and are ordained by Him. (VV 27:544)

As Kant goes on perhaps surprisingly to note, if, in addition to the moral law, we could think of a being who is able to unite appropriate consequences with the lawfulness of our actions – a being 'who is the author of our happiness, who knows and gauges our worthiness to fare well, or our moral worth' – then we would have an idea which would bolster a person's moral disposition and help to strengthen his resolve in the face of moral obstacles (27:545). The idea of divine command thus seems, on this account, to serve an important, yet secondary, motivational role, strengthening and supporting our interest in doing the right thing, not because of threats or promises, but

⁶⁷ The principle of autonomy is also most certainly at work in the Mrongovius II lectures. See, e.g. 29:628-29.

because connecting the law to God's will reminds us of the law's special status.⁶⁸ In no sense, however, do we find hints of the idea that God 'confers' the status of law onto principles where they did not already possess necessary validity. We might of course want to know more about what it *means* for an agent to declare that a law is in conformity with his will and what work this notion really does in Kant's thinking, yet such a question merely points toward an interesting avenue of thinking about Kant's distinction: it does not on its own force us to retreat to the distinction between the determination of content and the generation of obligatory force by 'conferring' authority.

2.5 *The Fate of Command*

In this chapter, I have been attempting to undermine the appeal of a particular version of the Command Thesis. I began by arguing that certain versions of Kantian constructivism which specify the thesis's content face quite serious difficulties when it comes to the reconciliation of their central claims and the necessity and universality of moral requirement. I have also attempted to show that passages which are often taken to support a reading of Kantian autonomy along these lines are inconclusive at best. Indeed, when it comes to Kant's distinction between the two senses in which God might be understood as the author of the moral law, we have good reason to think that his thinking on the matter reveals a conception of legislation which concerns neither the power to determine the content of the law nor the conferral of normative force.

⁶⁸ The obvious complications attached to this notion will be discussed further in chapter three.

This is not yet to show, of course, that this alternative conception of God's legislative authority can or ought to be applied in any sense to the concept of the autonomy of the imperfectly rational will. While I hope to have identified the space for an alternative to the tack taken by some constructivist views, further work must be done in order to show that such an alternative holds promise in the effort to understand Kantian self-legislation. It is to this task that we now turn.

PURE RATIONAL CONSTRAINT

The problems that accompany the Command Thesis would seem to hinge on its implicit voluntarism about the binding force of the moral law. Where the obligatoriness of a practical principle for a particular agent is taken to depend on a legislative act of her own will, the task of accommodating its universality and necessity becomes extremely difficult. What, then, of the Rationality Thesis, which points us toward a very different direction?

If we recall, this thesis states that the moral law is in some sense a ‘law of reason.’ No independent act of the will is required in order for it to be binding on an agent, for its obligatory force is secured entirely by the fact that it is a genuine *rational* constraint. At first glance, this way of thinking fits very well with many things that Kant says about the moral law. Kant continually emphasises the *a priori* nature of the fundamental moral principle, suggesting that if there is to be a necessarily binding moral law, it must determine the will via rational considerations alone, and not by reference to anything empirical. If there is such a law, then it must be connected, completely *a priori*, ‘with the concept of the will of a rational being as such’ (*G* 4:426). But what does it mean to say that a principle is connected *a priori* with the concept of the will

of a rational being *as such*? How might we develop this notion while preserving the notion that the moral law finds its source in *our very own will*?

I begin this chapter by exploring two recently popular conceptions of the sense in which the moral law is a ‘law of reason.’ As I will argue, these approaches to developing the Rationality Thesis fail for a number of important reasons. I then go on to advance a more orthodox approach to developing this thesis – one which relies on a better understanding of both Kant’s conception of a ‘ground of determination’ and his reasons for rejecting the determining grounds at the heart of previous moral theories. Finally, I consider the degree to which select elements of the Command Thesis may be included in this picture as well, with the result that Kantian autonomy may be viewed as incorporating aspects of both of the original theses presented in §1.1, though in very particular ways.

3.1 *A Teleology of Practical Reasoning*

It is helpful to begin by asking what it means to say that the moral law is the fundamental law of practical reason. On this front, a recent popular interpretation seeks to explain this notion in terms of an inherent connection between the moral law (the categorical imperative) and the *purpose* of the activity of practical reasoning. The reasoning behind this view is as follows.

Just like every human faculty, the faculty of practical reasoning has an essential end, and the norms that govern its exercise are determined by the nature of its proper func-

tioning. Where an organism possesses the faculty of vision, for example, the proper exercise of this faculty allows for the realisation of an end that is vital to its survival (i.e. an essential awareness of the objects in its surroundings). Successful functioning toward this aim is distinguished from a dysfunctional operation by what we might call the ‘norms of vision.’¹ Yet in just the same way, there is a certain aim present in every instance of practical reasoning, the nature of which gives rise to a normative principle that expresses how reasoning must proceed in order to realise it. In this sense, the moral law is the fundamental norm of this distinctive faculty, and this explains its obligatoriness for every finite rational being. To the degree that one is engaged in the activity of practical reasoning at all, one is committed to the realisation of its essential aim. To the degree that one is committed to *this*, one is committed to following the categorical imperative.

There might be two senses in which one is ‘committed’ to following the categorical imperative on this general approach. First, the claim might be that rational agents must *in fact* adopt the moral law as the principle of their behaviour in order to engage in practical reasoning at all. Second, the idea might be that any rational agent must, in the sense that he or she *ought*, to adopt the categorical imperative as her guiding principle of action, simply by virtue of engaging in the activity of practical reasoning. Let us consider each option in turn.

¹ See Wood, *Kantian Ethics* 115.

(i) *Constitutivism*

On this first line of thought, the key idea is that if you do not *actually* follow the moral law, you cannot be said genuinely to be engaged in the activity of practical reasoning. Because of this, the categorical imperative can be seen as a ‘constitutive’ norm of that very activity (a norm which makes the exercise of the faculty of practical reasoning possible).² To the degree that one is (or takes oneself to be) engaged in the practice of practical reasoning, this principle must be adopted.³ It is a standard which an agent must aim to meet simply in virtue of being an agent.

In coming to understand this idea, it is useful to think of an analogy. In the craft of medicine, for example, we say that a good doctor is one who makes his patients well (where he can). To the extent that this is so, health is the aim of the activity of medicine. On the view we are considering, however, to say this is just to say that someone who is genuinely engaged in the activity of medicine cannot simply put aside the question of whether the means she is taking are conducive to the health of those she is

² Alison Hills offers a view which might also be called ‘constitutivism,’ although her position focuses on the role that rational nature plays when it comes to the value of our non-moral ends. Hills’s constitutivism differs from the above view in important ways; rather than arguing that the moral law is a constitutive norm of practical reasoning as such, she argues that a given non-moral project is valuable to the degree that it is a constitutive means of exercising our rational nature. Although she does not offer a full-fledged vindication of the categorical imperative, she does acknowledge that her view gives support to the Formula of the End in Itself: on her view, the value of non-moral ends depends on the role that they play in the exercise of rational nature, which means that there is no reason for a person to pursue an end at the expense of rational nature. Because of this, we are required to respect rational nature in all of our actions. See Hills’s ‘Kantian Value Realism,’ *Ratio* 21 (2008): 182-200.

³ Versions of constitutivism concerning the categorical imperative have been offered by Andrews Reath (see in particular *Agency and Autonomy* ch. 5); Christine Korsgaard (‘Self-Constitution in the Ethics of Plato and Kant’ *The Journal of Ethics* [1999]: 1-29); and Barbara Herman (*Moral Literacy* [Cambridge, MA: Harvard University Press, 2007] ch. 7).

treating. Should she stray too far from the internal aim of medicine, she can no longer be said to be engaged in that activity at all. In just the same way, an agent cannot simply set aside the question of whether her maxims are universalisable, because this attaches directly to the question of whether her practical reasoning fulfils its aim. If a person rejects the standard that governs practical reasoning, she can no longer be said to be engaging in that activity at all. In this sense, even the most substandard agent must *try* to adopt universalisable maxims, for there simply is no other way to engage in the activity of practical reasoning.⁴

Is there evidence in Kant's writing to support the claim that this is his view? At first glance, the most promising support for this interpretation would seem to be found in a series of claims spanning *Groundwork* II and III. Kant begins the latter section by defining the will as a kind of causality of rational beings insofar as they are rational. Cast in negative terms, the freedom of this will consists in independence from determination by 'alien causes'; yet because a lawless will is an absurdity [*ein Unding*], a positive statement of the will's freedom 'flows' from this negative conception. As a form of causality, the will must be determined by some law or principle.⁵ Yet it must also determine itself independently of externally given laws. It must, in other words,

⁴ Korsgaard, 'Self-Constitution' 15. On Korsgaard's view, genuinely willing a maxim (through the faculty of practical reason) requires that one at least *take* one's maxim to be universal (27). Willing which is undertaken with this aim in sight, but which in fact falls short of this ideal, is still a species of willing, only 'badly done.'

⁵ Kant's insistence that in general a cause must operate according to laws (so that every will must operate in accordance with some law or another) is of course controversial. Here, it would seem to be presented by Kant as self-evident, though, as Karl Ameriks notes in his study of *Groundwork* III, Kant does seem to oppose freedom and law at one point in the first *Critique*, which raises doubts concerning its immediate self-evidence (see 'Kant's *Groundwork* III Argument Reconsidered,' *Interpreting Kant's Critiques* 233n.13).

express determination by a principle that it gives to itself (*G* 4:446-7). But what this seems to imply is that the free exercise of practical reason requires that the will be guided by its own internal principle – a principle which, as Reath puts it, ‘positively constitutes the capacity for free volition’ and which, when specified for the imperfect will, corresponds to the universal law formulation of the categorical imperative.⁶ In this way, the categorical imperative can be thought of as the principle which captures the general form of free practical reasoning.⁷

Notoriously, this argument faces a cluster of very serious objections. Particularly worrying for the constitutivist who appeals to it is the fact that she must be able to make sense of a necessary and immediate connection between the activity of practical reasoning *as such* and the activity of *free* practical reasoning (in the relevant sense of ‘free’). Here, things get especially murky.⁸ Yet even if these standard worries could

⁶ Reath, *Agency and Autonomy* 154-55.

⁷ On Reath’s view, Kant maintains that all action is necessarily undertaken under the idea of freedom (*G* 4:448). A commitment to one’s freedom is an inescapable feature of the deliberative perspective, such that if one is engaging in practical reasoning at all, one is committed to viewing oneself as free, and so as guided by the categorical imperative. See *Agency and Autonomy* 156-7. For an important criticism of this kind of argument, see Ameriks, *Interpreting Kant’s Critiques* ch. 9.

⁸ This is especially true for views like Reath’s, where the link is established via an appeal to how we must ‘take’ ourselves to be functioning whenever we engage in deliberation. See also Korsgaard, ‘Self-Constitution’ 23: ‘You take *yourself*, rather than the incentive on which you choose to act, to be the *cause* of your action. And Kant thinks that in order for this to be so, you must act on a universal law. You cannot regard yourself as the *cause* of your action – you cannot regard the action as the product of your will – unless you will universally.’ As Karl Ameriks quite rightly points out, what is needed in order to make the argument work is something more than this phenomenological point, even if at least *this* could be established: ‘despite our operating often “through” reason in a phenomenological sense – that is, simply with explicit thought (or later “rationalisations”) of rational grounds for what we are doing – it might be the case that the real power that makes the events that are those thoughts “effective” is nothing more than nature ... in its subrational power. In that case, we might be said to

be resolved, the constitutivist would nonetheless face the problem of explaining how clear-headed moral wrongdoing is possible on its own terms, and here the prospects look even worse.⁹ On this view, in order to be reasoning at all, a rational agent must in some sense adopt the categorical imperative. If this is so, then any actual deviation from the moral law must be explained either as a case in which one is not in fact engaged in practical reasoning or as one in which one has adopted the categorical imperative but has made a mistake in its use. In the first case, the category of practical wrongdoing cannot be applied, for ‘moral transgression’ is possible only where behaviour has resulted from practical reasoning. Yet the same conclusion would seem to hold true for the second case as well, to the extent that it must explain genuine instances of practical reasoning which miss the mark (which do not in fact conform to the categorical imperative) as involving the *misuse* of the categorical imperative, and so as involving a cognitive error rather than a rejection of the moral law.¹⁰ This is troubling both to the degree that Kant’s conception of moral wrongdoing depends heavily on the idea of an active rejection of or turning away from the moral law – a real *choosing* of evil as an alternative to goodness – and to the extent that clearheaded

be working “through” the appearance of reason but not “really by” it, and our whole everyday intentional life would be little more than a mere “stance,” an epiphenomenon at best’ (*Interpreting Kant’s Critiques* 235).

⁹ By ‘clearheaded wrongdoing’ I mean deviation from the moral law which is not explainable as a mere cognitive mistake – as a misapplication of the moral law.

¹⁰ As Barbara Herman explains (re: her own particular version of constitutivism), ‘[i]f by free action one means, as Kant does, action derived from the will’s own principle, then the [morally bad] action is free in that sense. (Analogously, we say that mistakes in addition are mistakes *in addition* – that is, a result is derived from relevant arithmetic principles misapplied, perhaps with self-serving lack of care. Unfree action is like a child’s assertion of a sum, for no arithmetic reason.) Faulty action is thus imputable to the autonomous will, since the principle of the maxim is a representation – albeit a *misrepresentation* – of the will’s own law’ (*Moral Literacy* 172).

wrongdoing is, in general and beyond what Kant himself wants to say, paradigmatic of culpable action.¹¹

In effect, this view collapses the distinction between two very different kinds of standard, and it is precisely this move which effects the preclusion of the possibility of genuine wrongdoing. Consider, again, our physician analogy. Suppose that a good and an inferior physician both aim at the same end: the health of the patient (let us call the good physician A and the inferior one B). In this case, the relevant difference between A and B lies only in their *effectiveness*: one operates well toward this end, the other badly. Now consider a third physician, C. C is as skilled as A in terms of technique and medical know-how and is as effective a doctor, but in a certain circumstance C rejects the aim of healing her patient and adopts the aim of poisoning him instead. She makes use of all kinds of medical technique, in some sense acting just as a doctor would: knowing what she is doing and why, exercising her skill. In another sense, however, she is not acting as a doctor would at all. But what our dual sense of this situation illustrates is the fact that there is a distinction between what we might call standards of efficiency (where if you fall too far below, you really aren't engaging in the relevant activity) and standards that have to do with the achievement of the proper end of that activity (standards that express what we must do if our exercise of that activity is to promote its essential aim). The actions of the first 'bad' doctor (B) conform to the second but not to the first standard, while the actions of our second

¹¹ As Kant makes clear at *G* 4:424.15-20, though we acknowledge the validity of the categorical imperative, our problem is that we sometimes 'permit ourselves (with all respect for it) just a few exceptions ...' See also *Rel* 6:22n, where Kant quite explicitly argues that any lack of agreement with the moral law is possible only as an active resistance on the part of *Willkür* to the law (an 'antagonism to the good').

‘bad’ doctor (C) conform to the first but not to the second.¹² Yet unless there is room in our thinking about practical reasoning for a character parallel to C, who is in some sense effectively engaged in the activity of practical reasoning but who has rejected that activity’s characteristic aim, genuine wrongdoing becomes impossible. To the degree that we want to be able to make room for such a notion, our theory must be able to distinguish between the end(s) sought in good practical reasoning and the end(s) sought in bad practical reasoning.¹³

(ii) The Aim of Practical Reasoning

Suppose that we try to avoid these difficulties by adapting the above view. One way to do this is to propose that the moral law, while not constitutive of practical reasoning as such, simply expresses what we must do in order to realise the essential aim of practical reasoning, where this aim is to be understood as the realisation of the proper functioning of practical reasoning. Does this bring us any closer to a plausible conception of the way in which the categorical imperative might be viewed as a law of practical reasoning?

¹² In his discussion of the meaning of ‘perfection’ in his lectures on ethics, Kant draws a related distinction between perfection and morality. Where the ‘perfect’ villain possesses a capacity or fitness for all of his ends (talent, skill, craft, or know-how, we might say), moral goodness consists in the property of making *good* use of these perfections (see *VC* 27:265-66 and *VV* 27:517-18). See also *KpV* 5:41. As Aristotle acknowledges in Book VI of the *Nicomachean Ethics*, a person of bad character might deliberate ‘correctly’ in one sense to the degree that he achieves what his project requires. In another sense, however, he of course deliberates *badly*, where this badness has to do with the relation of his action to a particular end (acting morally *well*). See 1142b18-20. See also Aristotle’s discussion of cleverness at VI.12.1144a24-29.

¹³ Ultimately, as we will see, the categorical imperative is neither a norm of efficiency nor a standard that has to do with the achievement of the end of an activity.

This view faces a twofold task: first, it must be able to explain the connection between the aim of practical reasoning and the categorical imperative; second, it must preserve the sense in which the moral law is a universally and necessarily binding practical principle. In the end, this approach faces serious difficulties on both of these fronts.

When it comes to explaining the connection between the aim of practical reason and the categorical imperative, the problem is that we would seem to be able to draw the relevant connection only at the cost of reducing this aim exclusively to the generation of moral maxims. On a very natural view, we might suppose that practical reasoning is an activity that aims principally at bringing about a certain state of affairs – at changing the world in some way or another. This is a morally neutral conception of the point or purpose of practical reasoning and one that does not invoke the idea of universalisability of maxims in the least. I might effectively bring about an object of the will (a state of affairs) by acting on the principle of self-love, and on this conception my doing so constitutes the achievement of the will's central aim. In order to rule this out as a candidate conception of the aim of practical reasoning, however, we must restrict the aim of practical reasoning to action which expresses universalisable maxims. In effect, we must say that the chief end of practical reasoning consists in the generation of morally good actions.¹⁴

¹⁴ It might, as a matter of psychological fact, be the case that some of us aim in our action to be morally good. Kant's example of the scoundrel who is generally accustomed to using reason to help him to pursue his ends, yet who, when presented with examples of great benevolence and moral strength of will, wishes that he could be so disposed, seems to be someone of this sort (*G* 4:454.21-27). Even here, however, Kant would seem to be describing a psychological point about human beings rather than providing an argument for the necessity of

Once we make this move, however, this approach loses one of its key assets. One of the benefits of the approach I have been sketching is that, if successful, it provides us with a way of vindicating the bindingness of the moral law. On this view, the categorical imperative procedure stems analytically from an articulation of the essential concern of practical reasoning, such that we can say that the nature of practical reasoning yields a law that can guide its own proper exercise. To the degree that we ought to pursue this activity's central aim, we ought to follow the categorical imperative. Yet to the extent that this view is forced to define the aim of practical reasoning precisely in terms of universalisable maxim selection (in order to draw the appropriate connection between the aim and the moral law), the key move in this vindictory project – the introduction of the idea of an aim of practical reasoning – loses all its usefulness. When an agent asks why she ought to act only on those maxims such that she can will at the same time that they be a universal law, the reply on this view must be that she ought to because she ought to pursue the aim of practical reasoning. When pushed on what this aim is, the supporter of this view will be forced to say that it consists in acting on maxims that can be willed as a universal law. But then we are back to where we were with our first question, and our search for vindication, should we wish to pursue it, has not really progressed at all.

Most serious of all, however, is an objection that applies to both versions of this teleological conception of the activity of practical reasoning and which concerns their ability to account for the necessity of the moral law. On both interpretations, the law is conceived as a norm that directs us to the realisation of the essential aim of practical reasoning. Our faculty of practical reason, the power by which we engage in the

acting morally well. The gap between instrumental and non-instrumental reasoning is and ought to remain unbridgeable, for Kant.

activity of practical reasoning, has an essential function, and the moral law expresses the principle it must follow in order to realise this function.¹⁵ One glaring problem with this approach, which I will not explore here, is the difficulty of specifying a single aim or function of practical reasoning. Yet even if such an aim can be identified, we face a more serious problem – for a story like this is unable to preserve the absolute universality and necessity that Kant requires of morality. On this view, the normativity of the moral principle is only as strong as the normativity of the aim of practical reasoning. The moral law's status as a norm we ought to follow is grounded in the fact that practical reasoning is an activity that commits us to pursuing a specific aim. If this fact is contingent, such that our aim could have been otherwise, then our subjection to the moral law is rendered contingent as well. The moral law is reduced to a problematic imperative of skill rather than an apodictic imperative of morality, for it specifies the means one must take in order to realise a purpose that human beings contingently pursue.¹⁶ Conversely, if we deny the contingency of our having this aim and suppose instead that this fact represents a natural necessity of sorts (such that our adopting this aim is an empirically necessary part of human nature), we are no better off, for the moral law is still rendered hypothetical (conditional on the agent's aims). What we end up with, in effect, is an assertoric imperative, for the moral law must then be taken to specify the means one must take in order to realise a purpose that human beings non-contingently pursue.¹⁷

¹⁵ For this kind of story, see Wood, *Kantian Ethics* 114-16.

¹⁶ For Kant's association of problematic imperatives and the idea of *possible* purpose, see *G* 4:414-15.

¹⁷ For Kant's association of assertoric imperatives and the idea of necessary aims or ends, see *G* 4:415-16. See also *KpV* 5:25.12-13, where Kant articulates the notion that happiness is

A version of this teleological approach has been offered by Paul Guyer, who, in a self-consciously controversial move, interprets Kant as having intended adherence to universal law to be understood as a means to the realisation of freedom.¹⁸ On Guyer's interpretation, the value of freedom is prior to the moral law; it provides the end to be realised by following the moral law and the reason why it has authority over us. Quite interestingly, however, Kant criticises just this very sort of move in the Dialectic of the second *Critique*, where the target is the Stoic belief that consciousness of strength of soul might constitute the ultimate end of morally good action, serving as that which gives it value.¹⁹ Although the Stoic doctrine constitutes an improvement compared to the view offered by the Epicureans, according to Kant, the Stoics make that which is merely a by-product of moral action (the 'elevation of one's cast of mind') the end to be realised. It might well be the case that morally good action gives a person strength of soul – or, as Guyer emphasises, makes a person truly free – but these rightly valued consequences cannot be the grounds of determination when it comes to morally worthy choice. Exposing people to examples of the connec-

'necessarily the demand of every rational but finite being ...' As Kant goes on to explain, one problem with assertoric principles of prudence in particular is that happiness is just a general concept, and the pursuit of this end will take various forms for various agents. Because of this, the assertoric imperative cannot be regarded as genuine law, since a law 'must contain the *very same determining ground* of the will in all cases and for all human beings' (5:25). Even if all rational beings were somehow to be agreed when it comes to the content of happiness, however, we could not therefore conclude that the principle of happiness is a genuine practical law, for this unanimity would be merely contingent. The agreement itself would be an empirically universal fact, in which case the relevant law would only contain subjective and not objective necessity (*KpV* 5:26).

¹⁸ See, e.g. *Kant on Freedom, Law, and Happiness* (New York: Cambridge University Press, 2000) 1. For Guyer's articulation of his position, see esp. ch. 3. See also Andrews Reath's commentary on Guyer's view in 'Value and Law in Kant's Moral Theory,' *Ethics* 114 (2003): 127-155.

¹⁹ See *KpV* 5:127n.

tion between moral action and freedom, happiness, wisdom, and so on, might help them to act morally, and so might play a secondary motivational role, but this is not to say that we should conceive of such things as the end of following the categorical imperative – as that for the sake of which good people act morally well.

3.2 *The Rational Grounds of Morality*

The central danger affecting the above approach is that it effectively seeks to transform Kantian ethics into a teleological ethics. On such a view, where priority is given to some goal an agent might or must pursue (free agency, for example), the moral law can only ever serve as a *conditional* imperative; good actions are only ever instrumentally good, relative to their role in the realisation of some end. Yet from the very start of the *Groundwork* Kant draws a strict and clear distinction between conditional and unconditional goodness, and here moral goodness is placed squarely on the side of the latter. The good will, he writes,

is good not because of what it effects or accomplishes, because of its fitness to attain some proposed end, but only because of its volition, that is, it is good in itself and, regarded for itself, is to be valued incomparably higher than all that could merely be brought about by it ... (4:394.13-18)

Although it may be presented by certain of Kant's interpreters as a particularly worthy or 'necessary' end, free agency (or properly functioning free agency) on the above view constitutes a goal the achievement of which is that upon which the value of morally good action must be seen to depend. Where the above approach offers what at first seemed a promising way of explaining the idea that the moral law is a law of practical reason, it thus ultimately fails to secure this law's unconditional

bindingness. As such, if we are to pursue the Rationality Thesis as a viable option in coming to understand Kantian autonomy, it seems as though we must take a different tack.

In one sense, a practical principle might be considered a law of reason to the extent that its *ground of determination* (*Bestimmungsgrund*) is rational rather than empirical.²⁰ On Kant's view, moral theories that ground the moral law in happiness must necessarily posit as a condition of a rule's being a practical law *empirical* determining grounds. That is to say, they must posit an empirically determined end in order to explain why a given precept is a rule that we ought to follow and in order to explain how we can recognize what we ought to do. On such theories, the ability of the rule formally to determine my will in this way is conditional on the relation of a represented object to the feeling of pleasure or displeasure, and this can only ever be

²⁰ A ground of determination can be thought of as that which 'contains' the law (see Timmermann, *Kant's Groundwork of the Metaphysics of Morals: A Commentary* 181). In general, we can think of a ground of determination as that which accounts for why a rule is a law for a person (a principle with overriding normative force). In this sense, it will explain how we can recognise what our duty is (by examining the ground, we derive the duty) and it will play a role in explaining why we might be motivated to follow this rule. See, e.g. *Rel* 6:4: '...since its laws bind through the mere form of universal lawfulness as the highest condition (itself unconditional) of all ends, morality needs absolutely no material determining ground of the free power of choice, that is no end, either in order to recognize what duty is or to impel its performance; on the contrary, when duty is the issue, morality can perfectly well abstract from ends altogether, and ought so to do. For example, to know whether I should (or even can) be truthful in my testimony before a court of justice, or faithful when someone else's goods entrusted to me are being reclaimed, there is no need to demand an end which I might perhaps propose to myself to realise my declaration, for what sort of end this would be does not matter at all; rather, one who still finds it necessary to look around for some end when his testimony is rightfully demanded of him, is in this respect already contemptible.'

specified empirically.²¹ As such, my cognition of what I ought to do and the explanation of why it is a rule can only ever hinge on empirical considerations.

By contrast, Kant takes the tradition of Leibnizian and Wolffian perfectionism, along with the theological perfectionism articulated by Crusius, to draw the principle of morality from something which can only be thought of by means of *rational* concepts, and so from rational grounds of determination. In the first case, the binding force of the moral law (its status *as* law) is derived from the rational concept of perfection as such, whereas in the second, it is drawn from the concept of supreme perfection represented in substance (God and his will).²² On the straightforward perfectionist view, our duty is to make ourselves perfect. On the theological view, our duty is to bring our will into agreement with the will of God, and only indirectly to seek our perfection (which he wills).

As is familiar, Kant in the end includes these two approaches in his list of theories that rely on material principles and which cannot serve as universal and necessary practical laws.²³ He thus opens a space for a kind of rational yet material practical rule. As he describes things in the second *Critique*, the practical concept of perfection signifies ‘the fitness or adequacy of a thing for all sorts of ends,’ whereas the concept

²¹ See *KpV* 5:27 and *G* 4:442. According to Kant, the principle of happiness is built upon physical *or* moral feeling, both of which depend on ‘*the special constitution of human nature* or the contingent circumstances in which it is placed.’ Moral feeling is here included under the principle of happiness because it is an empirical interest which ‘promises to contribute to our well-being by the agreeableness that something affords, whether this happens immediately and without a view to advantage or with regard for it’ (*G* 4:442n)

²² See *KpV* 5:40-41 and *VM* 29:622.

²³ A material principle contains the ground of determination by its matter and not by its form (*KpV* 5:27).

of perfection in substance (God) signifies ‘the adequacy of this being to all ends in general’ (*KpV* 5:41.10-11, 14-15). Yet if the concept of perfection signifies a fitness for possible *ends*, then its ability to serve as a ground of determination depends ultimately on an expectation of happiness. As Kant explains,

...if ends must first be given to us, in relation to which alone the concept of *perfection* (whether internal in ourselves or external in God) can be the determining ground of the will; and if an end as an *object* which must precede the determination of the will by a practical rule and contain the ground of the possibility of such a determination – hence as the *matter* of the will taken as its determining ground – is always empirical; then it can serve as the Epicurean principle of the doctrine of happiness but never as the pure rational principle of the doctrine of morals and of duty (so too, talents and their development only because they contribute to the advantages of life, or the will of God if agreement with it is taken as the object of the will without an antecedent practical principle independent of this idea, can become motives of the will only by means of the happiness we expect from them). (*KpV* 5:41)²⁴

On the perfectionist story, the direct determining ground of the will is rational insofar as it takes the form of one of two rational concepts. In order for this concept to function in this capacity, however, we must presuppose a given end, and so the principle turns out to be material. In order for this rational concept to play the determining role that the perfectionist theory supposes that it does, an antecedent object of the will must be given.

²⁴ Further criticisms of this strand of thinking, many of which hinge on what Kant takes to be an emptiness and indeterminacy when it comes to the concept of perfection (along with the claim that, when it *is* specified, it actually *presupposes* the moral law which it is meant to explain), can be found in *G* 4:443, *Rel* 6:3n., *VC* 27:265-66, *VV* 27:517-19, 544, and *VM* 29:626-27.

Consider, for example, a version of the theological approach that Kant attributes to Crusius. On this picture, the moral law is a law that God has imposed on us, and it directs us to do what is in accordance with his perfection. What is more, we are able to grasp morality's dictates because God has implanted innate representations of the law in each of us. This grasping of the content of morality is thus rational, in the sense that we need not turn to experience in order to determine what we ought to do. In addition, however, this law is grounded in a concept (the perfection of God's will) which we can also only grasp via reason, and so the principle is 'rational' (rather than empirical) in this sense as well. Now the primary issue that Kant takes with such a view is that this concept (God's perfection) can only serve as a determining ground of the will to the degree that we have antecedently adopted the end of being in agreement in our volition with God's will.²⁵ When an agent is considering what to do, she will be able, in theory, to tell herself a story about the justification of acting as this innate law tells her to which does not rely on any appeal to empirically discernable facts. There is a theological account of the origins of the innate law which she represents to herself in deliberation, and which is thus rationally accessible to her.²⁶ Yet in order to function as a ground for the adoption of this law, this theological account must be available to her *and* must coincide with a desire on her part to harmonise her will with the will of God. Although in this case the ground (the reason why it is a law

²⁵ The other charge that Kant mentions is that the concept of perfection is empty: that it is not a determinate concept at all, and must be filled in by a more informative notion.

²⁶ See *VM* 29:627: 'The theological principle depends on a being whose existence is inferred from reason...' As Kant makes very clear in the *Religion*, the idea of God cannot play a *foundational* epistemological or motivational role in moral agency: 'So far as morality is based on the conception of the human being as one who is free but who also, just because of that, binds himself through his reason to unconditional laws, it is in need neither of the idea of another being above him in order that he recognise his duty, nor, that he observe it, of an incentive other than the law itself. At least it is the human being's own fault if such a need is found in him' (6:3).

to her) is *grasped* by reason alone, the agent's ultimate motivation to comply with the law can only ever be mixed: tainted with a view to the desirability of attaining the object picked out by the concept of perfection.²⁷ What this shows is that the principle is really a material principle, and so is unfit to serve as a necessary and universally binding law.

With this brief explanation in tow, we can now make better sense of what Kant thought he had to say about a truly necessary law of reason. What we would seem to need is a story according to which i) the determining ground of the will contained in that law is based in reason, *and* ii) this ground does not derive its normative force from elsewhere (e.g. an antecedent desire). Even theories that satisfy this first condition can in principle offer principles which are not universally adoptable on Kant's account, and this points to the idea that an adequate understanding of what makes his conception of reason's lawgiving distinctive hinges in particular on making sense of how it accounts for the second condition.

Let us begin by thinking about the sense in which Kant deals with the first condition, which concerns the rational nature of the ground of the moral law. According to Kant, the moral law is not grounded in any object of either reason or the senses. To suppose this would be immediately to reduce the moral law to a material practical principle, which, as we have seen, Kant cannot do. But there must be *some* ground of determi-

²⁷ Note that, in addition to this concern about the possibility of pure moral motivation, the objection concerning the hypothetical status of such a law is equally applicable here. To the degree that such a law is grounded in a given object or state of affairs, it can only command assertorically.

nation when it comes to this law – *something* which contains it and which explains why it is a law for me. What could this be?

Kant's answer to this question may seem curious at first. As he puts his idea, the determining ground of action in accordance with the moral law can be nothing other than universal lawfulness as such (*KpV* 5:27.12-14). If no object (or material) of the will can ground a necessary moral law, then the only alternative available to us is to locate its determining ground in a *formal* feature of that principle: in the feature of universal lawfulness as such, or in the mere form of law (*KpV* 5:28.31).²⁸ The mere form of lawfulness can be represented only by reason, and is therefore not an object of the senses (*KpV* 5:28.34-35). But to say this is just to say that the moral law finds its source in pure rational considerations alone, and in this specific sense is a law 'of reason.'

Now since the ground of determination associated with this principle is nothing other than its form (its universal bindingness), we can say that, in a very real sense, the law *itself* provides its own determining ground. If the determining ground were to consist in something outside of that law (for instance, an object or state of affairs represented through it), then we would have to draw a real distinction between the law and its ground. This is not the case when it comes to the categorical imperative, however, for here the ground consists in something essential to that very law, which cannot be separated from it. In this sense, the idea of a possible giving of universal law, universal lawfulness, the form of law, *and* the representation of the law itself are all differ-

²⁸ See Lewis White Beck, *A Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960) 116. Kant sometimes explains this ground in terms of 'the a priori thought of a possible giving of universal law' through one's maxims (*KpV* 5: 31, 41).

ent ways of describing morality's ground of determination.²⁹ To the degree that the moral law supplies *its own ground*, cognisable through reason alone, its authority for us is non-derivative.

Notice that this kind of answer does not provide a genuinely satisfying response to the question of why we should think that the moral law is universally binding in the first place. To say that the universal lawfulness of the moral law is its own determining ground is to say that universal lawfulness is what explains why it is a law for me, or for anyone who might pose the question and to whom it applies. Where I ask, 'why is it that *I* must obey this principle?' the answer is that it is a principle which applies with the force of law to every morally responsible rational being (this is the relevant formal feature, and I just happen to be a morally responsible rational being).³⁰ But if I then ask why this principle is universally binding in the first place, I cannot again appeal to this fact about its form – that is, to its universal bindingness – without arguing in an obvious circle. What this shows is that this formal feature's justificatory force is only in play from within a standpoint that already recognizes the universal

²⁹ See *G* 4:401-2 for an example of where Kant speaks of the representation of the law as that which must determine the will.

³⁰ As we will see in §3.3, Kant raises the possibility of a fully rational being who would not, strictly speaking, be subject to the moral law (see *Rel* 6:26.n). This imagined rational being possesses an intellectual understanding of whether his maxims are suited to universal legislation, but he can only be moved to act by incentives provided by the objects of inclination. As Kant suggests in this footnote, the independence of our power of choice from determination by all non-moral incentives is what makes us accountable for our actions. This would suggest that, in order for the moral law to bind me with normative authority, it must function as an incentive for me. Whilst the formal features listed above provide the complete intellectual/theoretical explanation for the bindingness of the moral law – one that any rational being whatever might recognise – it is not the full story. Something more must obtain in order for the moral law genuinely to apply to a rational agent as a constraint: its ability to motivate of itself. More on this in what follows.

validity of the moral law. The fact that God wills that I pursue my own perfection might serve to justify the principle of perfection to those who do not already acknowledge its validity, and in this sense it serves (according to the relevant moral theory) as an *external* ground. If I am to view the universal applicability of the categorical imperative as a reason for me to obey it and as a justification for its applicability to me, however, then I must already acknowledge its universal validity, and in this sense this formal feature can only ever serve as an *internal* ground (a ground that is internal to the moral point of view). In describing the moral law as supplying its own determining ground, Kant is not pointing toward an external justification of its normative force. The moral law's authority for us is non-derivative, but this is precisely to say that there is no external justificatory story available to us when the sceptic rears her head.

Kant criticises the theological view presented above for the reason that God's will can only function as a determining ground if we presuppose an antecedent end (agreement with this will). In order to distinguish his fundamental principle from those rationalist principles that turn out to be material, we must therefore conceive of the moral law as non-derivative in a second sense: in the sense that its ground (the form of universal lawfulness), does not derive its determining power from a more fundamental source elsewhere. Put differently, we must conceive of this ground as providing the *complete* story when it comes to the validity of this principle for a morally responsible rational agent. We need not assume, for example, that its determining force depends on our having an antecedent purpose or desire to be lawful (if such a desire is possible). When it comes to universal lawfulness, this thought alone – this

purely rational consideration – must be sufficient to explain why I ought to obey the categorical imperative.³¹

What this means is that, if there is to be a universally and necessarily binding practical principle, i) its ground must consist in mere universal lawfulness as such, and ii) this consideration must be a sufficient determining condition on its own, without any support from presupposed ends or attitudes. To the degree that this is Kant's view, we have arrived at one way of specifying the Rationality Thesis: the moral law *is* a law of practical reason insofar as it finds its source in a purely rational, non-derivative ground, which accounts entirely for its normative force and which establishes it as a law for those to whom it applies.

3.3 *The Legislation of Reason*

By investigating the sense in which the moral law can be viewed as a law of reason, we have located an account of the source of its authority, though only from the inter-

³¹ Strong hints that Kant at one time held a different (pre-autonomy) view can be found in the *Critique of Pure Reason's* Doctrine of Method, where he argues that, although '[i]t is necessary that our whole course of life be subordinated to moral maxims ... it would at the same time be impossible for this to happen if reason did not connect with the moral law, which is a mere idea, an efficient cause which determines for the conduct in accord with this law an outcome precisely corresponding to our highest ends, whether in this or in another life. Thus without a God and a world that is now not visible to us but is hoped for, the majestic ideas of morality are, to be sure, objects of approbation and admiration but not incentives for resolve and realisation, because they would not fulfil the whole end that is natural for every rational being and determined *a priori* and necessarily through the very same pure reason' (A 813/B 841). Here, the idea is that the moral incentive is dependent on some end, in the absence of which the moral law can indeed be an object of admiration, but not an incentive for 'resolve and realisation.'

nal point of view mentioned above. Where the Command Thesis posited that this source is to be located in a legislative act of the will, the above account suggests that what explains the authority of an unconditionally binding principle (why it counts *as a law for me*) is its universal form. A personification of the form of law into ‘reason,’ the legislatrix of the moral law, gives us a sense in which the mixed will can be said to give the law to itself. This version of the Rationality Thesis thus presents a direct alternative to the Command Thesis, and one which finds support in Kant’s writings.

With this said, however, we might still wonder whether this account really helps us to spell out the concept of Kantian self-legislation – for in what sense is this legislation in which I *myself* engage? In what sense does this account involve the idea of law-giving at all? In the Vigilantius lecture notes, Kant offers a very helpful explanation of how the idea of self-legislation might be brought to bear on the above account. As he writes:

Although the obligation is established by reason, it is nevertheless assumed that in the performance of our duty we have to regard ourselves as passive beings, and that another person must be present, who necessitates us to duty. Crusius found this necessitating person in God, and Baumgarten likewise in the divine will, albeit known through reason, and not positively, and on this principle a particular moral system has been erected. If, however, we pay heed to self-regarding duties, then man is presented in his physical nature, i.e., insofar as he is subject to the laws of nature, as the obligated, and rightly so; but if the obligator is personified as an ideal being or moral person, it can be none other than the legislation of reason; this, then, is man considered solely as an intelligible being, who here obligates man as a sensory being, and we thus have a relationship of man *qua* phenomenon to himself

qua noumenon. The situation is similar in obligations toward others. (LV 27:510)³²

The supposition that we are necessitated to action *by another* when it comes to duty is described here as something of a background thought that accompanies our experience of the moral law. Where Crusius located this other in God, this is not our only option, for reason herself may be personified and represented as an active legislating – that is, obligating – agent. This idea (the personification of ‘reason’) is equivalent to the idea of *oneself*, although ‘considered solely as an intelligible being,’ such that the idea of duty brings with it the idea of a relationship between oneself *qua* legislating, intelligible being to oneself, *qua* sensory being.

Here, the idea of self-legislation is presented as a highly abstracted metaphor. Kant introduces self-legislation in this context as a rhetorical device of sorts: as a certain way of representing the foundation of duty to ourselves. Where the notion of a legislating ‘self’ comes into the picture at all, the relevant concept is the self as a purely intelligible being, and so a concept which does not resemble in the least the self with which we are in any way familiar via observation, sensation or reflection. Given all that he has said here, we need not assume that there are actually two persons who stand in a relation of obligation to each other. Yet given all that we have said so far concerning the sense in which the moral law can be understood as a law ‘of reason,’ the great leap from talking about reason as a metaphorically ‘legislating’ ground of determination to talking about *ourselves* as legislators might seem a bit much.

³² See also my discussion of duties to the self, which begins on p. 24. The above passage is, of course, highly reminiscent of the standpoint discussion in *Groundwork* III.

Happily, there is a less abstractly metaphorical, more concrete sense in which we can describe the rational will (*our* own particular faculty of practical reason) as legislating the moral law. Whilst we might on the one hand conceive of ‘reason’ as an abstract domain of rational considerations (such as the formal features of universality and necessity), we are also able to speak of our own cognitive and volitional powers, by which we discern what we ought to do and find the motivation to pursue certain options, as belonging to the broad faculty of practical reason: to ‘the will’ in the broad sense of the term. As we saw in chapter one, Kant often distinguishes between two separate elements of agents’ wills: the legislative faculty of *Wille* and the faculty of choice, *Willkür*. The personification of reason need not only consist in the personification of an abstract realm of rational considerations, but might just as significantly consist in the personification of the latter: of that pure power in each of us by which we are able epistemically to determine the content of morality and from which are issued representations of its fundamental principle for the consideration of the faculty of choice.³³

Kant quite explicitly maintains that the common understanding experiences the world with the moral law ‘in hand,’ such that each of us ‘knows very well how to distinguish in every case that comes up what is good and what is evil’ (*G* 4:403). One way to interpret this notion is to suppose that moments of moral choice involve the explicit bringing to mind of the moral law itself, via a representation of pure reason, and indeed Kant’s language often suggests that this is his story. Thus in the second *Critique*, Kant speaks of consciousness of this law as something which ‘forces itself upon us of itself as a pure synthetic a priori proposition,’ by which reason ‘announces

³³ See Ameriks, *Interpreting Kant’s Critiques* 252.

itself as originally lawgiving' (5:31). Of course, to suppose that everyday practical cognition must involve an *explicit* and *conscious* representation of the moral law is to represent a very unfamiliar picture of practical reasoning, and in fact Kant's talk of the *a priori* representation of the moral law need not be interpreted as pointing to an explicit conception. As Kant makes clear, *philosophical* reflection on our standard moral beliefs yields an explicit articulation of their underlying principle, but we may quite reasonably suppose that the common understanding exhibits an inherent sensitivity to this principle and a respect for its authority even in the absence of a conscious representation of the law (just as a person is able to grasp the punch line of a joke without being able to explain the principles – should there be any – that ultimately ground its comedic force).³⁴

When observing this fact about common practical reasoners, and when asked how we might explain their inherent sensitivity to what philosophical reflection can reveal as a single, fundamental principle of reason, the idea of an implicit *a priori* representation of some sort – at times something of which we are conscious, at times not – becomes quite useful. Even if a person is unable to articulate the basic formula of the categorical imperative, it is central to Kant's thinking that she must at least have a clear and immediate grasp of the moral impossibility of certain courses of action: she must be able appropriately to *use* the moral law, to reason in light of its instruction and to appreciate its overriding authority, even if she is unable to report it explicitly. In order to grasp the moral impossibility of certain maxims, an agent must be able in

³⁴ I am indebted, for this example, to Onora O'Neill, 'Autonomy and the Fact of Reason in the Kritik der praktischen Vernunft,' *Kritik der praktischen Vernunft*, ed. Otfried Höffe (Berlin: Akademie Verlag, 2002) 82. For more on the implicit sensitivity of the common understanding, see G 4:403-404.

some sense to arrive at a representation of the *form* of lawfulness (or unlawfulness), for this is precisely the criterion by which moral distinctions are determined. What is more, if she is to have the chance of being morally good, she must in her judgement be able to appreciate the normative force of possible universality and to adopt this reason as the pure basis of her action. This complex cognition of a given maxim's being unfit for universalisation and the necessity of refraining from adopting it is nothing other than the representation of the necessity of universal lawfulness – a cognition which, on Kant's account, generates a brand new practical incentive.

If we turn to Kant's description of the sense in which God might be understood as the legislator, but not the author, of the moral law, we find that it has much to do with just the activity we have been considering: the representation of the law and the generation of an incentive. As we have seen, Kant suggests in the Collins lectures that 'anyone who declares that a law in conformity with his will obliges others to obey it, is giving a law' (27:282.36-283.1). Here, the idea is that a legislator of non-positive law is one who addresses a group of agents: one who represents the law to others and who pronounces that it is at the same time in conformity with his will. What is more, his declaration provides an incentive to obedience – one, it is important to note, that ultimately stems from a recognition on his part of the law's authority, and not one that *generates* this authority. In a limited sense, the legislator's declaration can be said actively to 'give law' to those whom he addresses, because it provides a connection, in the form of an incentive, between their wills and the law that he represents to them.³⁵ Put differently, his activity (representing the law as a law in agreement with his will) generates a motive to comply that his audience can latch on to, providing

³⁵ This need not be the only incentive or connection that there is. What the lawgiver in this context ensures is that there is at least one such incentive.

one avenue by which that law might be *practically* relevant for them. His command thus ensures that the law is a law *for* his audience – one that can be chosen, or effective. In a very limited and specific sense, he is an ‘agent of obligation’ (one who ‘obliges’), insofar as his commanding activity can be thought of as a means by which those who are subject to the law are able to grasp its authority and to be motivated to act as it requires.

Now if we think of the representation of the necessity of universal lawfulness described above as stemming from a particular faculty, the personification of this faculty as a legislator *in just this sense* may easily be constructed. In a metaphorical sense, pure reason itself, as the faculty by which representations of universal lawfulness and its attendant necessity are made possible, may be thought of as ‘addressing’ the faculty of choice in precisely this manner, and so may be represented as engaged in the activity of giving law (though only in this limited, non-authorial sense). To the degree that this faculty is *mine*, a part of my own cognitive make-up, I can represent the moral law as a principle that I ‘give’ to myself. My own will’s representational activity is the means by which I grasp the authority of the moral law and am offered a motive to comply with it (in this case, one that is internal to the law itself and not external, as would be the case where God functions as this kind of legislator). In this sense I am the ‘agent’ of my very own obligation.

In the Doctrine of Method of the first *Critique*, Kant proposed that God must play an indispensable motivational role when it comes to morality. By the time of the *Groundwork*, however, he had abandoned this view for one according to which the motive to comply with the moral law is located squarely in the human will, whatever other external motives there might be, and in particular in its representation of the

moral law to itself. What is more, Kant makes clear throughout his writings that being obligated by the moral law and having this direct motive go hand in hand. In an important footnote to the first section of the *Religion*, Kant asks us to imagine a hypothetical rational being who, although he is able to grasp that this or that maxim conforms to universal law, has no motivation to act on this information.³⁶ Because the moral law does not ‘announce itself’ as an incentive for this creature, he needs incentives of inclination in order to act. He can indeed grasp the fact that a maxim is either morally allowed or morally forbidden, since there is nothing wrong with his cognition of the moral law, but because he cannot be motivated to act on this information, the cognition of the moral law is nothing more than an intellectual exercise for him. Put differently, although he is able to see that a certain course of action is morally required, he is unable to connect this purely rational consideration to his will. Yet as Kant points out, to the degree that he is unable to act in genuine responsiveness to the moral law, he cannot really be held accountable for following it. Where the moral law really obligates, the motivation to act in accordance must be available, for in the absence of this special kind of activity, the moral law would simply not be a law *for us*. As such, the will’s motive-offering activity is of central importance in the broader story of moral obligation.

Of course, to say all this is not to say that the will’s motive-offering activity in any way ‘confers’ the status of law onto the categorical imperative in the first place (or that it ‘creates,’ ‘generates,’ or constitutes the ‘source’ of the authority of the moral law). As we have seen, the categorical imperative is non-derivative, in the sense that it does not draw its authority from (or find its ‘source’ in) anything but itself. It is true

³⁶ See *Rel* 6:26n.

that we would not be obligated by the moral law if we did not engage in this particular kind of activity (the representation of the law and the subsequent offering to ourselves of the moral motive), but this is only because this activity allows us to count as the kind of being for whom the moral law can be effective: as the kind of being for whom obeying the moral law can really be a practical option. All that Kant's thoughts on the hypothetical rational agent described above imply is that, where there really is an obligation, there must be a motive to comply. They do not imply that we must think of the will's activity as that which *confers* normative authority onto the law. Indeed, that the activity of motive-generation does not confer normative authority onto the moral law on Kant's picture is ensured by the fact that, on this view, the internally generated moral motive *results* from our recognition (or representation) of the moral law's overriding and non-derivative authority. Yet this is just to say that the will commands that the faculty of choice follow the moral law *because* the latter is absolutely authoritative; the moral law is not absolutely authoritative because it is commanded.

With this more concrete interpretation of self-legislation in hand, we can incorporate the significance of aspects of the Command Thesis without supposing that our engaging in this active representation of the moral law is what *makes* it binding for morally accountable rational agents. The legislation story of §3.2, which hinges on the personification of reason conceived as the ground of morality, constitutes one way of describing the source of the moral law and its binding force. When thinking of legislation in this sense, we see how autonomy explains why the moral law is valid for a responsible rational being. By contrast, the legislation story offered in this section provides a way of expressing the kind of activity in which everyday moral agents engage when they determine what they ought to do, and thus it is tied much more

closely to moral epistemology and motivation. It offers a picture of an activity of pure reason in relation to obligation which does not reduce to creation or construction. Where the account given in §3.2 focuses on the reason *why* a morally accountable rational being is bound by the moral law, the account offered in this section focuses on the role that reason's activity plays in generating a motive to comply with the moral law – one which explains another sense in which we might be said to 'give' law to ourselves and to be the agents of our own obligation. In the end, both the Rationality Thesis and the Command Thesis, suitably modified, contain elements that help us to understand what Kant meant by self-legislation, though we must be careful to appreciate the distinctness of their contributions.

3.4 From Autonomy to the Dualism of Practical Interest

In contrast to the radical individualism and concern with self-expression that characterises some interpretations of Kantian autonomy, I have here been offering a more minimal and sober account of the moral law's status as a law of reason. The story traced thus far develops threads from both the Command and the Rationality Theses outlined in §1.1, and I hope to have made evident its power to withstand, in particular, the deep concerns affecting the former. Yet this more modest, rationalist account raises important concerns of its own, some of which we have already identified (with reference to the Rationality Thesis) in §1.5. Attention to these complications will be the focus of chapters four to six.

I argued in §3.1 that one of the more important reasons for rejecting the constitutivist model is the fact that it leaves no room for *clearheaded* moral wrongdoing. Yet as we

will see in chapter five, one implication of the view I am offering is that there is no space in Kant's ethics for truly *innocent* deviation from the moral law. The epistemological and motivational elements of autonomy entail that the moral option is always something that we represent to ourselves (where such an option is in play) and the authority of which we can always grasp, even if we do not act on it. This fact brings with it important implications for Kant's conception of character, attention to which will reveal the deep connections between this aspect of autonomy, a particular species of moral rigorism, and Kant's denial of moral luck.

Prior to exploring these issues, however, it is important to trace an equally interesting implication of the first element of autonomy mentioned in this chapter, which if we recall is the fact that the moral law does not have any ground beyond its universality and necessity. In chapter four, we will be exploring the connections between this feature of the moral law and its content (what the law tells us to do). As I will argue, this in turn entails what is best described as a strict *dualism* when it comes to practical concern – a principle which is connected to an important worry about the extent to which obedience to morality on Kant's view might entail an unpalatable dichotomisation of our interests and an unnatural denial of our affective nature.

MATERIAL PRINCIPLES AND THE FORM OF LAW

In the preceding chapter, I argued that Kant is best interpreted as holding that the moral law provides its very own determining ground. To the extent that the latter is located in its mere lawfulness rather than any presupposed interest, the moral law is a *formal* principle of reason, to be contrasted with both rational-material and empirical-material norms. In this chapter, I consider the implications of this aspect of the autonomy of practical reason for Kant's conception of what we might call the 'content' of morality (what the moral law commands us to do). In particular, my aim is to become clearer on the relationship in Kant's ethical thinking between this aspect of autonomy and the role reserved for natural desire (inclination) in morally good maxim selection.

Because of his strict separation of form and matter, Kant is often interpreted as proposing that the moral law stands in necessary conflict with naturally grounded material principles.¹ The autonomous agent is in some sense a divided creature, in a

¹ That Kant's ethics often seems to be an ethics of repression of what is natural to us can be partially accounted for by the fact that he sought primarily to separate elements that had been

constant struggle to maintain the proper order between her affective and her rational nature. As a result, it can seem as though the central Kantian moment consists in a species of self-denial – of subordination of concern for happiness to the moral incentive provided by reason’s very own law. Of course, this division ought not to be exaggerated. Kant allows that the harmonizing of inclination and reason is genuinely to be welcomed and pursued.² What is more, Kant’s picture is one according to which it is common that natural ends and purposes will be satisfied both in action which accords with and in that which is done from duty.³ Yet even with these points noted, it is central to Kant’s view that moral agents ought utterly to set aside considerations stemming from their sensuous nature wherever duty is at stake, and for this reason concerns about whether there is any real space in his ethics for the natural side of our being are not completely unfounded.

The chapter consists of four sections. First, I trace Kant’s derivation of the categorical imperative as it appears in Sections I and II of the *Groundwork*. As I argue, this derivation hinges on the intimate connection between a conception of the moral law as a formal law of reason, as discussed in §3.2, and the command that we consider only the form of our maxims. I then go on in §§4.2 and 4.3 to consider two recent interpretations of Kant’s conception of moral interest, which seek to undermine the tradi-

identified as morally fundamental by other philosophers (Beck, *A Commentary on Kant’s Critique of Practical Reason* 120).

² See, e.g. *Rel* 6:23n. See also *KpV* 5:93: ‘But this *distinction* of the principle of happiness from that of morality is not [...] at once an *opposition* between them, and pure practical reason does not require that one should *renounce* claims to happiness but only that as soon as duty is in question one should *take no account* of them.’

³ Think, for example, of the sympathetic soul and the honest shopkeeper of the *Groundwork* 4:397-98.

tional denial of a morally relevant role to desire in moral choice. I argue that, given the conclusions of §4.1 (along with other substantial worries), such approaches must be rejected as plausible interpretations of Kant's view. Kant's dichotomy between reason and desire, between form and matter, may be strict; indeed, it raises quite interesting worries, considered in chapters five and six, concerning the place within his moral theory for notions such as character and development. Given the relationship between this strict dichotomisation and a conception of the moral law as a genuine law of reason, however, we abandon the former only at the peril of remaining inconsistent with the latter.

4.1 Form and Matter

As we saw in §3.2, Kant conceives of the determining ground of the moral law as lying solely in its form. The argument for this thesis hinges heavily on the claim that no material principle – no principle which presupposes an object of desire (matter) as the determining ground of the will – can furnish a practical law. Where the determining ground of some principle is an object, whether empirically or rationally cognisable, the reason why it is a law must in the end reduce to the idea that following that principle allows us to achieve the end in question. In this case, its normative force for us holds only on the condition that we have adopted that end (that we have an interest in that object). The Epicurean view, for example, presents an empirical object – happiness – as intrinsically good, and the relevant rule of conduct is then derived from it. On this system, however, the Epicurean principle can be conceived as a rule for an agent only to the degree that the represented object bears a certain relation to her psychological disposition: only if she happens to take an interest in it. Where this re-

lation obtains, its occurrence is a matter of contingent fact, in which case the rule cannot be said to possess the strict necessity and universality required by the concept of a law.⁴ Because of this, no principle of pursuing a desire-based object can play the role of a practical law.⁵

Since no material can serve as the determining ground of a practical law, Kant's conclusion is that pure reason 'must be practical of itself alone': that is to say, it must be able to determine the will simply by the form of a practical rule, without presupposing any empirical condition (*KpV* 5:24, 27). The moral law is formal (in this specific sense) because it presupposes no end that an agent intends to achieve.⁶ In fact, it determines the end that agents ought to aim at in their maxims. We need not appeal to any object (matter) of the will in order to explain why it is binding; we must simply consider the idea of a mere law (a categorical imperative as such) and what it entails.

⁴ See also the discussion of the object of reason in §2.3.

⁵ For an articulation of the claim that *only* a formal law can be an *a priori* determining ground of practical reason, see *KpV* 5:64.22-25. Importantly, and as we saw in §3.2, the conclusion that no material principle can play the role of a practical law must hold equally for those whose material is an object that is represented by reason or the understanding (as opposed to the senses). According to Kant, '[i]f a representation, even though it may have its seat and origin in the understanding, can determine choice only by presupposing a feeling of pleasure in the subject, its being a determining ground of choice is wholly dependent upon the nature of the inner sense, namely that this can be agreeably affected by the representation. However dissimilar representations of objects may be – they may be representations of the understanding or even of reason, in contrast to representations of sense – the feeling of pleasure by which alone they properly constitute the determining ground of the will (the agreeableness, the gratification expected from the object, which impels activity to produce it) is nevertheless of one and the same kind ...' (*KpV* 5:23).

⁶ Kant appeals to the language of a 'merely formal law' (as opposed to a material practical rule) at *KpV* 5:22.29. See also *G* 4:400.8-16.

Both in sections I and II of the *Groundwork*, Kant moves from thinking about the nature of an absolute and formal practical law of this sort (a law that presupposes no particular end) to thinking about its content when addressed to finite rational beings: about what this law *commands*.⁷ His intention at this point is to discover what a categorical imperative commands by investigating its ‘concept.’⁸ The hope is not to derive a conception of *token* commands simply by examining the concept of a moral law, but rather to arrive at a conception of the fundamental or *general* law by which agents are able to determine in more concrete terms what they ought to do in particular circumstances.

Kant’s conclusion on this front is of course that the categorical imperative directs us to ‘[a]ct only in accordance with that maxim through which you can at the same time will that it become a universal law’ (*G* 4:421; emphasis changed).⁹ This general for-

⁷ In the discussion to follow, the focus is the content of the categorical *imperative* in particular, which must be distinguished from the moral law. The latter applies to every rational being as such, whereas the former is the form in which the moral law presents itself to beings like us, who have interests that might conflict with morality.

⁸ As he puts it, his task at this point ‘is to enquire whether the mere concept of a categorical imperative may not also provide its formula containing the proposition which alone can be a categorical imperative’ (4:420.18-21). See also 4:420.26-7: ‘...when I think of a *categorical* imperative I know at once what it contains.’

⁹ See also 4:402: ‘I ought never to act except in such a way that I could also will that my maxim should become a universal law.’ As Kant makes clear in Section II, there are two kinds of failure that might affect a maxim in this regard: ‘Some actions are so constituted that their maxim cannot even be *thought* without contradiction as a universal law of nature, far less could one *will* that it become such. In the case of others that inner permissibility is indeed not to be found, but it is still impossible to *will* that their maxim be raised to the universality of a law of nature because such a will would contradict itself’ (4:424). Those actions which cannot be thought without contradiction are the ones that strict duty requires us to avoid, while those that violate the second kind of consistency test are opposed to wide duty (4:424). In this characterisation, Kant has already identified the first variant of the general formula,

mula, to which Kant eventually assigns three variants, asks us to consider what we *can* will with consistency as a universal law, not what we *do* will or *would* in certain circumstances will as a universal law. Because of this, it makes no reference to what agents happen to want (or would happen to want).¹⁰

To universalise is to consider whether what one proposes as a maxim for oneself could be adopted by all others. It is to consider whether we can simultaneously adopt the maxim and will that all others do so as well. As a result, the universality test does not ask us to focus on the content of our maxims, but rather on the *mutual consistency* of sets of maxims.¹¹ When employed in our practical reasoning, it indicates whether we are choosing to act in a way that requires or presupposes that the same choice be denied to some others. Though it does not ground Kant's derivation of this formula, the intuitive thought behind the claim that a universalisability condition provides a canon of moral permissibility is the idea that morally bad action involves singling oneself out for special treatment. With this noted, however, we might nonetheless wonder how Kant moves from the mere concept of a categorical imperative to this particular prescriptive formula.¹² In this regard, commentators often suspect that there is a gap in the argument, attributable to an unbridged leap from the thought that the moral law is universally binding to the thought that we ought only to act on

according to which we are instructed to act only 'as if the maxim of [our] action were to become by [our] will a universal law of nature' (4:421).

¹⁰ As we will see, this is a very significant point.

¹¹ O'Neill, *Constructions of Reason* 103.

¹² Kant sketches his derivation of this principle in the *Groundwork* at 4:402 and 420.

maxims that can be willed as universal laws.¹³ It is one thing, after all, to propose that we ought to conform our maxims to whatever universally binding laws might apply to us, and so to act in accordance with law, but quite another to claim that our maxims must *themselves* be universalisable, or willable *as law*.¹⁴

Kant's critics often turn to two rival principles as candidates for 'universal law': a principle prescribing the impartial maximisation of utility, and an egoistic principle directing agents to pursue their own best interest.¹⁵ That we can eliminate these rivals should now be clear, however. The egoist's principle is not actually a contender at this stage in Kant's argument precisely because in cases where this principle is a plausible practical principle at all, its normative force can be traced to the desirable *effect* to be brought about by acting on it (e.g. the satisfaction of one's wants, the achievement of one's objective good). Where we think that this principle might rea-

¹³ Note that this is distinct from the charge of emptiness and lack of action-guidance that more commonly plagues the general formulation of the categorical imperative. (For a compelling defence of the principle's ability to yield action-guiding content in the absence of heteronomous considerations, see O'Neill, *Constructions of Reason* ch. 5.)

¹⁴ See Stephen Engstrom, *The Form of Practical Knowledge* 6. See also Paton, *The Categorical Imperative* (Philadelphia: University of Pennsylvania Press, 1971) 72; Bruce Aune, *Kant's Theory of Morals* (Princeton, NJ: Princeton University Press, 1979) 29-34, 86-90; Allen Wood, *Kant's Ethical Thought* 78-82 and *Hegel's Ethical Thought* 163-7; and Henry Allison, *Idealism and Freedom* (Cambridge: Cambridge University Press, 1996) 143-54. The whole debate is nicely discussed in Samuel Kerstein's *Kant's Search for the Supreme Principle of Morality* (Cambridge: Cambridge University Press, 2002) 73-94.

¹⁵ What would seem to make these two principles attractive candidates is the fact that neither depends *immediately* for its validity on the desires that an agent happens to have. If the utilitarian principle applies at all, we suppose that it does so even if an agent does not happen to want to maximise the universal happiness, in the ordinary sense of 'wanting.' Similarly, if the egoistic principle is valid, it is natural to suppose that it must be so independently of the agent's desiring his own wellbeing, in the ordinary sense of 'desiring' something. As I have been arguing, however, genuine categoricity and universality depend on more than satisfying this specific criterion.

sonably serve as a norm, we do not suppose that selfishness is *itself* normative; rather, we suppose that a principle of selfishness might be a good one to follow *because* of the goodness of getting what one wants.¹⁶ As such, the principle is grounded in the value of obtaining this effect or end, and so constitutes a material rule. *Given* that this effect is desirable, one ought to act only in one's own interest. To the degree that the principle finds its determining force in some object of the will, however, it can never be *categorical* in the sense relevant to Kant, for its validity is conditional on the goodness of getting what one wants. The egoistic principle may indeed possess a sort of universality, in the sense that it may be valid to those for whom this effect really is good or desirable, but its universality is ultimately *conditioned* by this restriction.¹⁷

Similar considerations show that the principle of impartially maximising utility also fails to count as a universally binding law. Here, again, the normative force of the principle directing agents to maximise utility can only stem from the goodness or desirability of its effect: the achievement of maximal universal happiness. On the utilitarian view, it is the independent value of universal happiness that grounds the normative force of the principle directing us to its realisation.¹⁸ If this is so, however,

¹⁶ As Jens Timmermann points out, '[a]n egoist acting for the sake of – rather than in mere conformity with – the *principle* of selfishness would be a philosophical curiosity and, arguably, not even selfish in any recognisable sense at all' (*Kant's Groundwork of the Metaphysics of Morals: A Commentary*, 74n.49).

¹⁷ Note that, on Kant's view, the goodness of getting what one wants will always be *conditional* on the possession of a good will (see *G* 4:393-4).

¹⁸ Again, Timmermann points out that it is the happy effect of acting on the utilitarian principle that is picked out by the theory as good (rather than conformity to this principle as such). As he asks, 'could a thoroughgoing consequentialist really applaud actions for the sake of a utilitarian law if they did not realise the (indirectly) intended effect?' (*Kant's Groundwork: A Commentary* 74).

then the utilitarian principle must be considered instrumental, its universality conditioned by the goodness of universal happiness.

The universality of rules which direct agents to the realisation of a specific end – even an end from which an agent cannot arbitrarily release herself, such as happiness or perfection – is restricted precisely by that end. Because Kant realises that no such law can constitute a *truly* categorical (and so unconditioned) imperative, he shifts our thinking away from the idea of a specific material rule – one that picks out an end as something to be pursued – altogether. As he writes:

...since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary. (*G* 4:421)

Beyond the law represented by reason as applying universally, the imperative for which we are searching contains *only* the necessity of conforming to this universal law. If it were to contain more – a material component (a prescribed object of the will) – then its universality and necessity would be conditioned by the value of this component, in which case it would not count as a categorically binding law. Yet precisely because the moral law can contain no material component of this sort, it cannot command us to realise specific states of affairs.¹⁹ As a result, nothing is left to which our maxim must conform except ‘the universality of a law as such.’

¹⁹ *If*, that is to say, there really is such a law that applies to us. This question is left open until Section III of the *Groundwork*.

What does it mean to *conform* my maxim to universality as such, and what does this have to do with being able to will that it should *be* a universal law? The supposed gap in the argument can be bridged by allowing a key starting point, stemming from Kant's moral psychology.²⁰ As Kant suggests in an important footnote at *G* 4:420, action on maxims of any kind aims to bring subjective norms into agreement with objective ones. As Kant puts this idea, every maxim 'contains' a practical *rule* that conforms in some sense with the 'conditions of the subject.' When it comes to action on hypothetical imperatives, we aim to conform our maxim to the (objective) laws of nature – to the regularities that we discover in experience – in order to realise an end represented as good. A desire presents itself to me, and I ask: Is my proposed action really the kind of thing that is *generally* and *regularly* conducive to my end? To ask this, however, is parallel to asking whether my maxim (to perform *this* kind of action for the sake of *this* end) could be a *rule* for everyone who happens to share my end. Here, regularity in nature when it comes to the connection between means and ends brings us easily to the notion of a rule or general (in this case, empirical) law; because of this, even action on hypothetical imperatives involves a universalisation test of sorts on Kant's view – one according to which universality is assessed relative to a specific group of agents.²¹

²⁰ I am indebted for this interpretation to Jens Timmermann, *Kant's Groundwork: A Commentary* 74-5.

²¹ Andrews Reath explores a similar continuity between non-moral and moral reasoning in 'Kant's Conception of Practical Rationality,' *Agency and Autonomy in Kant's Moral Theory* esp. 71-84. According to Reath, non-moral choice involves taking our reasons to justify our actions to those who share our ends, whereas moral choice involves taking our reasons to justify our actions to any agent, however situated. The formula of universal law is then presented as a procedure for determining when our reasons are able to justify our actions to any agent. Because Reath relies so heavily on the language of *reasons*, however, it is difficult to

In action which complies with a categorical imperative, however, universality is unrestricted. This much is ensured by the conclusion that such an imperative must be a 'law of reason' in the sense described above. In this case, rather than asking whether my maxim can serve as a rule for everyone who happens to share my end, I ask whether my maxim can serve as a rule for every rational agent *as such*, regardless of whatever needs, desires and ends might be in play. Reasoning in accordance with a hypothetical imperative requires that we determine whether there is a sufficient match, cashed out in terms of natural regularity, between an action and an end; it thus requires that we focus our attention on the nature of the objects of desire and on the empirical regularities associated with the means to their realisation. *Given* a particular end, we ask whether our maxim possesses the formal feature of possible law-likeness. In moral reasoning, however, our concern is *purely* formal. We ask simply whether our maxim contains the formal feature of being a possible law for all rational agents, and we do not refer at all to the nature of any end or to observed regularities. Thus there is a second sense in which the moral law may be conceived as a 'formal' principle, for it directs us to consider the mere form of our maxims rather than anything having to do with their matter.

This derivation can be questioned simply by bringing into doubt the psychological starting point on which it hinges. Moreover, we might question whether the connection between acting on a maxim that can serve as a universal rule and acting on a maxim that can be *willed* as a universal law is really as straightforward as I have so

make out the supposed connection between acting for reasons which justify universally and acting on maxims which can be willed as universal *laws*. Although I am sympathetic with the broad contours of Reath's analysis, I hope that my reconstruction has made this connection clearer.

far made it sound. Even with these worries acknowledged, however, what we can say for sure is that the categorical imperative, whatever its specific content, cannot possibly direct agents to consider the value of actions relative to ends. As we have seen, the universality of a rule that directs agents to the realisation of a specific end is conditioned by that end. Because a categorical imperative must be unconditioned (see §3.2), no such rule can constitute a categorical imperative. Yet if moral choice does not involve reasoning which considers relations of fit between objects of desire, or the matter of our maxims, it is natural to suppose that it must instead involve consideration of their formal features. What allows us to bridge the supposed gap in Kant's derivation of the categorical imperative is the fact that the moral law is a formal law of reason, in the sense that it does not presuppose any desire as a condition of its validity. The formal character of the moral law in one sense – the fact that it directs us to consider only the form of our maxims – is thus intimately related to the formal character of the law as discussed in chapter three.

4.2 Duty and Desire

The above helps to explain the importance that Kant places on setting aside considerations having to do with the relative value of states of affairs when duty is at stake. On Kant's view, action on moral grounds expresses a direct interest in our action's conformity to universality as such, and so in a very particular formal property of our maxim. By contrast, action on non-moral grounds expresses an ultimate interest in the object or effect to be realised by one's action. From this point of view, the act is valued instrumentally, as a means of securing some further end. Thus, according to

Kant, there is a strict difference between taking an interest in an action and acting *from* interest, or from inclination:

The first signifies *practical* interest in the action, the second, *pathological* interest in the object of the action. The former indicates only dependence of the will upon principles of reason for the sake of inclination, namely where reason supplies only the practical rule as to how to remedy the need of inclination. In the first case the action interests me; in the second, the object of the action (insofar as it is agreeable to me). We have seen in the first Section that in the case of an action from duty we must look not to interest in the object but merely to that in the action itself and its principle in reason (the law). (*G* 4:413-14.n)²²

Where pure reason directs us only to consider the universal validity of our proposed maxim, inclination – which relies on reason put to its empirical use – directs us to satisfy the interest that we take in an object. It is true that on Kant's account all action involves, in addition to a law, an object of volition. The crucial factor, however, is whether the agent gives priority to the formal or to the material element of her maxim in her choice. Where the agent gives unconditional concern to the form of her maxim and acts precisely from that concern, her action is morally good; where her concern rests on the object to be realised by so acting, it is not.

Because the moral law directs us only to consider the form of our maxims, inclination must stand aside where duty is at stake. As Kant writes,

²² See also *G* 4:459-60: 'Reason takes an immediate interest in an action only when the universal validity of the maxim of the action is a sufficient determining ground of the will. Only such an interest is pure. But if it can determine the will only by means of another object of desire or on the presupposition of a special feeling of the subject, then reason takes only a mediate interest in the action, and since reason all by itself, without experience, can discover neither objects of the will nor a special feeling lying at its basis, this latter interest would be only empirical and not a pure rational interest.'

the human being is not ... required to *renounce* his natural end, happiness, when it is a matter of complying with his duty ...; instead he must *abstract* altogether from this consideration when the command of duty arises; he must on no account make it the *condition* of his compliance with the law prescribed to him reason; indeed he must, as far as possible for him, strive to become aware that no *incentive* derived from that gets mixed, unnoticed, into the determination of duty ... (TP 8:278-79)²³

This conclusion is a direct consequence of accepting that the moral law must be a categorical imperative, and so a principle which is not conditioned by any object. Far from stemming from a deep-seated disdain for the inclinations, Kant's rejection of the moral worth of action from inclination finds its source, in large part, in his understanding of the unconditioned character of moral command and of what the moral law tells us to do.

With this said, however, many of Kant's most sympathetic interpreters find this insistence on purity when it comes to the concerns of moral deliberation either unpalatable or unrealistic.²⁴ There has thus been a move in recent criticism to find ways of accommodating a much more intimate partnership between the concerns of pure reason and those of inclination in a Kantian account of acting on the categorical imperative. Because two of these approaches in particular constitute such glaring deviations from the interpretation I have been offering here, it will be useful to consider them in some detail.

²³ See also *G* 4:400: 'For, the will stands between its a priori principle, which is formal, and its a posteriori incentive, which is material, as at a crossroads; and since it must still be determined by something, it must be determined by the formal principle of volition as such when an action is done from duty, where every material principle has been withdrawn from it ... Now, an action from duty is to put aside entirely the influence of inclination and with it every object of the will ...'

²⁴ I will be addressing the 'unrealistic' charge in chapter six.

(i) *Rules of Moral Salience*

One influential attempt to reconcile inclination with the categorical imperative focuses on the role that the development of our affective nature plays in the structuring of moral attention. On this view, part of our normal development as embodied rational beings involves the development of patterns of feeling and desire, which in turn results in the acquisition of what Barbara Herman has called ‘rules of moral salience’: rules that structure our perception of the world, bringing into relief the specifically moral features of our situations.²⁵ The key idea, here, is that human beings must be trained to view their situations in terms of their morally salient features if they are to be able correctly to apply moral principles to particular situations; the development of patterns of desire, which shape our attention, effects just this sort of training.

Given that the moral law is a formal principle of reason, however, directing us to consider only whether our maxim can serve as a universal law, it is difficult to determine just what kind of *moral* work these acquired ‘rules of moral salience’ are meant to do. It may of course be true that agents must apply the moral maxims they adopt, determining how they are to implement the commands of pure reason in their specific contexts, and in this respect attention to various salient features of their environments might well be incredibly significant. Yet when it comes to determining the more fundamental question of which maxims we should act on (and which we must not), a

²⁵ See esp. Herman’s *Practice of Moral Judgement* (Cambridge, MA: Harvard University Press, 1993) 73-93. This claim is echoed in Nancy Sherman’s account of the significance of emotional development for morality. On her view, emotions can serve a significant role ‘as modes of attention that help us to track what is morally salient in our circumstances, and thus locate possible moments for morally permissible and required actions’ (*Making a Necessity of Virtue* [Cambridge: Cambridge University Press, 1997] 145).

genuinely Kantian position must hold that all that agents require is the capacity to test their proposed maxims against the categorical imperative, and this common human reason can do no matter what the nature of the agent's affective disposition.²⁶ Since the moral law directs us only with respect to the selection of maxims – that is, since the selection of maxims is the only truly *morally* relevant moment on Kant's strict view – the moral importance of rules of salience can only ever be secondary.²⁷

(ii) *The Dispersal of Duty*

Rather than focusing on the way in which desiderative states might serve as moral aids, a second attempt to accommodate the significance of inclination in morally good action seeks to develop the claim that inclination and duty need not be conceived as separate incentives, such that action in which the agent takes an interest in an object of desire might in certain cases satisfy the conditions of being morally good.²⁸

²⁶ 'I do not, therefore, need any penetrating acuteness to see what I have to do in order that my volition be morally good. Inexperienced in the course of the world, incapable of being prepared for whatever might come to pass in it, I ask myself only: can you also will that your maxim become a universal law? ... [C]ommon human reason, with this compass in hand, knows very well how to distinguish in every case that comes up what is good and what is evil, what is in conformity with duty or contrary to duty, if, without in the least teaching it anything new, we only, as did Socrates, make it attentive to its own principle...' (G 4:403-4).

²⁷ For more on this kind of observation, see Patrick Frierson, *Freedom and Anthropology in Kant's Moral Philosophy* (Cambridge: Cambridge University Press, 2003) 73-4.

²⁸ Although the view discussed below has much in common with positions according to which the motive of duty ought to be read as a 'limiting condition' (a second-order motive which requires that actions prompted by more basic desires satisfy the condition it lays down), it constitutes an important development of this view in its own right. For more on the 'limiting condition' or 'backup motive' conception of the motive of duty, see Herman, 'On

For some, a pure interest in the formal character of our maxims – their possible suitability as universal law – will require a hefty amount of explanation. On what has come to be known as a traditional Humean view of moral motivation, natural desires generated under quite standard and familiar social conditions can act as the ultimate springs to activity that is generally recognised as morally admirable. Sadly for the Kantians, however, the motive of duty's strict separation from the empirical elements of human nature makes its origins altogether more mysterious. Where could this non-empirical motive possibly come from?

In response to this kind of worry, it may seem tempting to reinterpret the traditional Kantian story so as to allow for the idea that the cultivation of moral interest is something that can be traced through a naturalistic psychological story of normal human development. Something along these lines has recently been proposed by Barbara Herman, who suggests that certain of our motives, though they develop and 'draw content' from original natural desires, can also and at the same time develop in response to the requirements of practical reason. Because of the fact that natural desire can be shaped by our developing rational (and moral) capacities, she argues, Kantians need not endorse a strict opposition between inclination and reason.²⁹

This approach begins with the observation that we are by nature disposed to have certain desires, which can be developed and modified in accordance with reason.

the Value of Acting from the Motive of Duty,' *The Practice of Moral Judgement* 1-22; Paul Benson, 'Moral Worth,' *Philosophical Studies* 51 (1987): 365-382; and Richard Henson, 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action,' *Philosophical Review* 88 (1979): 39-54.

²⁹ This view is developed in Herman's 'Making Room for Character,' *Moral Literacy* 1-28.

Something like this idea is traced by Kant in the second *Critique's* discussion of the so-called 'higher' faculty of desire, where he acknowledges that certain of our desires do indeed fall under our control as opposed to others. Thus it is claimed, he writes,

that we can find satisfaction in the mere *exercise of our powers*, in consciousness of our strength of soul in overcoming obstacles opposed to our plans, in cultivating our talents of spirit, and so forth, and we correctly call these joys and delights *more refined* because they are more under our control than others, do not wear out but rather strengthen feeling for further enjoyment of them, and while they delight they at the same time cultivate. (5:24)

The metaphor here is of the *refinement* of desire, a gradual removal of impurities from the rough stock with which we originally find ourselves. Where Herman's view moves beyond Kant's observations is in the proposal that the moral law itself might play such a refining role, shaping and forming our desiderative make-up. On her view, we can think of the moral interest as something which is liable to be 'dispersed' in the naturally originating motives that have developed in response to the agent's representation of reason's law.³⁰ As we develop into rational adults, '[d]esire for drink becomes a desire for safe and pleasant (not dangerous or unpleasant) drink, and also a desire for an available (not otherwise possessed) drink'; 'my desire for a new computer or car does not (because it ought not) range over those already owned by others.'³¹ Though such motives may in the end yield the very same actions one would have performed for the sake of the original desire, the internal responsiveness

³⁰ *Moral Literacy* 21.

³¹ *Moral Literacy* 15-16. One question for Herman, here, is why we should think of the constraint of only wanting not otherwise possessed drink or not otherwise owned computers or cars as being responsive to a particularly *moral* principle. On the face of it, these might equally well be prudential constraints, having nothing to do with morality *per se*.

of the desire to the constraints of rationality means that the maxim is relevantly different.³²

Thus according to this view, action which expresses moral interest need not be conceived as grounded in an isolated spring to action, the genealogy of which utterly transcends the natural. Because desires can be developed in reason-respecting ways via their integration with moral principle, the morally good agent can be moved to action by grounds which are quite intimately connected to the original springs of inclination. This allows us to say more than the familiar claim that there need be nothing morally *wrong* with the mere presence of empirical desire alongside the moral motive. According to Herman, what we are more interestingly able to claim on this picture is that ‘[t]he evolution of desire is in this way the condition for having a moral character that does not necessarily involve the segregation (or oppression) of one’s affective life in general by one’s commitment to morality.’³³

4.3 *The Heterogeneity of Moral and Non-Moral Concern*

Although this trend in interpretation has its obvious attractions, we would do well to consider it in light of the conclusions of §4.1. In particular, it is important to ask whether a natural motive (desire) can ever actually be said to draw content from a formal law of reason. On Herman’s view, the formation of a motive involves an ev-

³² *Moral Literacy* 20. On Herman’s alternative, motives of self-interest can be completely rational. To the extent that a person’s desires are evaluated in light of principles which include the moral law, her self-interest is not what it would have otherwise been (23).

³³ *Moral Literacy* 20-21.

alulative process by which certain desires are complicated and conditioned. The raw material of our affective nature is transformed by its coming into the relevant kind of contact with rational principles, such that only a more complex version of the former remains. While this is the story, however, what does it mean for a desire to ‘respond’ appropriately to moral principle? Can a coherent account of such a phenomenon be found?

In a straightforward sense, familiar from Section I of the *Groundwork*, the *actions* undertaken by agents as a means of satisfying their desires can indeed be said to ‘conform’ to the moral law. The prudent shopkeeper of 4:397 acts on a desire to secure his own advantage, but the action that this desire happens to recommend in his situation – the necessary means toward his end in this very case – conforms to the requirement of duty (he acts honestly). Similarly, the sympathetically attuned soul of 4:398 acts on an immediate desire to help others, and again the action that this desire happens to recommend conforms in outward appearance to the moral law. One way to interpret the idea of a desire’s being ‘shaped’ by the moral law might thus involve viewing morally refined desires as those which recommend actions that conform to duty.

If this is the picture that Herman is suggesting, however, it faces serious problems. First, it is clear that any coincidence between an object of desire and a recommended action can only ever be *external* and highly contingent. Desires and the acts that might bring about their satisfaction do not lie in a one-to-one correspondence (or even a one-to-many correspondence): my desire to gain a good reputation, for example, might require any number of acts, depending on the company I keep, the options

available to me, and so on.³⁴ What this means, however, is that nothing about a desire for an object connects it inherently to a specific action (a means towards realising that object), and this makes it difficult to see why we should speak of a *desire's* conformity to the moral law at all, rather than the conformity of certain *actions* to the demands of reason.

Is there any sense in which desires *themselves* might be deemed morally good or bad, permissible or impermissible, according to the moral law? Does the moral law speak to the so-called 'content' of our desires at all, and is it in this sense that they might obey or disobey reason? Two possibilities on this front may be ruled out from the start. First, on any recognisably Kantian framework, desires are not something which could, in and of themselves, be deemed morally permissible or impermissible from the standpoint of the moral law. In a very trivial sense, of course, we are permitted to *have* any kind of desire we wish, but this is only because in this sphere the moral law remains silent. Because it imposes an 'ought,' the moral law can pronounce its verdicts only on that for which we might reasonably be held responsible; since the contingent fact that I possess a certain desire is not a consequence of genuine choice, but is rather a bare fact stemming from my sensuous nature, it is not a proper locus of moral assessment. Moreover, that we are the kinds of beings that have desires at all is a contingent fact about our makeup, and not something which Kant takes to apply to any rational being as such. Similarly, the range of the possible objects of the relevant desire cannot of themselves be deemed either permissible or forbidden from the

³⁴ This is not to say that, in a particular circumstance, a specific action might not be the only means available for the satisfaction of a desire. It is rather to say that the judgement that certain desires are morally permissible or impermissible simply because they are connected to certain acts cannot be held *a priori*.

standpoint of the moral law. For Kant, objects of desire are states of affairs to which the agent attaches a certain agreeableness: they are, as we might say, ways that the world might be. But again, because my acts of choice might well bring about any number of actual consequences, depending on factors well beyond my control, the way the world happens to be is not something which is wholly determined by the exercise of my freedom, and thus something for which I can be held morally responsible. States of affairs, like the suffering of desire, remain beyond the boundary of the morally assessable. On these two fronts, the moral law is silent.³⁵

The above observations constitute significant concerns. Given the conclusions of §4.1, however, we are in a position to identify an even more fundamental shortcoming of this interpretation. Precisely because the moral law pronounces only on the form of maxims, it directs agents to consider questions of universal consistency. As we have seen, this leads to the idea that the moral interest is fundamentally different from the non-moral: where the former constitutes a direct interest in the action to be performed, the latter can only ever constitute an *indirect* interest in the action as a means to securing a desired or desirable effect.³⁶ Herman's alternative upsets this distinction to the degree that it proposes that acting on the categorical imperative, from respect for the law, consists in action from properly formed, properly responsive, inclination. In addition to the worries presented above, it thus fails to accom-

³⁵ In the Doctrine of Virtue, Kant argues that it is an indirect duty to cultivate compassionate feelings (see, e.g. *MdS* 6:457, §35). The emphasis here, however, is on the importance of doing so as an *aid* to ameliorating bad situations. As Kant makes clear, where I cannot help a person, my sharing of his pain merely increases the ills in the world, and is a bad thing (see *MdS* 6:457, §34).

³⁶ This point is developed in Jens Timmermann, 'Acting from Duty: Inclination, Reason and Moral Worth,' *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*, ed. Jens Timmermann (Cambridge: Cambridge University Press, forthcoming).

moderate the very clear affirmation, repeated by Kant throughout his ethical writings, of the heterogeneity of moral and non-moral concern. Were the moral law at all concerned with our desires and their objects – were it able to ‘speak to them,’ and so to ‘shape’ their content – this insistence on the dualism of human volition would be an ultimately disposable element in Kant’s ethical thought. As we have seen, however, it stems directly from the formal nature of the moral law as discussed above, and so ultimately from Kant’s commitment to the autonomy of pure reason: to the fact that the moral law is not conditioned by any object. The endorsement of Herman’s alternative thus entails the abandonment of this fundamental Kantian principle.

It is of course true that Kant accepts the idea that our desires may be refined through reflection and maturity. Where he does so, however, he warns that action on refined desires must not ever be presented as anything other than action on inclination. Though the desire on which the agent acts might be more honourable and admired than others, action for the sake of the former expresses *the very same principle* as action for the sake of the latter, and so cannot be deemed more or less moral. As Kant puts this point,

Only those who would like to deny to pure reason the ability to determine the will without some feeling being presupposed could deviate so far from their own definition so as to explain as quite heterogeneous what they have themselves previously brought under one and the same principle. (*KpV* 5:24)

Where genuine heterogeneity applies to the moral and the non-moral interest, those who seek to anchor morality in a ‘higher’ faculty of desire suppose, wrongly, that there is genuine heterogeneity among the incentives of inclination: among our refined motives and ‘those of the coarsest senses.’ In this way, according to Kant, they are like ‘ignorant people who would like to dabble in metaphysics’: who ‘think of matter

so refined, so superrefined, that they make themselves giddy with it and then believe that in this way they have devised a *spiritual* and yet extended being' (*KpV* 5:24).

4.4 *Interest, Dialectic and Character*

What I hope to have shown thus far is that it is misguided to suppose that Kant's distinction between inclination and the motive of duty is grounded on a strange obsession with dichotomies or a deep-seated disdain for our sensuous nature. Rather, it rests on an acknowledgement that the very concept of a categorical imperative implies that the moral law must be formal in the dual sense outlined above, and thus that moral concern attaches solely to the form of our maxims rather than to any object of desire. For better or for worse, Kant has great confidence in the claim that the common understanding will recognise that actions undertaken for the sake of some effect lack moral worth.³⁷ That reason is not in the business of approving (or respecting) the effects of my actions is meant to be a conclusion arrived at from *within* common human cognition – or so he supposes.³⁸ His conclusions about the heterogeneity of duty and desire, however, do not in the least hinge on their intuitive plausibility. As we see in Section II of the *Groundwork*, Kant's derivation of the categorical imperative proceeds directly from his understanding of the moral law as a law of reason in conjunction with his conception of human moral psychology, and it is this understanding of the content of the moral command that ensures that the moral interest

³⁷ The examples at *G* 4:397-99 are meant to make explicit this intuition.

³⁸ See *G* 4:400.19-25.

must be absolutely and purely formal. An acceptance of the distinctive starting point of Kantian ethical theory thus entails this most controversial of tenets.

With all of this said, however, I do not wish to underplay the significance of the concern to explain how it is that human beings, part sensuous and part rational, are able to take an interest in the moral law and in compliance with its dictates. Kant himself was acutely aware of the mystery associated with explaining the possibility of a moral outlook. As he writes in the *Groundwork*, where ‘pathological’ interest is easily explained by citing sensuous provocation of some kind or other, ‘it is quite impossible to explain how and why the *universality of a maxim as law* and hence morality interests us’ (4:460). Unable to cognise how the thought of universality on its own should create in us an interest to do what the moral law commands, we can at best show that we *must* suppose that it does, if there is to be such a thing as moral necessity.³⁹ As philosophers we would of course *like* to investigate the grounds for believing in the moral interest, but in the end Kant has faith that its possibility (itself grounded in freedom) is an assumption that human reason must ultimately accept. With this query, we reveal the limits of moral philosophy.⁴⁰

³⁹ See also *G* 4:458-9: ‘But reason would overstep all its bounds if it took it upon itself to *explain how* pure reason can be practical...’; ‘The subjective impossibility of *explaining* the freedom of the will is the same as the impossibility of discovering and making comprehensible an *interest* which the human being can take in moral laws, and yet he does really take an interest in them, the foundation of which in us we call moral feeling ...’ (*G* 4:459-60). At 4:461-62, Kant goes on to hint that the problem of explaining moral interest is related to the formal character of the categorical imperative.

⁴⁰ See *G* 4:462.22. See also *SF* 7:58-9: ‘For there is something in us that we cannot cease to wonder at when we have once seen it ... We do not wonder at the fact that we are beings subject to moral laws and destined by our own reason to obey them, even if this means sacrificing whatever pleasures may conflict with them ... But we do wonder at our *ability* so to

A related and equally interesting problem concerns not the possibility of explaining the moral interest as such, but the possibility of developing an unwavering and settled commitment to morality's demands. Kant's moral law is described as an unforgiving, commanding and compelling force; its origin is radically different from any source known to us via empirical introspection, and to that degree it is likely to be confronted by ordinary agents as a foreign yoke – as a constraint imposed *on* the self by some 'other,' majestic force rather than a principle which finds its source *in* the self.⁴¹ In comparison, moreover, Kant continually observes our great attachment to the 'dear self': to the sensuous side of our nature, our identification with which renders moral obedience, at least at times, a species of 'self-denial' (*G* 4:407.23-28). As he continually acknowledges, though there is something 'splendid' about the innocence of the common human understanding, it is rather easily 'seduced.' Human beings feel within themselves 'a powerful counterweight to all the commands of duty,' such that there arises a propensity on the part of human reason to give preference to inclination: 'to rationalise [*vernünfteln*] against those strict laws of duty and to cast doubt upon their validity, or at least upon their strictness' (*G* 4:405). Thus there is a resistance in human reason to accepting its own fundamental principle; though the implicit use and recognition of the moral law is natural to us, just as inherent is the tendency both to be concerned with our own happiness and to attempt to justify this propensity by ad-

sacrifice our sensuous nature to morality that we *can* do what we quite readily and clearly conceive we *ought* to do. This ascendancy of the *supersensible* human being in us over the *sensible*, such that (when it comes to a conflict between them) the sensible is *nothing*, though in its own eyes it is *everything*, is an object of the greatest *wonder*; and our wonder at this moral predisposition in us, inseparable from our humanity, only increases the longer we contemplate this true (not fabricated) ideal.'

⁴¹ This charge is discussed in some detail in §6.1.

opting a heteronomous moral philosophy.⁴² The danger here is a lapse into what Kant calls a ‘*natural dialectic*’ of reason, where human reason surreptitiously allows itself to substitute the principle of happiness for the moral law.

Given that the experience of our relative attachment to the moral law and to inclination so easily fits this pattern on Kant’s view, how are we to explain anyone’s ever developing a lasting and complete commitment to morality? Kantian autonomy contains an inner tension when applied to finite rational beings to the degree that it combines the solemn majesty of the moral law – its unanswerability, its purely formal disciplinary force – with the power of every agent to be motivated to do as duty prescribes.⁴³ Moral concern involves refusing to engage in a search for reasons for compliance beyond those contained in the law itself. Yet because we are sensuous beings, vulnerable to distraction and to moral quibbling, we are easily led to resist the moral law and to become suspicious of its adequacy as an incentive. Questions concerning the possibility of the moral law’s gaining lasting access to the human heart as a sole incentive are thus highly important. This is a topic to which we shall return in chapter six.

⁴² Wisdom requires science, according to Kant, ‘not in order to learn from it but in order to provide access and durability for its precepts’ (*G* 4:405). For more on this interesting strand in Kant’s thinking, see Guyer, *Kant on Freedom, Law, and Happiness* 211-212 and Susan Meld Shell, *Kant and the Limits of Autonomy* (Cambridge, MA: Harvard University Press, 2009) 130-31. On the natural acceptance of the moral law, see *KpV* 5:8n: ‘A reviewer who wanted to say something censuring this work [the *Groundwork*] hit the mark better than he himself may have intended when he said that no new principle of morality is set forth in it but only a *new formula*. But who would even want to introduce a new principle of all morality and, as it were, first invent it? Just as if, before him, the world had been ignorant of what duty is or in thoroughgoing error about it.’

⁴³ Kant refers to the ‘*solemn majesty*’ of the moral law at *KpV* 5:77. The notion that we stand under a ‘*discipline* of reason’ is outlined at 5:82.

Before turning to these worries, however, it is useful to consider the extent to which this aspect of Kant's ethical thought really might expose him to the charge that he has no plausible account of moral *character*. Herman casts her alternative as an attempt to 'make room for character' in Kant's ethics, and here her focus is on whether there might be space within Kantian moral philosophy for a recognisable notion of moral *development* or *formation*. Whilst this issue takes us, in part, into the territory of chapter six, chapter five will focus on a cluster of prior questions concerning Kant's thoughts on the nature and origins of moral character. Does the orthodox Kantian picture really lack a coherent account of moral character? More importantly for our purposes, in what sense might such an account be influenced by his commitment to the autonomy of practical reason?

ON THE PURITY OF THE MORAL DISPOSITION

A member of the English Parliament exclaimed in the heart of debate: ‘Every man has his price, for which he sells himself.’¹ If this is true (and everyone can decide by himself), if nowhere is a virtue which no level of temptation can overthrow, if whether the good or evil spirit wins us over only depends on which bids the most and affords the promptest pay-off, then, what the Apostle says might indeed hold true of human beings universally, ‘There is no distinction here, they are all under sin – there is none righteous (in the spirit of the law), no, not one.’² (*Rel* 6:38.34-39.6)

As we have seen, autonomy entails the loss of a certain kind of innocence. The autonomous will can undertake the pursuit of happiness only subsequent to pure reason’s permission or condemnation, such that immoral action involves not merely deviation from the moral law, but also a refusal properly to abide by its authority. On Kant’s view, this fact leads to two related ideas: the notion that there is no intermedi-

¹ This exclamation is commonly attributed to Sir Robert Walpole (1676-1745).

² As the editors of the Cambridge edition of Kant’s writings on religion point out, this phrase can be found in Romans 3:9-10: ‘What then? are we better *than they*? No, in no wise: for we have before proved both Jews and Gentiles, that they are all under sin; As it is written, There is none righteous, no, not one.’

ate moral position between good and evil at the level of maxim selection, and the claim that there are no moral characters that are neither good nor bad.

Whilst Kant accepted these implications, they are likely at first glance to seem overly restrictive. Kant's sharp distinction leaves no room for the idea that an action can be morally good in some ways and morally bad in others – a notion at the heart of much commonsense moral thinking. In practice, we would seem to find evidence for the idea that those around us take their reasons for acting morally well from a host of complex sources – some having to do with reverence for the law, others from more earthly concerns – and we do so without supposing that human action must be reducible to only two distinct types. Even less, however, do we suppose that people's underlying moral *dispositions* fall into two distinct categories. Rather, we tend to suppose that there are more than two fundamental moral types of people; that on-the-whole morally decent individuals may nevertheless differ in terms of the degree to which they are good; and that the moral improvement of character occurs most commonly as a gradual process of personal improvement.

In this chapter, I consider the deep connections in Kant's thinking between autonomy of the will, as we have been considering this property, and the possession of a moral disposition. For many readers of Kant's moral philosophy, his account of the latter notion is likely to seem impoverished. With this acknowledged, my aim is to defend Kant's rigoristic account of moral disposition by bringing clarity to the question of why Kant felt the need to describe our moral possibilities so starkly.

This chapter is divided into four sections. In section one, I consider Kant's conception of a supreme maxim, which he understands as a fundamental ground for the adoption of more particular maxims. Continuing to draw from *Religion Within the Boundaries of Mere Reason*, I then go on to discuss the implications of this notion for Kant's conception of intelligible moral character, focusing in particular on the rigoristic conclusions to which Kant is attracted. In section three, I evaluate, and ultimately reject, an alternative to Kantian moral rigorism proposed by Stephen Engstrom. As I argue, reflection on the inadequacy of Engstrom's proposal reveals the very interesting relationship in Kant's thinking between the moral worth of actions and good character.

5.1 *The Concept of a Supreme Maxim*

As we learn in the *Religion*, Kant takes human nature to possess an original predisposition (*Anlage*) to good, which is innate and so not chosen, and which contains three elements, themselves predispositions.³ The predisposition to *animality* is associated with a kind of self-love for which reason is not required and with the natural drives for self-preservation, the propagation of the species, and for community with other human beings (*Rel* 6:26.12-18). The predisposition to *humanity*, by contrast, is subsumed under the general title of self-love, which is not merely instinctual but also requires reason. It grounds an inclination to be seen as worthy (in some cases to be

³ To say that these are predispositions to good is to say that they enjoin observance of the moral law (*Rel* 6:28.12-14); no human being is good simply in virtue of possessing them, for we are not accountable for having them, but according to Kant they nonetheless demand compliance with the moral law to the extent that the ends they pick out are either best or solely realised under the condition of having chosen within the bounds of morality.

seen as superior) in the eyes of others, and since it involves comparison, it is rooted in a use of reason which is practical yet subservient to other incentives (*Rel* 6:27.4-12, 28.9-10).⁴ Finally, the predisposition to *personality* is described by Kant as ‘the susceptibility to respect for the moral law *as of itself a sufficient incentive to the power of choice*’ (*Rel* 6:27.27.29). It is because we possess this predisposition that it is possible for us to incorporate the moral incentive into our maxims. As Kant writes in the second *Critique*, personality is the very origin of duty:

[It is] freedom and independence from the mechanism of the whole of nature, regarded nevertheless as also a capacity of a being subject to special laws – namely pure practical laws given by his own reason, so that a person as belonging to the sensible world is subject to his own personality insofar as he also belongs to the intelligible world ... (*KpV* 5:87)

Our predispositions demand compliance with the moral law. How, then, are we to explain the possibility of morally evil maxims? What could possibly serve as the ground of the adoption of an evil maxim – as the explanation of why it, and not a good maxim, has been selected? One way of thinking about this is to seek the origin of evil in some notion of inherited sin (*Rel* 6:40.8-12). That this cannot be the solution on a Kantian approach, however, should be evident, for the notion that human beings could have inherited this way of selecting maxims from our ‘first parents’ precludes the possibility of imputing evil to individual human beings. In fact, it is Kant’s view that the origin of moral evil cannot be traced to *any* state which precedes the free exercise of the agent’s power of choice: whatever our ancestors may have chosen, whatever the natural causes influencing us, our actions must be free from determination by external causes if we are to be morally responsible. The morally evil ef-

⁴ It is a familiar Kantian theme that, other things being equal – and within the bounds of morality – the fulfilment of inclination is part of what is good for us and part of what nature intends for us. See, e.g. *Rel* 6:58.1-6.

fect under examination (the adoption of evil maxims) must be bound to its cause by the laws of freedom, in which case the ground of the morally evil exercise of choice must be sought, not in time, but in the representations of reason (*Rel* 6:39.15-41.26). We must take the human being to have fallen into evil from a state of innocence as the result of an original exercise of the power of choice, for only then can he be held responsible for the moral character of his actions.⁵

Similarly, the ground of evil cannot be placed in the mere possession of natural inclinations inviting the adoption of non-moral ends. This is because it is in terms of this ground that we impute moral goodness or badness to a person, and nothing can be imputed if it is not the result of the use of his freedom. If the fundamental ground for our adoption of a particular maxim were a natural impulse, the exercise of our faculty of choice could be traced back to a determination through natural causes, and this too would contradict freedom and imputability (*Rel* 6:215-18). On Kant's view, natural inclinations in and of themselves are no threat to the conformity of the will with the moral law, for the concept of moral evil applies to how we respond to our inherited drives.⁶ The ground of evil maxims must therefore consist in a ground of choice, and Kant explains this in terms of the enduring moral *disposition* (*Gesinnung*) of the hu-

⁵ In addition, Kant emphasises that the adoption of evil maxims cannot be blamed on a corruption of morally legislative reason. If this were to occur, the human being would be a diabolical (*teuflisch*) being – an *evil reason* or an *absolutely evil will*, which for Kant is impossible for us (*Rel* 6:35.20-26). It is hard to see how an agent whose morally legislative reason was corrupted in this way could genuinely count as a moral agent.

⁶ As Allan Wood puts this idea, inclinations are necessary for choice, and so for evil, but they are not its source. See *Kant's Moral Religion* (Ithaca: Cornell University Press, 1970) 112.

man being – a disposition which, since it must be attributable in the strong sense required by moral responsibility, is itself a maxim (*Rel* 6:20.33).⁷

As we saw in chapter three, because we are beings whose will contains a pure, legislative power, duty is an incentive for us. Whenever inclination bids us transgress morality, we have available to us a sufficient motive for refraining from doing so, which finds its source not in nature but in the will's own representation of universal law. As Kant puts this idea, the moral law imposes itself 'irresistibly' on even the one who transgresses it because of her moral predisposition, such that if no other incentive were present and working against the moral motive, she would be devoted to it as a wholly sufficient determination of her power of choice (*Rel* 6:36). Yet as a finite rational being, this agent is also dependent on the incentives of her sensuous nature, and so is necessarily devoted to their satisfaction (that is, to her happiness). We can thus think of the human being as possessing two fundamental practical commitments: on the one hand she is necessarily concerned with her own happiness; on the other, with acting on the demands of morality. To the degree that these two commitments are ineluctable, constituting two necessary implications of the human being's mixed nature, we can think of them as combining to make up something along the lines of a basic practical outlook – an inescapable standpoint against which decisions about how it

⁷ The idea of a moral disposition is a concept which Kant takes to be relatively intuitive to his reader: as he notes in a reply to one of his critics, the distinction between *virtus phaenomenon* and *virtus noumenon* is both contained and understood (though in other words) in the most common teachings and sermons, such that even young children are able to distinguish a person's facility in acting in conformity with duty from the possession of an underlying and constant disposition toward such actions from duty (*Rel* 6:14). This claim appears as a part of Kant's reply to a critic who complained that the *Religion* would be of no use or concern at all to those who are not acquainted with his entire system (the editors of the Cambridge translation to this work helpfully point out that this criticism appeared in the *Neueste Kritische Nachrichten* 29 [1793]: 225-29).

would be good to live are made.⁸ This is not to say that every particular choice *actually* satisfies each concern, but it is to say that particular choice necessarily reflects a decision about *what to do* with regards to the satisfaction of each, where both are in play. Human beings cannot fail to take into account both types of good when reckoning what to do (where both types are relevant), but what they actually choose to do about the pressure exercised by each is not itself determined by this fact about our nature.

Kant's view on the proper relationship between these competing incentives is clear. Objectively speaking, morality has lexical priority over happiness. The goodness of happiness is conditional on the moral goodness of the relevant agent, since it becomes bad or evil if its pursuit conflicts with the requirements of morality. Different conditionally good things can be measured against each other; I might, for instance, weigh up the goodness of satisfying my inclination for good health against my inclination for thrill seeking and find that the latter wins out over the former. Similarly, the realisation of happiness admits of degrees: a certain course of behaviour can be better or worse for me from the standpoint of prudence. Moral goodness, however, does not admit of degrees and actually blocks the value of conflicting options. It is incomparable, in this sense, and functions as a strict limit on the goodness of happiness to the extent that prudent action possesses genuine value only on the condition of my moral worth as an agent.⁹ Because of this, the moral law finds its rightful place far above

⁸ See *KpV* 5:25: 'To be happy is necessarily the demand of every rational but finite being and therefore an unavoidable determining ground of its faculty of desire.' See also *G* 4:414n, where Kant defines inclination as '[t]he dependence of the faculty of desire upon feelings.'

⁹ See *G* 4:393, *TP* 8:283.

happiness on the table of kinds of value, and a well-ordered motivational disposition will reflect its lexical priority.

When we say that a human being is evil, according to Kant, we mean that ‘he is conscious of the moral law and yet has incorporated into his maxim the (occasional) deviation from it’ (*Rel* 6:32.13-16).¹⁰ That is to say, we mean that he reverses the proper (or objective) order of the incentives furnished by morality and self-love respectively, making the promotion of self-love the condition of morally acceptable activity.¹¹ Every morally relevant choice may be represented as reflecting an underlying sense of priority, and so as involving the subordination of one to the other ‘as its supreme condition’ (*Rel* 6:36.28). Yet when it comes to the evil human being,

he makes the incentives of self-love and their inclinations the condition of compliance with the moral law – whereas it is this latter that, as the supreme condition of the satisfaction of the former, should have been incorporated into the universal maxim of the power of choice as the sole incentive.’ (*Rel* 6:36.26-33)

¹⁰ For helpful discussions of Kant’s conception of evil character, see Allison, *Kant’s Theory of Freedom* ch. 8; Allen Wood, *Kant’s Moral Religion* ch. 6; Patrick Frierson, *Freedom and Anthropology in Kant’s Moral Philosophy* ch. 5; G. Felicitas Munzel, *Kant’s Conception of Moral Character* (Chicago: The University of Chicago Press, 1999) esp. ch. 3; and Gordon E. Michalson, Jr., *Fallen Freedom* (Cambridge: Cambridge University Press, 1990) esp. chs. 2 and 3. For a defence of Kant’s claim that the propensity to evil is possessed universally by human beings, see Seiriol Morgan, ‘The Missing Formal Proof of Humanity’s Radical Evil in Kant’s “Religion”,’ *The Philosophical Review* 114 (2005): 63-114.

¹¹ In his discussion of *Gesinnung*, Henry Allison suggests that we understand this notion in terms of ‘an underlying set of intentions, beliefs, interests, and so on’ (*Kant’s Theory of Freedom* 136), which together constitute a ‘pattern of willing’ (137). Rather than appealing to this cluster of terms, however, I suspect that it is more helpful to focus on the notion of a choice in which the two kinds of incentive which influence the mixed will are ordered and prioritised. A *Gesinnung* might well be a pattern of willing and an underlying set of interests in some sense, but it is primarily a prioritization of two competing incentives, with which the human being finds himself forced to contend.

A good maxim will be one in which the incentives of self-love are subordinated to the incentive of respect for the law – it will be one in which activity in conformity with the pursuit of happiness is undertaken only on the condition that it is morally permissible. An evil maxim, by contrast, will reverse the relation of subordination, making the satisfaction of the moral law conditional on its relation to the promotion or preservation of happiness. A person's fundamental maxim will thus express her most fundamental practical orientation: it will express her commitment either to the priority of morality or to the supremacy of self-love.¹² For Kant, it is in this sense that a person's fundamental maxim may be described as reflecting her character.

5.2 *Empirical and Intelligible Character*

As I suggested above, experience would seem to confirm that human beings are by nature good in some ways or parts, but evil in others. In any case, our common practices of moral discrimination, of praise and blame, generally presuppose that there are degrees of goodness and evil when it comes to the overall assessment of a person. Yet Kant makes clear that he wishes to exclude anything intermediate when it comes to the moral assessment of actions and character, and so to preclude the possibility of

¹² In this way, as Henry Allison observes, the Kantian choice of *Gesinnung* differs from a Sartrean *projet fondamental*. To the extent that the moral disposition consists in a ranking of the incentives provided by pure reason and inclination, the idea of a good or evil disposition presupposes the dictates of *Wille*. See *Kant's Theory of Freedom* 142-3. At this point, we might wonder whether there is room in Kant's ethics for one-off morally good actions, even when they stem from a character that is morally bad. It seems natural to think, after all, that even the most heard-hearted scoundrel might have a moment of moral clarity, seeing what he morally must do and doing it because the moral law tells him to. This possibility will be addressed in §5.3.

moral characters which are morally good in some ways and evil in others, as well as those which are neither morally good nor evil (*Rel* 6:22.17-18). On his view, which he calls moral rigorism, there are only two types of moral character, the good and the evil. Although we may find comfort in the intermediate view, which allows us to assure ourselves that there is some goodness in ourselves and those around us though we are evil in other ways, it ought to be rejected.

The denial of the morally indifferent follows straightforwardly from an acceptance of the notion that the moral law is always an incentive for the free power of choice where duty is at stake. As Kant explains in a lengthy footnote in the *Religion*, if the moral law were not an incentive of the power of choice, the agreement of the latter with the moral law would be good and any disagreement a morally indifferent *lack* of goodness – an action which is not good, but not because of that evil. But in the human being the law *is* an incentive, and so any lack of agreement with the moral law is possible only as an active resistance on the part of *Willkür* to that law, which constitutes, not a mere lack of good, but rather an *antagonism* to the good. As such, ‘between an evil and a good disposition (the inner principle of maxims) according to which the morality of an action must be judged, there is no intermediate position’ (*Rel* 6:22n). Where the moral law fails to determine a person’s free power of choice, an incentive from inclination must have influence. Yet because morally good choice is a possibility for us – because the moral law never fails to announce itself to the faculty of choice as supremely authoritative where morality is at stake and because it never fails to offer an incentive to compliance with it – the incorporation of a sensuous incentive *just is* the incorporation of the deviation from the moral law into one’s maxim. It follows from this that a person’s disposition as regards the moral law can never be indifferent (neither good nor evil) (*Rel* 6:24.7-15).

That human dispositions and actions cannot be conceived as good in some ways and evil in others is ensured, on Kant's view, by the fact that a person's underlying disposition must express one of only two relations of priority at the level of incentives. This relation of subordination is absolute and without qualification; the one value (morality or happiness) is made the condition of the pursuit of the other, and as such the attitude constituted by this ordering represents a commitment to the priority of *either* morality or self-love as the spirit in which one makes one's choices. Precisely because this spirit is defined by an exclusive choice *between* two fundamental values or standpoints, each reflecting a basic element of our mixed nature, it grounds actions which are either good or evil, but never both.¹³

This designation of only two fundamental orderings entails the surprising conclusion that it is actually morally evil (though perhaps in conformity with the letter of the law) to incorporate an incentive from self-love as sufficient motivation for action in accordance with duty. As Kant makes clear, his particular conception of a person's being evil (*böse*) is a designation from the standpoint of morality rather than appearances, and although the word evil may at first seem overly strong when applied to

¹³ As Kant points out in an illuminating footnote, it is quite fitting for the Christian tradition to represent the moral good (and its principle) as differing from the moral evil (and its principle of self-love), 'not as heaven from *earth*, but as heaven from *hell*. This is indeed a figurative representation and, as such, a stirring one, yet not any the less philosophically correct in meaning – For it serves to prevent us from thinking of good and evil, the realm of light and the realm of darkness, as bordering on each other and losing themselves into one another by gradual steps (of greater and lesser brightness); but rather to represent them as separated by an immeasurable gap. The total dissimilarity of the basic principles by which one can be subject to either one or the other of these two realms, and also the danger associated with the illusion of a close relationship between the characteristics that qualify somebody for one or the other, justify this form of representation which, though containing an element of horror, is nonetheless sublime' (*Rel* 6:60 n).

cases of dutiful action from inclination, it designates a kind of moral failure which, according to Kant, even small children are able to recognise as sin: an orientation of spirit which *at base* places morality below self-love. Kant explains his position as follows:

[W]henver incentives other than the law itself (e.g. ambition, self-love in general, yes, even a kindly instinct such as sympathy) are necessary to determine the power of choice to lawful actions, it is purely accidental that these actions agree with the law, for the incentives might equally well incite its violation. The maxim, by the goodness of which all the moral worth of the person must be assessed, is therefore still contrary to law, and the human being, despite all his good actions, is nevertheless evil. (*Rel* 6:30.34-31.5)¹⁴

Thus while an externally well-behaved human being and a morally good human being will betray no difference at the level of the agreement of their actions with the moral law, a fundamental difference lies in the fact that the actions of the former do not always have (or perhaps ever have) the law as their sole and supreme incentive, ‘whereas those of the latter *always* do’ (*Rel* 6:30.28-29).

In a footnote commenting on certain ancient philosophers, Kant suggests that they were right to believe both that virtue cannot be learned and that it is unitary.¹⁵ If virtue could be taught and learned, this would imply that the human being in its nature is indifferent to virtue and vice, which as we have seen must be denied if we are to make room for the possibility of autonomy. Similarly, to deny the unity of the virtues would allow for the thought that human beings can be good in some parts but bad in

¹⁴ By appealing to the idea of ‘good actions,’ here, Kant must be referring to actions that appear to be good, or that conform to the moral law in their outward nature.

¹⁵ *Rel* 6:24n. It is not made clear who, exactly, Kant is referring to here when he speaks of the ‘ancient moral philosophers.’

others. These philosophers were right to hold these positions because they were considering virtue *in itself*, in the *idea* of reason. Yet Kant is also careful to note that if we wish to pass moral judgement on the human being as she appears in nature, and so as experience allows us to understand something about her, we are licensed to deny these ancient tenets. The person as appearance is to be judged according to empirical standards, by a human judge, and ‘not by the scales of pure reason (before a divine court of justice)’ (*Rel* 6:24n), and what this distinction allows for is a potential softening of the ascription of evil at the level of *Gesinnung*. Where an agent may be called evil when judged by the scales of pure reason, human judges must pronounce their own verdict, though it is made according to different (empirical) standards. The principle of the exclusion of a mean between good and evil must be the basis of one level of moral evaluation (in the end, that which genuinely deserves to be called *moral*), whereas it need not be seen to apply at the level of empirical judgement, which involves the discrimination of empirical deeds (actual doings or not doings). At the level of appearances, we find both ‘a negative mean of indifference prior to all education’ and a positive mean – ‘a mixture of being partly good and partly evil.’ This second judgement, however, ‘concerns only human morality as appearance, and in a final judgement must be subordinated to the first’ (*Rel* 6:39n).¹⁶

¹⁶ ‘Morality as appearance’ is something that we do not find in Kant’s earlier works. His use of the term in the *Religion* would suggest that he is moving toward the allowance of something like a phenomenal morality (perhaps a small-m ‘morality’ as opposed to the capital-m ‘Morality’ with which he is concerned in the *Groundwork* and the second *Critique*). His use of this idea in his later works might, on the other hand, simply be a way of accommodating how people speak (‘his actions might be good, but ooh, what a bad man he is!’) and of appealing to an idea which is engrained in the common understanding, even though there is, strictly speaking, no room for *morality* as appearance in his theory.

Upbringing and habituation can save empirical character, but they cannot by themselves save intelligible character. Here, Kant quite importantly points out the element of moral luck involved in the acquisition of empirical character – an element which is expressly and just as importantly denied when it comes to the free will, which selects for itself an intelligible disposition. People who act in accordance with the moral law but who do not consult that law or allow it to count the most in the selection of their maxims only *luckily* slip by the evil consequences which might well have accompanied their choices had circumstances been different. Such people often enjoy a measure of peace of mind and fancy themselves blameless in comparison to those who are not so lucky, although they do so, according to Kant, without asking whether the credit is perhaps better given to good luck or whether their underlying attitude would have just as easily led them to similar (empirical) vices, ‘had they not been kept away from them by impotence, temperament, upbringing, and tempting circumstances of time and place (things which, one and all, cannot be imputed to us)’ (*Rel* 6:38.12-23). Intelligible character will express itself in the world in different ways, depending on factors such as these, and although on the whole we call a human being morally good because of how his character *appears* to us as an expression in this world, we ought not to lose sight of the fact that the appearance of character can only provide a means for the all-too-fallible inference of a certain kind of intelligible disposition. Empirical virtue requires the cooperation of stepmotherly nature, yet genuine moral virtue must be freely chosen by the will on its own. To ascribe empirical virtue is thus to point out a particular kind of good fortune, while to ascribe good intelligible character is to suppose that the agent’s apparently good actions express a purity of disposition which wholly transcends fortune’s grasp. We often forget the significance of luck at the earthly level, especially when it comes to our own apparently good actions. Yet one

of our first and most significant lessons as human beings lies in the teaching that genuine moral discrimination concerns not the world, but the heart.¹⁷

It is important to note that one consequence of Kant's view on this matter is that a person's fundamental moral identity – the inner disposition which grounds her adoption of maxims and which distinguishes her as possessing a certain sort of intelligible character – is not to be confused with how she sees herself when reflecting on the empirical expression of her choices. Nor are we to think of good intelligible character as inferable from the presence of peace of mind or the absence of a guilty conscience. Kant does in fact think that immoral action results in condemnation by the inner judge, from which we might run but can never fully escape, but he also acknowledges the complex ways, almost always involving self-deception, in which we can silence this voice at the level of consciousness.¹⁸ How we fancy ourselves as moral agents

¹⁷ As Henry Allison notes, this talk of empirical and intelligible character – along with the requirement that *Gesinnung* must be freely chosen in an intelligible act – does seem to suggest a commitment to a conception of 'a merely intelligible subject that creates its own character, together with its phenomenal manifestations, through an unconditional timeless activity' (*Kant's Theory of Freedom* 138). In the end, however, Allison endorses a reading of Kant whereby his references to 'intelligible character' and 'causality as noumenon' are part of a conceptual claim about how human beings must take themselves if they are to hold themselves morally responsible rather than a metaphysical explanation. It is a claim, that is, 'which is grounded in a reflection of the conditions of the possibility of imputation' (144). I am not sure that this move is the right one to make, but since this is not the place for a full-fledged discussion of Kant's metaphysical ambitions, I must put this question to the side.

¹⁸ At *G* 4:404, Kant describes the common understanding as engaging in 'quibbling tricks' with conscience. In the second *Critique*, Kant emphasises conscience's ultimate upper hand in the matter: 'A human being may use what art he will to paint some unlawful conduct he remembers as an unintentional fault ... and to declare himself innocent of it; he nevertheless finds that the advocate who speaks in his favour can by no means reduce to silence the prosecutor within him, if only he is aware that at the time he did this wrong he was in his senses, that is, had the use of his freedom; and while he *explains* his misconduct by certain bad habits

tells us nothing about who we really are and about the orientation of our fundamental maxim. That image of moral goodness with which we might *identify* on self-examination can differ quite dramatically from the moral identity chosen by our free power of choice, and while the ascription of empirical virtue makes up a large part of the practice of morality in everyday human life, we ought not to fool ourselves when it comes to its relation to what really matters from the standpoint of God and pure reason.

5.3 *Conditioned Goodness*

Up to this point, I have been explaining how Kant's rigorism flows naturally from the idea that human beings have at their disposal two types of incentive to action and that moral character depends on adopting a fundamental orientation toward their relation. We might, however, question the latter of these background assumptions. In a paper on Kantian autonomy, for instance, Stephen Engstrom wonders whether Kant's rigoristic division of character into good and evil might spring from a far too idealised conception of the finite rational agent. As he writes,

[i]f agents are viewed as having rendered all of their specific maxims consistent with a single, unifying highest maxim in which some determinate, ordered relationship between the two abstract principles is worked out, then rigorism may seem quite plausible. But when our question is about the characters of actual human beings, such idealisation seems out of place.¹⁹

... yet this cannot protect him from the reproach and censure he casts upon himself" (5:98). For related discussions of conscience, see also *MdS* 6:438-440 and *VC* 27:351.

¹⁹ 'Conditioned Autonomy.' *Philosophy and Phenomenological Research* 48 (1988): 443.

According to Engstrom's criticism, it is rarely the case that human beings are altogether committed either to morality or to self-love. His central example is that of a man who refrains from telling a lie on a certain number of occasions, though he would not have done so had the circumstances been relevantly different – if, in particular, the cost to himself had been much higher. The point of his illustration is to show that even though in situations of moderate difficulty this man acts honestly – and does so because he thinks that deception runs against the moral law – his motive in so acting is nevertheless not purely moral in Kant's sense. When he views the possibility of practicing deception in the abstract, he is able to see that it is morally unacceptable. Yet where he finds himself in a situation where deception would allow for him to avert a personal tragedy, he deceives. Because of this, we can see that he has not made morality a *necessary* condition of his adopting his maxim: 'though in ordinary circumstances his actions are in accordance with duty, he nevertheless does not act for the sake of duty.'²⁰ The man is not wholly 'wedded' to self-love, but neither does he make morality a supreme condition of his pursuit of self-interest. Indeed, what we seem to find lacking in this case is any sense of a clear-cut and fundamental maxim ranking the one over the other.

In light of the intuitive plausibility of his example, Engstrom's proposed alternative is a picture in which evil and goodness do not exhaust our options when it comes to character. As he puts his view, to deny that an agent is completely good is not to assert that happiness is a necessary condition of his conduct, 'but only that morality is not.' Similarly, to deny that an agent is completely evil 'is not to assert that morality

²⁰ 'Conditioned Autonomy' 445.

is a necessary condition,’ but only ‘that happiness is not.’²¹ Yet to say this is to make room for an intermediate option when it comes to moral character, for what we might now say is that there are some agents who commit *neither* to morality *nor* to happiness as supreme conditions for choice. Unlike the completely good agent, those agents are not wholly wedded to morality; unlike the completely bad agent, they are not wholly wedded to happiness. Rather, they are moved both by duty and by happiness and so are attached imperfectly – though we need not suppose in equal measure – to each.²²

Engstrom’s alternative is intriguing, particularly to the extent that, if tenable, it would seem to provide us with a way of escaping the rigorism which threatens the acceptability of Kantian ethics in light of our common intuitions. Attention to how this alternative is spelled out, however, reveals quite interesting difficulties.

Let us look again at the case of the generally honest man. According to Engstrom’s explanation, this man is not purely good; were he pressed by severe financial need, he would choose to act dishonestly. In his ordinary dealings, he remains honest and dependable, and his honesty extends to situations of even moderate difficulty. That is to say, he does make some sacrifices from the standpoint of happiness because he chooses not to deceive, and so at times transgresses the recommendations of prudence. Yet when he rejects dishonesty in such cases, he does so on moral principle,

²¹ ‘Conditioned Autonomy’ 448.

²² ‘Conditioned Autonomy’ 448.

and so reveals that his choice is not conditioned by the principle of prudence. Were the risk to his well-being more grave, he would choose to deceive.²³

It is important to ask about the kind of principle that the generally-but-not-perfectly honest man adopts when he does choose to tell the truth against the bidding of prudence. On Engstrom's account, this is not to be understood as a principle of self-love, but nor is it to be understood as a pure moral maxim in Kant's strict sense of this notion. That is to say, it is presumably meant to carry a certain *degree* of goodness, but not perfect goodness. But what, when all is said and done, is *morally* good about a maxim of this sort?

On the orthodox Kantian view of things, we find a clear explanation of why a maxim which expresses a moral price or breaking point must be considered a through-and-through maxim of self-love, and so as grounded in an evil fundamental maxim. When it comes to such a maxim, it need not be the case that one's sole reason for acting concerns the straightforward promotion of one's happiness (in this case in particular, the idea is that one's immediate happiness *is* put at risk by telling the truth). Nevertheless, what is striking about this maxim is that it betrays an underlying attitude which allows concerns stemming from self-love to have the *ultimate* say over

²³ 'Conditioned Autonomy' 449. On Engstrom's alternative, 'autonomy' is to be ascribed to wills for whom morality 'has sufficient reality as a maxim,' such that 'personal happiness is limited or sacrificed where necessary for the sake of conformity with the requirements of duty.' Autonomy is 'conditioned,' however, if this limitation is incomplete: 'if temptations, fears, and the like are overcome, but only so long as the stakes are not too high.' This occurs when there is a limit to the amount of pressure that the power of choice can withstand before it gives way to self-love, and so ought to be ascribed to those who, while generally honest and good in their day-to-day dealings, have their price in the manner described in the epigraph to this chapter.

whether the agent will be honest, and this is why it must count for Kant as immoral. As it happens, the agent's concern for happiness is such that it allows a certain amount of happiness to be sacrificed before it demands that he deviate from the letter of the law. Yet the fact that there is a threshold of this sort lurking behind the apparently honest action means that a fundamental prioritisation of a very certain sort can indeed be attributed to the agent (though it may never be consciously considered) and that this fundamental standpoint— the standpoint from which the 'final say' is pronounced — is ultimately (although perhaps in a lenient way) concerned with consequences for happiness.²⁴

It does not help us to say that the maxim which is allowed by the agent's concern for his own happiness is a maxim of duty (although the underlying character behind it is not wholly good), for there are good reasons to think that this cannot be the case on Kant's view. When an agent acts from duty, she acts on the basis of respect for the law, which involves not simply the idea that so acting is required by morality, but also, and most importantly, an appreciation of the supreme and unconditional authority of morality in relation to happiness. As Kant writes in the *Groundwork*, acting from duty is equivalent to acting from reverence for the law. Objects which are the effects of our actions can be the object of inclination, but never reverence, insofar as they are 'merely an effect and not an activity of a will' (4:400). While a state of affairs might well come about as a contingent result of the activity of a will, reverence is reserved for that which properly and *directly* belongs to a will: its own activity. In

²⁴ In this way, a fundamental maxim is *implied* in the agent's choice — because she is autonomous and the moral law is an incentive for her — even if it is not expressly adopted. It would certainly count as an 'idealistic' view of the moral agent to suppose that the latter must be the case, but the former idea is much less likely to raise this concern.

addition, however, reverence is described by Kant as an attitude which is reserved for that which one takes to exclude inclination (4:400). As he writes, it signifies consciousness of a very special kind of worth: ‘a worth that infringes upon my self-love’ (4:402n) and that far outweighs any worth claimed by happiness.²⁵ Because reverence consists in the recognition and affirmation of the incomparable worth of the moral law in comparison to happiness, Kant describes action from duty as involving the absolute exclusion of inclination in deliberation. In action from duty, he writes, one ‘puts aside entirely the influence of inclination and with it every object of the will’ (4:400), such that it is the pure thought of duty alone which determines the will to action.

Our ability to see anything good in the agent who tells the truth when the price is not too high depends on our being able to see at least that instance of truthfulness as being done from duty. Yet as the above shows, this in turn depends on our being able to view his situation as one in which the influence of inclination on his decision is completely put aside. Can we do this? On the face of things, this seems doubtful. Though we need not construe this as a case where the agent acts from an immediate sense that truthfulness will bring him happiness, it is natural to think that inclination does have a very important influence on his choice, for it is central to this example that if the price of being truthful had been just high enough, the agent would have resorted to deception instead. To the degree that the almost-perfectly-good agent in our example has

²⁵ See also 4:403: ‘Although I do not yet *see* what this respect is based upon (this the philosopher may investigate), I at least understand this much: that it is an estimation of a worth that far outweighs any worth of what is recommended by inclination, and that the necessity of my action from *pure* respect for the practical law is what constitutes duty, to which every other motive must give way because it is the condition of a will good *in itself*, the worth of which surpasses all else.’ This theme is continued in the second *Critique* (see 5:72-78).

his price, he allows considerations that have to do with his happiness ('would this cause catastrophic damage to my well-being?') to function as an ultimate limiting condition on his truthfulness. Yet to the degree that he allows *this*, he allows inclination to influence his will in a way that would seem to be incompatible with acting from duty in the specific sense outlined above.²⁶

If a person puts any price on morality, though he might at some level appreciate good actions and the examples of virtue he encounters, he does not accord them the status that they ought to have, which is, as Kant puts it, beyond price. To put a price on morality is in effect to assign a price to oneself: a point at which one can be bought off. Take, for example, Kant's description of the gallows case in the second *Critique*.²⁷ The situation here is one where we are asked to imagine someone who asserts that he cannot control his 'lustful inclination'; would he find his inclination so irresistible were we to threaten him with a hanging should he satisfy it? We are meant to suppose the he would not. We are then asked to imagine whether, on pain of the same penalty, this person would not give false testimony against an innocent man. Here we are likely to be less certain about our answer, but what we can at least recognise is that the man who does not allow even loss of life to limit his commitment to doing the right thing possesses a kind of honour and self-mastery which, though achievable by each of us, is possibly very rare indeed. Were his price to be met when tempted with

²⁶ An important question for Engstrom would thus concern the nature of this agent's incentive; for if it is not explainable as a species of reverence for the law, and if it is not self-love (as Engstrom maintains), then what could it be?

²⁷ See the bottom half of *KpV* 5:30. See also the example of Anne Boleyn at *KpV* 5: 155-6, where the thought is that our esteem for the virtuous agent who refuses all gifts and who suffers all punishments is so high precisely because his virtue costs him so much, and not because of any benefit he receives.

the object of his inclination or with the sparing of his life, there would be a point at which *he* could be bought off, and so a point at which he could be manipulated by anyone in the position to pay the price he has put on himself.

In defence of Engstrom's alternative, it might be argued that what Kant really means to preclude in cases of action from duty is the *direct* influence of inclination on decision-making (as is the case when an agent tells the truth *because* doing so will directly promote her non-moral ends), rather than the indirect influence of happiness as an ultimate and distant limiting condition. If this is so, then we might be able to preserve a sense in which the truthful man who has his price is still acting from duty when he tells the truth. On closer inspection, however, there is good reason to think that this cannot be the case on a recognisably Kantian approach to moral goodness. As we saw above, Kant allows for a kind of moral luck at the level of empirical character – the *appearance* of virtue or vice. He allows that there are all kinds of situations where a person who does not act from duty luckily avoids the evil consequences that might have resulted from his underlying maxim had the circumstances been different. Thus an underlying maxim of self-preservation might be fulfilled by the pursuit of all kinds of apparently noble actions (e.g. honesty, helpfulness, industriousness); that such noble pursuits constitute the best means to self-preservation for a person, however, is determined completely by luck – by the fortuitous constitution of the universe – such that the connection between the motive of self-preservation and these pursuits is entirely contingent. On Kant's account, a person who finds herself in lucky circumstances is likely to possess a peace of mind about her character, whereas she who is not so lucky, and whose circumstances are such that self-preservation recommends all kinds of apparently vicious wiles, is likely to be tormented by the inner judge. In the former case, however, it is clear to Kant that all credit must be

given to fortune rather than anything having to do with the agent's will.²⁸ Because genuine moral goodness must result from an act of freedom in Kant's very strict sense, the lucky agent is far from good.

If we look closely, we see that Kant's insistence that morally good actions be done for the sake of duty hinges on this very concern about the contingent fit between certain motives and apparently praiseworthy kinds of action (actions that conform to duty). As he notes in the preface to the *Groundwork*,

... in the case of what is to be morally good it is not enough that it *conform* with the moral law, but it must also be done *for the sake of the law*; without this, that conformity is only very contingent and precarious, since a ground that is not moral will indeed now and then produce actions in conformity with the law, but it will also often produce actions contrary to the law. (4:390)

Where the motive is not duty itself, a person's maxim will be such that the conformity of his behaviour to duty is only accidental, ensured by the state of his circumstances and not by any necessary connection. It is in this sense that the honesty of Kant's prudent shopkeeper does not evidence a good will, for had honesty not been the best policy in his particular case, his maxim of prudence would have committed him to some other, and perhaps more dubious, course of behaviour.²⁹ Nor can Kant's naturally sympathetic man be held as an example of moral goodness; had the circumstances dictated that sympathy was to be best served by acts contrary to duty, his maxim would have committed him to taking these sinister means. The connection between the agent's benevolence – his conformity to duty – and his underlying mo-

²⁸ See again *Rel* 6:38.

²⁹ See *G* 4:397.

tive is entirely accidental.³⁰ By contrast, in action from duty, it is no accident that the agent's motive connects him to dutiful acts. Because the motive consists in reverence for the moral law itself and not some object to be achieved by compliance with it, it *necessarily* grounds actions which conform to duty. No cosmic cooperation is required in order for the match to be made, and thus the credit for compliance belongs entirely to the agent.

What we see, then, is that action from duty is such that conformity to duty is no lucky accident. In the case of Engstrom's truthful man, however, there is an important sense in which his conformity to law really *is* contingent in precisely the way that concerns Kant in his famous examples. Although Engstrom wishes to say that this character's particular maxims possess some degree of moral worth when the stakes are not too high, in the end this man is simply *lucky* whenever this occurs. Whenever he performs a dutiful action, it is a matter of fortune that the circumstances are not overly dire and that he need not sell himself for the price which is nonetheless set on his allegiance to duty. Just as fortune smiles on the prudent shopkeeper when honesty turns out to be the best policy, so too does it smile on the one who is not asked to sacrifice beyond his limit by complying with the moral law. Engstrom describes this man as acting on moral principle *in this very case*, and he thinks that because of this we can speak of his moral goodness as something which he possesses as a matter of degree. Yet to the extent that we agree with Kant and maintain that action from duty must

³⁰ See *G* 4:398.

avoid accidental conformity to law, it cannot be the case that this man acts on moral principle, in this situation or in any one like it.³¹

This is a criticism of Engstrom's proposal, but it is also illustrative of the deep connections in Kant's thinking between moral worth, moral egalitarianism, and the rather peculiar conception of character outlined in §§5.1 and 5.2. The only way to ensure that dutiful action follows in a non-contingent way from the nature of an agent's maxim, and so has moral worth, is to guarantee that she does not have a price. Where there is an upper limit to her commitment to morality, no matter how high, it is a contingent fact that her number has not been called here and now, and so a contingent fact that her action has conformed to morality. Yet to say that an agent has no moral 'price' is precisely to say that a fundamental attitude toward morality may be attributed to her as an implied practical commitment, such that she is unconditionally 'wed,' to use Engstrom's phrase, to the supremacy of the moral law. It is only if she possesses an underlying maxim of this sort, which subordinates the satisfaction of inclination to respect, that her action is non-contingently, and thus *morally*, good. Put differently, it is only if she possesses an underlying maxim of this sort that she may be described as having acted from duty in Kant's specific sense. The possibility of moral worth in Kant's sense thus implies the idea of a fundamental maxim, suggesting that Kant's conception of character is far from incidental to his moral philosophy.

³¹ In this sense, Kant must reject the possibility of what Bernard Williams and Thomas Nagel have referred to as 'circumstantial' moral luck: the moral luck that affects the kinds of situations and problems one faces. See 'Moral Luck,' *Proceedings of the Aristotelian Society, Supplementary Volumes* 50 (1976): 115-135, 137-51.

5.4 *Assessment and Luck*

Though Kant's rigorism is deeply connected to his commitment to autonomy and to his conception of moral worth, many of his readers are liable to find this an unwelcome result. Kant allows, as we have seen, that his view accommodates the *appearance* of differing degrees of goodness, such that we cannot in fairness charge him with contradicting the evidence given by experience. Yet what his view does imply is that, in a very real sense, all good and all evil is as it were on par, for all cases reduce to the adoption of one of only two fundamental maxims. What this means is that, in one important sense, the general who leads the massacre of innocent villagers is really no worse, *morally speaking*, than the one who tells a lie to get out of trouble. Both are evil for what turns out at base to be just the same reason: they have adopted into their fundamental maxim the occasional deviation from duty in favour of self-love, such that any *apparent* moral difference between them can only be attributed to luck. Where one finds himself in a situation where his fundamental maxim is expressed by telling a lie, the other's situation is such that his maxim finds its expression in a much more serious way (at least from the standpoint of wellbeing).

This is not to say, however, that we cannot draw important distinctions about these characters at the level of what Kant comes in his later writings to call *empirical* character. Where we must deny a mean between good and evil character in the intellectual judgement of human beings, we may nonetheless apply this principle to the empirical judgement of humankind, based wholly on observation of their deeds. Thus Kant allows that there is 'a negative mean of indifference' which applies to agents' actions 'prior to all education': to the degree that a person has not yet been exposed to the relevant empirical prompts and experiences, she is neither good nor bad. Equally,

there is a ‘positive’ mean, as when an agent’s character is judged to contain a mixture of good and evil (*Rel* 6:39.n). Yet in both cases the judgement concerns only ‘human morality as appearance’ – it must, as such, be subordinated to the intellectual judgement of human beings, and so to the juridical judgement of ‘one who knows the heart’ (*Rel* 6:72).³²

As I suggested above, ‘human morality as appearance’ is a notion that does not find a place in Kant’s earlier ethical writings. Though there is much talk in the *Groundwork* and the second *Critique* of actions which merely accord with the moral law, judgements of conformity are not properly *moral* in the strict sense which concerns Kant throughout these works. His continued mention of ‘empirical character’ and ‘morality as appearance’ in the *Religion* thus hints at the idea that he might have been willing to make room in his moral theory (though only slight, and ultimately subordinate room) for something like Engstrom’s suggestion of the possibility of intermediate moral character, and so for a level of moral evaluation beyond the strict scrutiny of hearts. With this said, however, Kant’s shift in language may in the end stem merely from an interest on his part in accommodating our everyday way of speaking about moral matters. Talk of empirical ‘morality,’ tied as closely as it is to mere outward conformity with the law, may be best read as representing a popular way of capturing what is at base a morally irrelevant phenomenon. Nevertheless, Kant’s repeated return to this notion in the *Religion* points at least to his recognition of the importance, in ordinary moral judgement, of this external axis of evaluation. Whilst his moral theory cannot allow for the genuinely *moral* significance of outward conformity, as

³² God is again described as fulfilling the role of ‘scrutiniser of hearts’ (via pure intellectual intuition) at *Rel* 6:67. Our own inability to ‘fathom the depths’ of the human heart is asserted at 6:63.

contingent as it is on circumstances, Kant seems at least to have allowed for the notion that this level of judgement is something for which the everyday agent, in a misplaced and often self-serving way, shows a great deal of concern.³³

Whether or not Kant ultimately wished to leave space for a genuinely empirical level of morality, Engstrom's alternative points to the fact that those who do 'have their price' are liable to differ in terms of the actual level at which they are willing to transgress the letter of the moral law. Where Kant might not in the end have found this morally interesting, many of us are likely to insist that this constitutes an important moral distinction: one connected to a particular form of strength and integrity. In this chapter, I hope only to have provided reasons for thinking that Engstrom's non-rigoristic proposal conflicts fundamentally *both* with significant elements in Kant's conception of autonomy and with the moral egalitarianism at the heart of his moral theory. Where §5.1 established a connection between the epistemological-motivational element of autonomy and the idea of a supreme maxim, §5.2 explained why the endorsement of autonomy also leads to the view that there are only two types of fundamental character. The argument of §5.3 provides independent support for the necessity of Kant's conception of character, in the sense that it relies not on an appeal to the apparatus of autonomy, but on the Kantian conception of moral worth. Because it rules out a strong and intuitive way of conceiving of a middle ground between complete goodness and complete evil at the level of character, it also provides independent support for Kant's rigorism.

³³ On the relationship between being satisfied with this level of evaluation and self-serving self-deception, see *Rel* 6:38.

Engstrom's worry is, in part, that Kant's conclusions about character and moral worth are in tension with experience, and so represent an unreasonable ideal against which few will measure favourably. Although I have been defending what I take to be the orthodox Kantian view against his alternative, it is nonetheless interesting to consider the degree to which Kant might have been concerned about our ability to develop a lasting, settled and comfortable commitment to morality, despite our obvious and natural attachment to the sensuous side of our nature. How might human beings, saddled as they are to desire and to the propensity to make excuses for themselves, move beyond experiencing the moral law as a harsh and foreign inconvenience? How might they begin to develop a lasting attachment to the voice of pure reason and the interests to which it directs them? These questions set the agenda for chapter six.

MORAL IDENTITY

As we noted in chapter one, certain of Kant's followers objected to what they perceived to be an unpalatable tyranny of reason and its law. Put simply, this worry hinges on the idea that there is a fundamental tension between the supposed freedom of the human being and its being subject to the dictate of reason.¹ Though Kant locates the source of the moral law in the human being's own will, this law does not stem from the self that we know or experience; rather, it finds its source in a mysterious, otherworldly, and utterly impersonal faculty, which imposes its law with unforgiving strictness. But if this is truly the Kantian picture – if the moral command finds its seat in some magisterial 'higher self' rather than *my* free choice – then what sense is there in the assertion that this self is *mine*? Is 'autonomy' anything more than a clever deception?

¹ This tension is echoed by Hannah Arendt: 'Does reason then command the will? In that case the will would no longer be free but would stand under the dictate of reason. Reason can only tell the will: this is good, in accordance with reason; if you wish to attain it you ought to act accordingly' ('Some Questions of Moral Philosophy,' *Responsibility and Judgement*, ed. Jerome Kohn [New York: Schocken Books, 2003] 71). Note that this comment would seem to assume that a command actually compels.

This strand of concern is deeply connected to a related worry, alluded to in previous chapters, about the possibility of developing a lasting and settled commitment to morality in the face of the obstacles afforded by our sensuous nature. Kant accepts that we tend to feel most at home with the affective side of our being – that we are powerfully and inescapably concerned with the satisfaction of our desires. When accepted in conjunction with the idea that we are likely to experience the moral law as a tyrannical injunction to self-denial, however, it becomes relevant to ask how we might find it within us to form a thorough and sincere attachment to the moral path in the face of life's many obstacles. How is the deep-seated moral hesitation and self-deception which so plagues finite rational beings to be overcome?

I have attempted thus far to show that we have good reason to embrace many species of Kant's strictness and rigorism, given a commitment to the categoricity of the moral law and to autonomy. To the degree that we are genuinely attached to the latter and accept that we are both sensuous and moral beings, we cannot coherently object to the features we have been examining. This, indeed, is just one of Kant's most central points. The moral law and the principle of happiness point to two radically different practical options, and there is no midpoint between them. Yet with this acknowledged, my aim in this chapter is to bring us back to the objections identified at the beginning of this thesis and to consider whether the Kantian might, after all, be able to offer solace to those who advance them beyond pointing out that his system requires it. My central questions will thus concern whether we have any reason to think that Kant was concerned about the apparent despotic character of reason in his ethical theory and, perhaps more importantly, whether there are resources in his writings which might allow us to address this worry on its own terms.

I hope to show that there is indeed scope within Kant's ethical thinking for treating this concern seriously. In particular, I will be suggesting that attention to Kant's writings on moral education yields an interesting and subtle picture of the means by which human beings might experience a fundamental shift in their self-understandings, and so to identify the voice of pure reason as that of their own best self rather than a foreign commanding power. What we find in this area of Kant's writing is the thought that coming to grasp the awesome and utterly unique nature of the moral law is actually *transformative* of self-understanding, such that the experienced foreignness of its source is not inevitable. Rather than shaping his conception of the source and nature of the moral law to fit the self-conceptions and attachments of the common human being *as they are*, Kant supposes that it is our responsibility to develop our self-understandings in accordance with what his transcendental arguments have shown morality to be like. What is more, however, I will argue that this is precisely the fit we should maintain in developing a truly normative and prescriptive Kantian moral philosophy. Given that the self-understandings of average human agents might not accurately represent what and who we are capable of being – given that identity can be highly distorted and limiting – any moral theory that insists on coherence with our settled sense of self as a condition of adequacy for its central tenets is bound to be unacceptably passive.² Kant's theoretical starting point is not the question of who we happen to think we are, but rather what we think we *ought* to do and to become. To the degree that this is so, the aim is first to determine the *a priori* principles on which genuinely moral action might be grounded, al-

² This kind of concern about passivity and a lapse into descriptive ethics is voiced by Barbara Herman in 'Responsibility and Moral Competence' (*Moral Literacy* 92), where she criticises the trend of deriving our conception of what a person could have reason to do from a conception of the 'however-formed' adult, with her given dispositions and motivations.

lowing this discovery, in the end, to tell us something significant about the capacities we share, the dispositions we might develop, and the kinds of beings we essentially are. As I hope to show, Kant is indeed concerned about the possibility of feeling at one with the voice of pure reason – of experiencing autonomy as *self*-legislation. What he rightly refuses to do, however, is to privilege the attachments and identities we might happen to have formed as human beings and to adjust his supreme principle to that with which they allow us to be comfortable.

6.1 *The Tyranny of Reason*

At various points in his moral philosophy, Kant's language straightforwardly confirms the role of pure reason as a compelling, confrontational force. In Chapter III of the second *Critique*, for example, the formation of a moral maxim is presented as presupposing a need to be 'impelled to activity' by some power because of the opposition provided by inclination (5:79). In effect, this opposition accounts for the imperatival form of the moral law; it is because we do not possess a holy will from which moral action would proceed analytically, but rather a mixed will, affected by needs and sensuous motives, that our relation to the moral law is one of dependence and constraint (5:32). While the human will is not necessarily 'obedient' when it comes to the moral law, obedience is demanded of it.

Similar passages emphasising the centrality of notions such as command, obedience, dependence, submission, and constraint can be found throughout Kant's ethical writings, a particularly clear example of which is to be found in his discussion of the nature of reverence for the law:

There is something so singular in the boundless esteem for the pure moral law stripped of all advantage – as practical reason, whose voice makes even the boldest evildoer tremble and forces him to hide from its sight, presents it to us for obedience – that one cannot wonder at finding this influence of a mere intellectual idea on feeling quite impenetrable for speculative reason and at having to be satisfied that one can yet see *a priori* this much: that such a feeling is inseparably connected with the representation of the moral law in every finite rational being. (*KpV* 5:79-80)

Passages such as this evince Kant's rhetorical flair, and we must be cautious in attempting too literal a reading of reason's commanding power. Nevertheless, Kant's appeal to the terrifying power of pure reason may prompt us to wonder whether he avoids the spectre of natural necessity only to saddle himself with a version of moral tyranny.

Far from constituting a novel observation, the charge of moral tyranny can be traced back to criticisms made of Kant's moral philosophy by thinkers who were more or less his contemporaries. Hegel's 'Spirit of Christianity and Its Fate' (1798-1799), for instance, uses imagery of just this sort when considering Kant's ethical project. This work begins with a criticism of the self-alienation at the heart of the Hebrew religion, which Hegel compares to Kantian morality. Between those who are subject to positive religion and the Kantian agent, he writes, 'the difference is not that the former make themselves slaves, while the latter is free, but that the former have their lord outside themselves, while the latter carries his lord in himself, yet at the same time is his own slave.'³ Far from embodying the idea of harmonious self-expression, au-

³ In *On Christianity: Early Theological Writings*, trans T. M. Knox (New York: Harper, 1961) 211. For the original German, see 'Der Geist des Christentums und sein Schicksal,' *Hegels theologische Jugendschriften*, ed. Herman Nohl (Tübingen: J. C. B. Mohr, 1907) 266. Quoted in Henry Allison, *Kant's Theory of Freedom* 185; and Allen Wood, 'The Emptiness of the Moral Will,' *The Monist* 72 (1989): 457.

tonomy is described in this work as involving a tyranny of the rational over the sensuous part of the self.⁴ The Kantian moral law is not a law legislated by the whole self, but is rather the *a priori* law of a supersensible self, which imposes its rule on the baser elements of the will without regard to their habits or tendencies. As such, the charge is that the Kantian doctrine of autonomy internalises moral bondage rather than abolishing it, gilded in the rhetoric of freedom though it may be.⁵

Prior to Hegel's criticism, Friedrich Schiller identified in a strikingly similar way what he took to be an unduly harsh picture of the moral life in Kant's ethics. Schiller's most famous criticism of Kant is contained in his epigram on the 'scruples of conscience,' where the focus is Kant's thoughts on moral worth.⁶ In addition, how-

⁴ Allison, *Kant's Theory of Freedom* 185.

⁵ This criticism is articulated in Wood, *Hegel's Ethical Thought* 128-29.

⁶ It is a matter of controversy whether this epigram represents a genuine attack on Kant's moral philosophy or a parody of those who criticise it. For helpful discussions on Kant and Schiller, see Jens Timmermann, *Kant's Groundwork of the Metaphysics of Morals: A Commentary* 152-54, Allen Wood, *Kant's Ethical Thought* 28-9, H. J. Paton, *The Categorical Imperative* 48-50, Henry Allison, *Kant's Theory of Freedom* 180-84, and Marcia Baron, 'Acting from Duty,' *Groundwork for the Metaphysics of Morals*. Ed. Allen Wood (Binghamton, NY: Yale University Press, 2002) 92. The translation runs as follows:

Scruples of Conscience

Gladly I serve my friends, but alas I do it with inclination.
And thus I am frequently nagged by my lack of virtue.

Decision

There is no other advice, thou must seek to despise them,
And do with disgust what thy duty commands.

This translation is provided by Jens Timmermann (*Kant's Groundwork: A Commentary* 152). The original German reads: 'Gewissensskrupel / Gerne dien ich den Freunden, doch tu ich es leider mit Neigung / Und so wurmt es mir oft, daß ich nicht tugendhaft bin. // Decisum / Da ist kein anderer Rat, du mußt suchen, sie zu verachten. / Und mit Abscheu alsdann tun, wie die

ever, an important and lesser-known challenge to Kantian autonomy is also developed in the essay, ‘On Grace and Dignity,’ written in 1793.⁷ Schiller’s comments in this text are most likely directed at Kant’s mode of representation rather than the substance of his doctrine;⁸ whilst he expresses a basic agreement with Kant when it comes to the shortcomings of eudaimonism, he nonetheless complains that, because of Kant’s extreme rhetoric – in particular, because of the imperativ form which he insists on giving the moral law – he creates the impression that the principle of morality is a ‘foreign’ law, through which reason tyrannises over an oppressed sensuous side of the self. As Schiller writes, where the moral weakling (*Weichling*) relaxes the moral law and grounds it in pleasure, Kant seems to have exaggerated its harshness, causing the fulfilment of duty to be interpreted by the great number of his readers as little more than a laudable form of servitude.⁹ As a result of the harshness of Kant’s presentation and the strictness with which he separates the will’s two fundamental interests, the supreme principle of morality was bound to take on the appearance of a

Pflicht dir gebeut.’ (Friedrich Schiller, *Werke*, vol. I, ed. Julius Peterson and Friedrich Beißner [Weimar: Hermann Böhlau Nachfolger, 1943] 357).

⁷ For much of this discussion I am indebted to Henry Allison, *Kant’s Theory of Freedom* 180–84.

⁸ Allison, *Kant’s Theory of Freedom* 180. It is not entirely clear whether Schiller’s conception of grace (developed in this essay) is meant to constitute a criticism of Kant’s conception of moral worth. There are many significant passages in Schiller’s text which indicate that he agreed with Kant on the idea that moral worth must concern only the agent’s principle of action, adding only that we ought to ask more of the moral agent. For more on this interpretation, see Guyer, *Kant and the Experience of Freedom* 352.

⁹ *Über Anmut und Würde* 219.28–220.12. Presumably, the reference to the ‘moral weakling’ here is directed at the Epicurean (this is, indeed, how some translate ‘*Weichling*’ in this context). The relevant German reads as follows: ‘*Weil der moralische Weichling dem Gesetze der Vernunft gern eine Laxität geben möchte, die es zum Spielwerk seiner Konvenienz macht, mußte ihm darum eine Rigidität beigelegt werden, die die kraftvollste Äußerung moralischer Freiheit nur in eine rühmlichere Art von Knechtschaft verwandelt?*’

foreign (fremden) and positive law,¹⁰ inherently opposed to and in conflict with our sensuous nature, and so incapable of participating in a harmony of the self.

According to Schiller, ‘Kant was the Draco of his age, because the latter seemed to him to be not yet worthy of a Solon.’¹¹ Thus he explains the harsh imperatival tone given to the moral law as a consequence of Kant’s concern radically to discredit and to distance himself from moral theories which ground the moral law in the principle of happiness. Kant’s theory can indeed be saved from the charge of tyranny with its central tenets intact, on Schiller’s view, but only in the wake of a careful elision of the strict separation of the rational and the sensuous which so deeply characterises his ethical works.

6.2 *Confronting the Moral Law*

At first glance, the above criticisms may seem imprecise. As we saw in chapters four and five, Kant does not hold that pure reason commands anything at all to our sensuous nature *per se*, and nowhere in his ethics do we encounter the idea that pure reason is meant to keep inclination in check in any direct fashion. Indeed, pure reason has nothing at all to say to our inclinations as they are, but speaks only to the faculty of choice, representing the moral law to this power when the latter has been prompted

¹⁰ *Über Anmut und Würde* 219.28-220.12.

¹¹ *Über Anmut und Würde* 219.16-23. As has been frequently noted, Schiller’s ethical thought can be thought of as an apology of sorts for Kant’s harsh presentation. He thus took it upon himself to correct this harsh picture of morality by emphasising the role in the moral life for the sensuous side of human nature. For Schiller, the best soul is one in which reason and inclination exist in perfect harmony, directed toward the same end. Where this harmony is present, we have a ‘beautiful soul’ (*schöne Seele*). See 221. 4-223.31.

choice, representing the moral law to this power when the latter has been prompted by inclination to pursue a certain practical option. What the pure will or the legislative faculty forbears and keeps in check, then, is a tendency on the part of the faculty of choice to grant priority to the prompting of inclination, and so to adopt the principle of happiness.

This notion is spelled out quite clearly in Kant's discussion of ancient Stoicism in the *Religion*.¹² According to Kant, these 'valiant men' quite mistook their enemy when they identified the natural inclinations as being morally dangerous. If they can be faulted at all, inclinations merely lack discipline, presenting themselves to the will yet only ever affecting the faculty of choice, never determining it. Considered in themselves, they are not at all reprehensible, and to wish to be rid of them 'would not only be futile but harmful and blameworthy as well' (6:58). Instead, what they require is conditioning, organisation and harmonisation, whether this occurs by reasoning from prudence or morality:

Thus when the Stoic thought of the human moral battle simply as a human being's struggle with his inclination, so far as these (innocent in themselves) must be overcome as obstacles in the compliance to his duty, he could locate the cause of the transgression only in the omission to combat them, since he did not assume any special positive principle (evil in himself) ... (6:59)

The true opponent of goodness, with whom reason must stand in combat, is not the inclinations or our sensuous nature *per se*, but rather an underlying practical principle: the propensity to give favour to inclination over the moral law, which finds its source in the faculty of choice alone. As such, if the above criticisms are to be made more precise and more fitting to Kant's theory, they must represent the relevant

¹² See 6:57-59.

‘slave of reason’ as a faculty of choice which is fundamentally attached to the satisfaction of needs stemming from the human being’s sensuous nature. Only in this way can the criticism fit with Kant’s theory of practical reasoning and practical freedom.

Let us suppose, then, that such an amendment may be made. Like Schiller, we might wish to reserve this kind of criticism for Kant’s mode of presentation when it comes to the moral law. Quite interestingly, however, a number of passages in Kant’s texts would seem to suggest that a momentary experience of strangeness and alienation – of being inescapably subject to a law which finds its source in a place we cannot know – *really is* part and parcel of common moral experience. Consciousness of the moral law is described by Kant as ‘forcing itself upon’ our faculty of choice: as arising from an active force of pure reason ‘by which it announces itself as originally lawgiving’ (*KpV* 5:31). This consciousness is, in addition, ‘absolutely inexplicable from any data of the sensible world and from the whole compass of our theoretical use of reason’; we cannot trace it to the world of experience or to our nature as it appears to us, but rather are directed by means of this consciousness to a world that is altogether different from that which we experience with the senses (a ‘pure world of the understanding’) (*KpV* 5:43).¹³ That there is an active and invisible ‘something else’ lurking behind the objects of the senses is a conclusion that can be found in even

¹³ As Kant goes on to explain, theoretical reason cannot sustain speculation that reaches beyond experience (see, e.g. *KrV* A 244/ B 303), such that consciousness of the moral law points to a belief in something – perhaps even a fundamental aspect of the self – the existence of which cannot be proven theoretically. In the *Groundwork*, this connection is depicted as one half of a threatening circle in the argument of Section III: ‘For 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; but if we think of ourselves as put under obligation [*als verpflichtet*] we regard ourselves as belonging to the world of sense and yet at the same time to the world of understanding’ (4:453).

the most common understanding (though due to our attachment to the sensuous we continually spoil this train of thought, and our chances to be wiser, by representing the invisible as an object of intuition) (*G* 4:451-52). Though we do not *experience* ourselves as members of the intelligible realm, it is because we are conscious of ourselves as bound by the categorical imperative that we count ourselves as members of an intelligible, supernatural realm subject to its very own laws.

Despite this appeal to a genuine sense of otherness in the experience of duty, Kant seems to have thought that this sense of the otherworldly, of the foreignness of the moral law to our sensuous nature, need not – indeed, ought not – bring with it feelings of being unpalatably enslaved by reason. As he writes in a response to Schiller’s essay, we must of course portray the concept of duty as being concerned with unconditional necessity; as Kant readily admits, he is ‘unable to associate gracefulness with the concept of duty by reason of its very dignity.’ Yet with this said, he also holds the following:

[I]f we ask, ‘What is the aesthetic constitution, the temperament so to speak of virtue: is it courageous and hence joyous, or weighed down by fear and repressed?’ an answer is hardly necessary. The latter slavish frame of mind can never be found without a hidden hatred of the law, whereas a heart joyous in the compliance with its duty (not just complacency in the recognition of it) is the sign of genuineness in virtuous disposition ... (*Rel* 6:23-24.n)

Thus while Kant denies that the concept of duty can be connected with grace and that any sentiment can be a worthy motive for the performance of duty, he clearly allows that a general attitude of feeling anxious and repressed by duty indicates a lack of genuine commitment to morality. This is not to say that the human being does not experience duty as a genuine constraint on her actions. To deny this would be to deny the necessitating property of pure reason’s law. Rather, Kant distinguishes the kind of

constraint experienced by the agent who is properly committed to the moral law from that of the agent who has inwardly rejected the maxim of duty. Where the former experiences constraint without repression, the latter faces the moral law with a telling anxiety and dejection. Though it may be rare, a virtuous disposition will be manifest in a heart which is joyous.

We encounter a similar conjunction of otherworldliness and joyous elevation in Kant's many discussions on the moral feeling of reverence for the law. As is often noted, Kant initially characterises respect through the language of humiliation and pain. These strong feelings are a negative effect of the influence of practical reason, the result of pure reason's infringement 'upon the activity of the subject so far as inclinations are his determining grounds and hence upon the opinion of his personal worth' (*KpV* 5:78). As Kant explains,

[a]ll inclination and every sensible impulse is grounded on feeling, and the negative effect on feeling (through the check exercised on feelings) is itself a feeling. Consequently we can see a priori that the moral law as the determining ground of the will, since it works to the prejudice of all of our inclinations, must effect a feeling which can be called pain, and here we have the first, perhaps also the only case where we can determine the relation of a cognition (here it is one of pure practical reason) to the feeling of pleasure or displeasure a priori from concepts. (*KpV* 5:73)

The moral agent is confronted with the representation of the moral law and realises that he has a free choice to make it the determining ground of his will. The representation is attended by pain, to the degree that it engenders a comparison by the agent between 'the sensible propensity of his nature' and the moral law (*KpV* 5:74).¹⁴ When we compare ourselves with the 'holy' moral law, we discover how remote we

¹⁴ See also *KpV* 5:79.

are from complete accordance with it; the comparison is thus not between ourselves and other human beings, but between ourselves and the moral law.¹⁵

With this crucial aspect noted, however, the painful frustration of inclinations is by no means the only element in reverence for the law. According to Kant, in the process of feeling wonder and awe for the moral law, we also experience respect for ourselves, and indeed for all other rational beings. Because we recognise that we are able to deny our own inclinations, the painfulness of reverence reveals in us the presence of a higher faculty, pure practical reason, and this generates a dramatically positive feeling. Thus ‘the lowering of pretensions to moral self-esteem – that is, humiliation on the sensible side – is an elevation of the moral – that is, practical – esteem for the law itself on the intellectual side’ (*KpV* 5:79). Since the relevant constraint is exercised by the lawgiving power of the agent’s very own faculty of reason, ‘it also contains something elevating, and the subjective effect on feeling, inasmuch as pure practical reason is the sole cause of it, can thus be called self-approbation with reference to pure practical reason’ (*KpV* 5:80-81).¹⁶ This sense of independence from the mechanism of nature ‘elevates a human being above himself (as a part of the sensible world)’ and ‘connects’ him with an order of things that is radically different from the realm of sensible objects (*KpV* 5:86). Thus, while our experience of reverence involves both self-frustration and a sense of otherworldly awe, it is ultimately a feeling of connection to a lawgiving power which demands our unconditional compliance –

¹⁵ See *VC* 27:349.39-50.3.

¹⁶ As Paul Guyer notes (*Kant and the Experience of Freedom* 358), some commentators take Kant to be describing two separate feelings: the painful feeling of respect and a positive feeling (perhaps the ‘moral feeling’), based on admiration for the superiority of the moral principle. I agree with Guyer in rejecting this interpretative move.

of identity with an active, legislative force which transcends our immediate cognition. Far from merely pointing to a realm beyond the senses, moral experience is described by Kant as bringing us to a sense of belonging and self-appreciation.¹⁷ It is of course the moral law which is the primary object of our awe – an awe ‘which rouses respect of the subject toward his master.’ In our case, however, ‘since the master lies in us, it rouses a feeling of the sublimity of our own vocation that enraptures us more than any beauty’ (*Rel* 6:23.n).¹⁸

6.3 *Identity and Moral Education*

It might seem to us that Kant’s discussion of reverence overstates the feeling of connection and self-respect that accompanies consciousness of the moral law and its unconditional strictness. After all, if it is indeed possible for a sensibly bound human being to come to identify himself as a lawgiving member in a kingdom of ends, it is natural to think that this is likely to require much development and self-adjustment. Though the positive element of reverence is described by Kant as an experience open to any rational being whatever, it is clear that he is also concerned by the human tendency to weaken, or in some cases to eclipse altogether, its lasting effect on self-understanding. Obstacles to our identification with the voice of pure reason can come in many forms, some of which I have already mentioned (see §4.4). Quite in-

¹⁷ The complex duality of feeling involved in reverence (its combination of frustration and self-reflected pleasure) is akin to the complexity involved in the feeling of the mathematical and the dynamical sublime, as discussed in *The Critique of the Power of Judgement* (see esp. 5:248-264).

¹⁸ See also *KU* 5:257, 262.

terestingly, however, the movement towards this species of harmonious identification is a significant element in Kant's thoughts on moral education, where Kant depicts a connection between securing moral self-understanding and strengthening the influence of the moral law when it comes to the selection of maxims.

Here, we embark upon murky waters, for one of the more difficult issues facing those who wish to make sense of Kant's theory of moral education concerns its awkward fit with transcendental freedom and with Kant's claim that the intelligible world grounds the empirical. If it is true that moral character is the result of a radically free choice, independent of the chain of natural causality, what real effect can empirical lessons and practices have with regards to our fundamental moral disposition? What is to be avoided, it would seem, is the importation of determinism into the conditions of moral agency, such that the freedom of morality is in any way threatened. Yet just how this is to be effected, and how Kant's claims about moral education are to be explained within the limits set by transcendental freedom, are quite difficult questions.¹⁹ While answering these extremely difficult questions is beyond the scope of my im-

¹⁹ Various interesting attempts to deal with this question are found in the recent literature on the empirical side of Kant's moral theory. Barbara Herman, for instance, suggests that we ought to view the kinds of empirical helps mentioned in Kant's accounts of moral education as epistemic aids: as means of developing an understanding of the 'rules of moral salience,' by which the morally significant features of circumstances are brought into view (see *The Practice of Moral Judgement* ch. 4). Others, such as Robert Louden, Allen Wood, and Felicitas Munzel interpret these empirical aids as steps which either *help* to prepare free agents to choose well, or which *enable* agents to choose well, functioning as propaedeutics to moral agency (See Louden, *Kant's Impure Ethics* ch. 2; Wood, *Kant's Ethical Thought* esp. Part II; and Munzel, *Kant's Conception of Moral Character* esp. ch. 5). See also Patrick Frierson's recent account in *Freedom and Anthropology in Kant's Moral Philosophy*, where he argues that empirical helps and hindrances affect not the good will itself, but its appearance in the world, in a way which is nonetheless morally relevant. See also Beck, 'Kant on Education,' *Essays on Kant and Hume* (New Haven: Yale University Press, 1978).

mediate project, my aim is to investigate how Kant's thoughts on the development of moral self-understanding might also gesture toward a solution to the problem of moral commitment.

Kant's thoughts on moral education find their home in a variety of texts, and to touch on each example would take us much too far afield. For our purposes, then, we shall focus principally on the discussion of education set out in the second *Critique's* Doctrine of Method, as it is here that the idea I wish to explore is set out in its most clear and concise form.

As Kant explains in the opening lines of this section, a doctrine of the method of pure practical reason is a study of the way in which we can provide the moral law with 'access to the human mind and influence on its maxims' (*KpV* 5:151).²⁰ He acknowledges that it must seem to most of us improbable that the presentation of the moral law could come to have more power over our minds than 'all the deceptive allurements of enjoyment' (5:151). Even though the *Groundwork* and the previous sections

²⁰ As is often noted, the Doctrine of Method can be seen as an enquiry into the grand old Enlightenment questions: By what course are human beings brought to a condition where the moral disposition may be engendered and flourish? How can the child be brought from innocence to moral maturity? See, e.g. Beck, *A Commentary on Kant's Critique of Practical Reason* 233. It is important to emphasise that the Doctrine of Method differs quite substantially from the *Groundwork* and the more doctrinal passages of the second *Critique* in that it points to observations that have actual practical effects, whereas the latter two texts are primarily theoretical exercises. What Kant is doing, here, is raising a very particular set of questions that have to do with how we might become better moral agents. By contrast, he never thought that the primary purpose of reading the *Groundwork* is to become morally better (although there are elements, as I have noted, that would seem to serve this role at least in part). Although I focus on the Doctrine of Method of the second *Critique* in this chapter, see also the Doctrine of the Method of Ethics of the *Metaphysics of Morals*, which also outlines a programme for teaching people to be morally good.

of the second *Critique* have shown us that the possibility of a pure moral incentive must be presupposed as a condition of the possibility of morally good action, this does not in itself give us reason to think it probable that such an incentive will in fact have a more lasting influence on the human faculty of choice than those provided by inclination. We understand at this point that the *in principle* sufficiency of the moral incentive follows from accepting that we are bound by the moral law; knowing what we do about ourselves, however, why should we think it likely that any of us really *would* choose the hard and lonely road of virtue? Given the strength of our desires and the attractiveness of happiness, how could it ever seem reasonable to suppose that something as formal and seemingly empty as the moral law can provide creatures like us with a stronger incentive to action? Kant's hope in this section is to provide anecdotal proofs of the human receptivity to the moral incentive, along with a sketch of the method by which to facilitate the cultivation of a truly moral disposition (5:153).

For those whose minds are still uncultivated or already degraded, Kant allows that being brought onto the rails of morality requires being attracted by threats and promises connected to happiness. As soon as these 'leading strings' have had even the slightest effect, however, 'the pure moral motive must be brought to bear on the soul' (5:152).²¹ One of the exercises highlighted by Kant as being extremely useful ~~for moral improvement is engagement in argumentation concerning the moral worth~~

²¹ Here again, the moral motive is described by Kant as giving to the human being the power, 'unexpected even by himself,' to tear himself away from all sensible attachments so far as they want to rule over him and to find a rich compensation for the sacrifice he makes' (5:153). According to Kant, and somewhat problematically given what he says about freedom in the *Religion*, this is a method for founding and cultivating genuine moral 'dispositions' (5:153). What Kantians would seem to be unable to say is that such a method is what makes the choice of a good moral disposition of the sort discussed in chapter five *possible*. What there may be room for, however, is the idea that the undertaking of such a method makes it all the more *easy* for ordinary moral agents to develop into the kinds of people who are committed to morality above inclination. More will be said about this at the close of this section.

moral improvement is engagement in argumentation concerning the moral worth of people's deeds, as laid out in examples.²² Here, the primary aim is to become able to identify, as best one can, the nature and characteristics of good conduct. In addition to this skill, however, the further hope is that the agent will develop an attitude of approval of the examples of moral goodness picked out in discussion and argument and regret or contempt for examples of deviation from it.²³ Kant's emphasis on argument rather than passive instruction presumably hinges on the notion that the former method allows developing agents actively to engage in the search for, and construction of, reasons for ascribing moral predicates (good, evil) to certain examples of agents and their deeds.²⁴ As Kant argues, this process of engaging in argument and critical discussion with others is observed to result in a certain habit of responding to

²² For a very helpful discussion of Kant's thoughts on the role of examples in Kant's moral philosophy, see Robert Louden, 'Go-Carts of Judgement: Exemplars in Kantian Moral Education,' *Archiv für Geschichte der Philosophie* 74 (1992): 303-22. See also Louden's 'Making the Law Visible: The Role of Examples in Kant's Ethics,' *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*, ed. Jens Timmermann (Cambridge: Cambridge University Press, forthcoming).

²³ C.f. the *Lectures on Pedagogy*: 'Everything in education depends on establishing the right principles throughout and making them comprehensible and acceptable to children. Children must learn to substitute detesting things because they are disgusting and absurd for abhorrence out of hatred; they must learn to have inner aversion replace the external aversion to human beings or to divine punishment, to have self-estimation and inner worth replace the opinions of human beings, to have the inner value of actions and deeds replace words and emotions, understanding replace feeling, and joy and piety with good humour replace morose, timid, and gloomy devotion' (*UP* 9:492-3). Educators, ought, however, to leave out examples of noble or supermeritorious actions. '...whatever runs up into empty wishes and longings for inaccessible perfection produces mere heroes of romance who, while they pride themselves on feeling for extravagant greatness, release themselves in return from the observance of common and everyday obligation, which then seems to them insignificant and petty' (*KpV* 5:155).

²⁴ The importance of *activity* and *independence* of mind in moral education is a theme which lies at the centre of Kant's thinking on the matter more generally, and here there is no exception.

people's behaviour – a quite lasting impression of esteem for the good and disgust for the evil, which 'would make a good foundation for uprightness in the future conduct of life' (5:154-5). Gradually, the appraisal of actions becomes a 'natural occupation' – 'a habit accompanying all our own free actions as well as our observation of those of others' (5:159).²⁵ We are then able to sharpen our appraisal by asking first whether the action conforms with the moral law, and then the nature of the more specific principle with which it accords (does it, for instance, accord with a principle grounded on what the needs of human beings require of one?) (5:159).²⁶

In addition, however, Kant's method proposes that we direct our attention to whether the action was in fact done for the sake of the moral law. This involves a cultivation of our reason in making judgements about the practical, where the result is a gradual production of 'a certain interest in reason's law itself and hence in morally good actions.' As Kant goes on to suggest, at this stage

we finally come to like something the contemplation of which lets us feel a more extended use of our cognitive powers, which is especially furthered by that in which we find moral correctness,

²⁵ Though Kant speaks in terms of building up a habit of appraisal, by no means does he believe that moral education ought to have as its final aim the development of mere habit. As he writes in the *Lectures on Pedagogy*, moral education 'is based not on discipline but on maxims. Everything is spoiled if one tries to ground this culture on examples, threats, punishments, and so forth. Then it would be merely discipline. One must see to it that the pupil acts from his own maxims, not from habit, that he not only does the good, but that he does it because it is good. For the entire moral value of actions consists in the maxims concerning the good. Physical education differs from moral education in that the former is passive for the pupil while the latter is active. He must at all times comprehend the ground of the action and its derivation from the concepts of duty (UP 9:475).

²⁶ C.f. the *Lectures on Pedagogy*: 'If one wishes to form a character in children, it is very important to draw their attention to a certain plan in all things, certain laws, known to them, which they must follow exactly' (UP 9:481).

since only in such an order of things can reason, with its capacity to determine a priori in accordance with principles what ought to be done, find satisfaction. (5:160)

This exercise thus develops in us a *mediate* interest in moral correctness; we value the latter not in itself, but because it occasions the harmony of our cognitive powers. The existence of moral correctness in and of itself remains indifferent to us, because ‘it is viewed only as the occasion of our becoming aware of the tendency of talents in us which are elevated above animality’ (5:160).

The second major exercise in the course of the method involves bringing to attention – again, through the use of examples – the purity of the good will. Here, the primary focus is the fact that incentives of inclination have no influence on the will in morally good action. According to Kant, this works to free the agent from the feelings of dissatisfaction and disappointment that accompany concern for sensuous needs, such that his mind is made receptive to a radically different source of satisfaction. By this second exercise, then,

the pupil’s attention is fixed on the consciousness of his freedom and, although this renunciation excites an initial feeling of pain, nevertheless, by its withdrawing the pupil from the constraint of even true needs, there is made known to him at the same time a deliverance from the manifold dissatisfaction in which all those needs entangle him ... (5:160)

As Kant goes on to suggest, the pupil’s heart is by this method ‘freed and relieved of a burden that always and secretly presses upon it’ (5:161). The effect of setting before him examples of pure moral resolutions in which are revealed an otherwise unknown inner capacity is that he develops ‘the inner freedom to release himself from the impetuous importunity of inclinations so that none of them, not even the dearest, has any influence on a resolution for which we are now to make use of our rea-

son' (5:161).²⁷

As mentioned above, the observations of the Doctrine of Method are meant to show the reality of reverence for the law – proofs of our receptivity to it – and ‘the method of founding and cultivating genuine moral dispositions’ (5:153). It takes the form of a response to the articulation of a doubt – *can the moral law really effect more power over our limited human minds than inclination?* In this sense, this section can also be seen as answering the old charge that Kantian ethics is ‘false to human nature’ – that it puts unnatural demands on us to forsake our character and our sensuous concerns. What Kant’s writing suggests is that human beings might develop a strong pleasure and satisfaction in being able to stand above the sensuous needs which ‘entangle’ them: a sense of being able to withdraw from the constraints set out by nature and a sense of deliverance from the dissatisfactions which accompany our existence as natural, dependent beings. The lifting of this burden is something which, according to Kant, we human beings will welcome, once we’ve experienced it and trained our minds to be receptive to its significance. Yet this shift in consciousness, this development of a sense of immense relief and peace in the realisation of practical freedom, is the result of an altogether natural process of education, observable, or so Kant thinks, by any human being whatever. The development of a receptive feeling of satisfaction in a radically different object is quite easily explainable as the result of the activity of the mind in contemplating the principles and motivational springs which in some sense have always grounded our moral evaluations. From this state of

²⁷ Kant goes on to note that ‘through the positive worth that observance of it lets us feel, [the moral law] finds easier access through the respect for ourselves in the consciousness of our freedom ... When this is well-established ... then every good moral disposition can be grafted onto it, because this is the best, and indeed the sole, guard to prevent ignoble and corruptible impulses from breaking into the mind’ (5:161).

mind, the constraint effected by the moral law is distinctly not experienced as a form of repression or fundamental self-denial. It is, rather, experienced as the source of a kind of freedom from the exhausting demands of desire not available to non-rational beings – the source of a unique form of escape and ensuing peace of the soul. In short, it is experienced as the source of a unique and otherwise unavailable option for the living of a human life, and for this the well-adjusted pupil of morality feels not dejection, but immense self-satisfaction.

Attention to Kant's thoughts on moral education thus reveals his concern to articulate the means by which agents might learn to identify with the voice of pure reason. In addition to autonomy, such an agent will possess *integrity*: a sense of unity and completeness denied to those who choose to attach themselves to the baser element of their nature. As a result of having acquired this self-understanding, such an agent is likely to find that compliance with duty comes more easily, that it meets with less resistance, such that he dreads nothing more than the discovery that he has acted in a manner that is morally contemptible. It must of course always be *possible* for any agent to do what the moral law commands, for this is assured by the epistemological and motivational aspect of autonomy. Yet the degree to which compliance with the moral law is made difficult or effortless by one's inner temperament is an area that Kant allows to vary between individuals. We thus encounter, quite interestingly, a second area in which luck might be said to affect virtue on the Kantian view.²⁸ Though the moral worth of one's actions is not at all the kind of thing that can be subject to fortune on Kant's view (see §5.3), luck will indeed affect the nature of one's temperament and one's attachments, and so the level of ease with which one is

²⁸ Recall the notion that luck might be said to affect empirical character (see §5.2).

able to fulfil morality's demands.²⁹ As the Doctrine of Method aims to 'prove,' this feature of our lives is, for better or for worse, altogether conditioned by empirical influence (the having of certain experiences, exposure to the right kinds of teacher). Whilst autonomy promises that the path of virtue is in principle open to any moral agent, it does not guarantee that all shall follow it with parity of ease.

6.4 Identity and Normative Ethics

What the above discussion shows is that the supposed 'tyranny' of reason is not an inevitable fact of moral experience. Given the right frame of mind, finite rational beings are able to bring their self-understandings in line with pure reason, and so to achieve the integrity that characterises a settled comfort with the moral law. Such comfort is not equivalent to perfection or holiness, a property possessed only by the Supreme Being, but rather a self-satisfaction that expresses itself in a lack of moral resentment. Though the dichotomisation of reason and desire is strict, this need only bother those who suspect that our attachment to the latter is either desirable or incapable.

Will Kant's thoughts on the transformative power of moral education when it comes to self-understanding satisfy all who object to his insistence on purity? The answer must surely be no. What is perhaps most significant about Kant's approach to this problem, however, is his refusal to alter the rigoristic conclusions that stem from au-

²⁹ Because this form of luck deals with the kind of person you are (your attachments and temperament), it is similar to what Bernard Williams and Thomas Nagel describe as 'constitutive' luck. See, e.g. 'Moral Luck' 115-16, 140, 144-5.

tonomy in order better to fit the contours of our already-formed attachments and ways of approaching practical problems. As the Doctrine of Method makes clear, it is our responsibility to develop the kinds of environment in which morally helpful self-understandings might flourish. Rather than taking the character of his times as a moral starting point, Kant's moral philosophy begins by setting out *a priori*, from the very idea of unconditional goodness and its fundamental principle, the content of morality and all its implications. Beyond acknowledging that the moral law must address them as an imperative, it does not make special allowance for the particular characteristics of human beings. As a result, Kantian theory is one in which the 'ought' is given theoretical priority over the 'is': if the majority happen to find the confrontation with the moral law uncomfortable or distressing – a species of self-wrought slavery – then so much the worse for humanity.

In the *Lectures on Pedagogy*, Kant is reported to have said the following:

One principle of the art of education, which particularly those of men who are education planners should have before their eyes, is this: children should be educated not only with regard to the present but rather for a better condition of the human species that might be possible in the future; that is, in a manner appropriate to the idea of humanity and its complete vocation. This principle is of great importance. Parents usually educate their children merely so that they fit in with the present world, however corrupt it may be. However, they ought to educate them better, so that a future, better condition may thereby be brought forth. (UP 9:447)

Here, Kant makes clear that education is set the task of moving beyond the given in practical matters – of refusing to conform our practices to the 'present world.' Given that the concerns and attachments of average human agents might accurately represent neither what humanity is capable of becoming nor what it *ought* to become, a truly prescriptive (and in *this* sense normative) moral theory cannot strive merely for

coherence with our settled, if natural, sense of self. The charge of moral tyranny constitutes a valuable criticism of Kantian ethics only to the degree that we understand coherence with our sensuous nature to be a good thing. Where we have independent reason to suppose that conformity with the demands of the latter does not in fact represent our best possible self (our true vocation [*Bestimmung*]), a lack of fit between the prescriptions of moral theory and our given attachments, while it might of course raise important practical and pedagogical issues, does not in itself render suspect the theory's underlying principles. As I hope to have shown, Kant is indeed concerned about the possibility of feeling at one with the voice of pure reason and with experiencing autonomy as self-legislation. What he rightly refuses to do, however, is to privilege our already-formed attachments and identities. In true form, he does not put the end to be achieved – the securing of popular access for the law – before the law itself.³⁰

In his later years, Kant increasingly stressed the significance of the virtue of inward *sincerity* for moral progress and self-development. As he writes in the *Religion*, there is a certain dishonesty at work when people concern themselves only with the appearance of their empirical character and refuse to enquire into their inner maxims. By this means 'we throw dust into our own eyes,' hindering the development of a moral disposition and actually fostering a greater ease when it comes to the deception

³⁰ See, e.g. *G* 4:409: 'This descending to popular concepts is certainly very commendable, provided the ascent to the principles of pure reason has first taken place and has been carried through to complete satisfaction. That would mean that the doctrine of morals is first *grounded* on metaphysics and afterwards, when it has been firmly established, is provided with *access* by means of popularity. But it is quite absurd to want to comply with popularity in the first investigation, on which all correctness of basic principles depends.'

of others (6:38).³¹ A kind of self-deception is also at work, however, when, we allow our deep identification with the ‘dear self’ described in *G* 4:407 to interfere with the (morally significant) recognition that there is more to our nature. An ethics of autonomy is appealing for just this reason: it takes as its starting point a conception of how and who we ought to be, and from there it arrives at a conception of what we must be capable of becoming. It thus points forward to a way of being that, while perhaps difficult to achieve, must nonetheless be possible, rather than deriving prescriptions for living from a backward-looking conception of our settled patterns and dispositions as they happen to be. Because the moral law in an ethics of autonomy is non-derivative – because it does not depend on any antecedent correlation between some state of the world and our given dispositions – its prescriptions and prohibitions are bound to possess an unbounded normativity unparalleled in heteronomous theory. Of course, the most difficult task for such an ethics is to prove that our trust in the categorical imperative as a genuinely binding principle of action is *justified*, such that we have a *right* to think of ourselves as legislating members of an intelligible realm (for analysis alone cannot establish the validity of the moral law). I have not touched on this extremely difficult issue so far in this thesis, and nor shall I here. What I do wish to suggest, however, is that there is an inherent connection between an attraction to autonomous moral theory and a refusal to be bothered by a lack of fit, no matter how fundamental, between ‘ought’ and ‘happens to be.’ Though pure reason’s yoke is gentle, it is simply to be expected on such a framework that the bulk of us shall bear

³¹ As Kant goes on to note, this self-deception ‘puts out of tune the moral ability to judge what to think of a human being, and renders any imputability entirely uncertain...’ (6:38). See also *G* 4:407.11-12 on self-flattery. The other major deception that preoccupies Kant in both the *Groundwork* and the *Critique of Practical Reason* is the doctrine of determinism – yet another bit of sophistry by which we excuse ourselves from responsibility.

it reluctantly.³²

³² On the ‘gentle yoke’ of the moral law, see *KpV* 5:84-5.

CONCLUSION

Kant concludes his *Critique of Practical Reason* by writing what would come to be his most famous passage:

Two things fill the mind with ever new and increasing admiration and awe, the more often and more steadily one reflects on them: *the starry heavens above me and the moral law within me* ... The first begins from the place I occupy in the external world of magnitude with worlds upon worlds and systems upon systems ... 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. The first view of a countless multitude of worlds annihilates, as it were, my importance as an *animal creature*, which after it has been for a short time provided vital force (one knows not how) must give back to the planet (a mere speck in the universe) the matter from which it came. The second, on the contrary, 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 of my existence by this law, a determination not restricted to the conditions and boundaries of this life but reaching into the infinite. (5:161-2)

Kant was well aware that our legislative power, our being as *intelligence*, had always somehow to be reconciled with our sense of ourselves as contingent *animal creatures*. The dual movement that characterises this passage (the inner humiliation that follows the upward and outward gaze toward the infinity of worlds upon worlds; the elevation that follows the inward reflection upon one's own independence, as intelligible being, from the sensuous world) thus traces two views of humanity which Kant was acutely aware he had to accommodate in his moral theory as a whole. In this thesis, I have

sought to provide an explanation of the latter self-conception, a view of the self as a legislating member of a world of intelligences, which nonetheless addresses some of the more troubling problems that emerge from the intersection of this self-understanding with a conception of ourselves as members of the world of sensibility. Where the first half of this thesis sought to account for the claim that the moral law is a law that we ourselves legislate, the second half addressed the problem of the extent to which this conception of ourselves can be integrated with the more familiar side of our nature: with the fact that we have inclinations, temptations, are extremely fallible, and can be wavering in our commitment to even the most rational of principles.

Although I hope to have provided a compelling picture thus far, there are, of course, directions for further investigation that I could not pursue in this study, but that are clearly deserving of future reflection. In chapters two and three, for example, I presented Kant as one who is largely uninterested in providing a morally independent, foundationalist justification for the moral law. Whilst I argued that the moral law must be viewed as possessing a determining ground of some sort, which explains why it is a law for *my* will (for the will of any particular agent), this ground turned out to be nothing other than the categorical imperative's universality and necessity. I offered no independent justification on Kant's part for thinking that the categorical imperative is universally and necessarily binding, and indeed, I altogether avoided the question of how and whether Kant sought a vindication of this sort. There is thus much room for further thought on the question of whether Kant ought to be read as providing (or requiring) a justificatory picture of this more fundamental and ambitious kind and whether autonomy plays any role in it.

Second, there is certainly scope for further investigation into Kant's theory of moral education and development. Given my interests in this study, I avoided some of the more difficult and puzzling questions surrounding the very possibility of a Kantian theory of *moral* development, including especially the important role that he seems to reserve for empirical aids and hindrances to moral goodness in the Doctrine of Method and in his anthropological works. Though recent work in this area has shed light on some of these questions, there is still much work to be done when it comes to reconciling the importance that Kant seems to place on empirical social development and his commitment to the connection between morality and transcendental freedom.

Finally, although this project has been for the large part internal, seeking to provide an interpretation of autonomy that can and ought to be accepted by Kantians, it is certainly worth thinking about the challenges that the view I have been tracing might face from outside the comforts of Kantian ethics and whether (and how) this story might be maintained in light of them. I have not sought at this stage to defend the approach to ethics developed in this thesis against competing views regarding the source of moral requirement, though this would clearly constitute an important development of the work presented here. For those who do not share Kant's starting points, I hope in this study at least to have raised interest in the complexity of his vision of humanity, in the origins of his concern for purity, and in the source of his commitment to autonomy.

BIBLIOGRAPHY

TRANSLATIONS OF KANT'S WORKS

- Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. 1785. *Practical Philosophy*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996. 37-108.
- . *Critique of Practical Reason*. 1788. *Practical Philosophy*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996. 133-272.
- . *Critique of Pure Reason*. 1781, 1787. The Cambridge Edition of the Works of Immanuel Kant. Trans. Ed. Paul Guyer and Allen Wood. Cambridge: Cambridge University Press, 1998.
- . *Critique of the Power of Judgement*. 1790, 1793. The Cambridge Edition of the Works of Immanuel Kant. Trans. Eric Matthews and Paul Guyer. Ed. Paul Guyer. Cambridge: Cambridge University Press, 2000.
- . *The Metaphysics of Morals*. 1797. *Practical Philosophy*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996. 353-604.
- . 'Reflections on Moral Philosophy.' *Notes and Fragments*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Curtis Bowman, Paul Guyer, Frederick Rauscher. Ed. Paul Guyer. Cambridge: Cambridge University Press, 2005.
- . *Religion Within the Boundaries of Mere Reason*. 1793. *Religion and Rational Theology*. The Cambridge Edition of the Works of Immanuel Kant. Ed. Allen W. Wood and George Di Giovanni. Trans. George Di Giovanni. Cambridge: Cambridge University Press, 1996.
- . 'The Conflict of the Faculties.' 1798. *Religion and Rational Theology*. The Cambridge Edition of the Works of Immanuel Kant. Ed. Allen W. Wood and George

- Di Giovanni. Trans. Mary J. Gregor and Robert Anchor. Cambridge: Cambridge University Press, 1996. 233-328.
- . 'On the Common Saying: That may be correct in theory, but it is of no use in practice.' 1793. *Practical Philosophy*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996. 273-310.
- . *Lectures on Pedagogy*. 1803. *Anthropology, History, and Education*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Robert Louden. Ed. Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007. 434-485.
- . *Lectures on Ethics*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Peter Heath. Ed. Peter Heath and J. B. Schneewind. Cambridge: Cambridge University Press, 1997.
- . 'An Answer to the Question: What is Enlightenment?' (1784). *Practical Philosophy*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996. 11-22.

ALL OTHER WORKS CITED

- Adams, Robert Merrihew. *Finite and Infinite Goods: A Framework for Ethics*. New York: Oxford University Press, 1999.
- Allison, Henry. *Kant's Theory of Freedom*. New York: Cambridge University Press, 1990.
- Ameriks, Karl. *Kant and the Fate of Autonomy: Problems in the Appropriation of the Critical Philosophy*. Cambridge: Cambridge University Press, 2000.
- . *Interpreting Kant's Critiques*. Oxford: Clarendon Press, 2003.
- Anscombe, G. E. M. 'Modern Moral Philosophy.' *Philosophy* 33 (1958): 1-19.
- Arendt, Hannah. 'Some Questions of Moral Philosophy.' *Responsibility and Judgment*. Ed. Jerome Kohn. New York: Schocken Books, 2003. 49-146.
- Aristotle. *Nicomachean Ethics*. Trans. Ed. Terence Irwin. 2nd ed. Indianapolis, IN: Hackett Publishing Company Inc., 1999.
- Baron, Marcia. 'Acting from Duty.' *Groundwork for the Metaphysics of Morals*. Ed. Allen W. Wood. Binghamton, NY: Yale University Press, 2002. 92-110.

- Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason*. Chicago: University of Chicago Press, 1960.
- . 'Kant on Education.' *Essays on Kant and Hume*. New Haven: Yale University Press, 1978.
- Benson, Paul. 'Moral Worth.' *Philosophical Studies* 51 (1987): 365-382.
- Bittner, Rüdiger. *What Reason Demands*. 1983. Trans. Theodore Talbot. Cambridge: Cambridge University Press, 1989.
- Broad, C. D. *Five Types of Ethical Theory*. London: Kegan Paul, Trench, Trubner & Co. Ltd., 1930.
- Broadie, Sarah J. *Ethics with Aristotle*. New York: Oxford University Press, 1991.
- Cohen, G. A. 'Reason, Humanity and the Moral Law.' *The Sources of Normativity*. Ed. Onora O'Neill. New York: Cambridge University Press, 1996. 167-188.
- Darwall, Stephen. *The Second Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press, 2006.
- Descartes, René. *Meditations on First Philosophy*. 1641. Trans. Ed. John Cottingham. Revised ed. Cambridge: Cambridge University Press, 1996.
- Engstrom, Stephen. 'Conditioned Autonomy.' *Philosophy and Phenomenological Research* 48 (1988): 435-353.
- . *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Cambridge: Cambridge University Press, 2009.
- Frierson, Patrick R. *Freedom and Anthropology in Kant's Moral Philosophy*. Cambridge: Cambridge University Press, 2003.
- Guyer, Paul. *Kant on Freedom, Law, and Happiness*. New York: Cambridge University Press, 2000.
- Hegel, G. W. F. *Hegels theologische Jugendschriften*. Ed. Herman Nohl. Tübingen: J. C. B. Mohr, 1907.
- . *Early Theological Writings*. Ed. T. M. Knox. Chicago: University of Chicago Press, 1948.
- Henson, Richard. 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action.' *Philosophical Review* 88 (1979): 39-54.
- Herman, Barbara. *The Practice of Moral Judgement*. Cambridge, MA: Cambridge University Press, 1993.
- . *Moral Literacy*. Cambridge, MA: Harvard University Press, 2007.
- Hill, Thomas E., Jr. 'Kantian Constructivism in Ethics.' *Ethics* 99 (1989): 752-770.

-
- . *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca, NY: Cornell University Press, 1992.
- . 'Legislating the Moral Law and Taking One's Choices to be Good.' *Philosophical Books* 49 (2008): 97-106.
- Hills, Alison. 'Kantian Value Realism.' *Ratio* 21 (2008): 182-200.
- Hobbes, Thomas. *Leviathan*. 1651. Ed. Richard Tuck. Revised student ed. Cambridge: Cambridge University Press, 1996.
- Hume, David. *A Treatise of Human Nature*. 1739-40. Ed. L. A. Selby Bigge. Rev. P. H. Nidditch. Oxford: Clarendon Press, 1978.
- Kain, Patrick. 'Self-Legislation in Kant's Moral Philosophy.' *Archiv für Geschichte der Philosophie* 86 (2004): 257-306.
- Korsgaard, Christine M. 'Kant's Analysis of Obligation: the Argument of Groundwork I.' *Monist* 72 (1989): 311-40.
- . 'Self-Constitution in the Ethics of Plato and Kant.' *The Journal of Ethics* 3 (1999): 1-29.
- Korsgaard, Christine M. et al. *The Sources of Normativity*. Ed. Onora O'Neill. Cambridge: Cambridge University Press, 1996.
- Larmore, Charles. *The Autonomy of Morality*. Cambridge: Cambridge University Press, 2008.
- Leibniz, G. W. *Theodicy*. 1710. Trans. E. M. Hugard. Ed. Austin Farrer. LaSalle: Open Court Publishing Company, 1985.
- Louden, Robert B. *Morality and Moral Theory: A Reappraisal and Reaffirmation*. Oxford: Oxford University Press, 1992.
- . 'Go-Carts of Judgement: Exemplars in Kantian Moral Education.' *Archiv für Geschichte der Philosophie* 74 (1992): 303-22.
- . *Kant's Impure Ethics: From Rational Beings to Human Beings*. New York: Oxford University Press, 2000.
- . 'Making the Law Visible: The Role of Examples in Kant's Ethics.' *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*. Ed. Jens Timmermann. Cambridge: Cambridge University Press, forthcoming.
- Michalson, Gordon E., Jr. *Fallen Freedom*. Cambridge: Cambridge University Press, 1990.
- Morgan, Seiriol. 'The Missing Formal Proof of Humanity's Radical Evil in Kant's "Religion".' *The Philosophical Review* 114 (2005): 63-114.

- Munzel, G. Felicitas. *Kant's Conception of Moral Character: The 'Critical' Link of Morality, Anthropology, and Reflective Judgement*. Chicago: University of Chicago Press, 1999.
- O'Neill, Onora. *Constructions of Reason: Explorations in Kant's Practical Philosophy*. Cambridge: Cambridge University Press, 1989.
- . *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press, 2002.
- . 'Autonomy and the Fact of Reason in the Kritik der praktischen Vernunft.' *Kritik der praktischen Vernunft*. Ed. Otfried Höffe. Berlin: Akademie Verlag, 2002. 81-98.
- . 'Constructivism in Rawls and Kant' *The Cambridge Companion to Rawls*. Ed. Samuel Freedman. Cambridge: Cambridge University Press, 2002. 347-69.
- . 'Constructivism vs. Contractualism.' *Ratio*. 16 (2003): 319-331.
- . 'Autonomy: The Emperor's New Clothes.' *Aristotelian Society Supplementary Volume 77* (2003): 1-21.
- . 'Self-legislation, Autonomy and the Form of Law.' *Right, History, Religion: Kant's Present Significance*. International Symposium, Austrian Academy of Science. Vienna, 5th March, 2004. Printed in *Recht – Geschichte – Religion: Die Bedeutung für die Gegenwart*. Ed. Herta Nagl-Docekal and Rudolf Langthaler. Spec. issue of *Deutsche Zeitschrift für Philosophie* 9 (2004): 13-26.
- Paton, H. J. *The Categorical Imperative*. Philadelphia: University of Pennsylvania Press, 1971.
- . 'Analysis of the Argument.' *The Moral Law: Groundwork of the Metaphysics of Morals*. 1948. London: Routledge, 1991.
- Plato. 'Republic.' *Complete Works*. Trans. G. M. A. Grube. Rev. C. D. C. Reeve. Ed. John M. Cooper. Associate ed. E. S. Hutchinson. Indianapolis, IN: Hackett Publishing Company, Inc. 1997. 971-1223.
- Rawls, John. 'Kantian Constructivism in Moral Theory.' *Journal of Philosophy* 77 (1980): 515-72.
- . *Political Liberalism*. New York: Columbia University Press, 1993.
- . *A Theory of Justice*. 1971. Cambridge MA: Harvard University Press, 1999.
- . *Lectures on the History of Moral Philosophy*. Ed. Barbara Herman. Cambridge, MA: Harvard University Press, 2000.
- Reath, Andrews. 'Self-legislation and Duties to Oneself.' *Kant's Metaphysics of Morals: Interpretative Essays*. Ed. Mark Timmons. Oxford: Oxford University Press, 2002. 349-370.

-
- . 'Value and Law in Kant's Moral Theory.' *Ethics* 114 (2003): 127-155.
- . *Agency and Autonomy in Kant's Moral Theory*. Oxford: Oxford University Press, 2006.
- . 'Autonomy, Taking One's Choices to be Good, and Practical Law: Replies to Critics' *Philosophical Books* 49 (2008): 125-137.
- Reinhold, C. L. 'Erörterung des Begriffs von der Freiheit des Willens.' *Materialien zu Kants Kritik der praktischen Vernunft*. Ed. Rüdiger Bittner and Konrad Cramer. Frankfurt am Main: Suhrkamp, 1975. 252-74.
- Ross, W. D. *The Right and the Good*. Oxford: Clarendon Press, 1930.
- Rousseau, Jean-Jacques. *The Social Contract and Other Later Political Writings*. Trans. Ed. Victor Gourevitch. Cambridge: Cambridge University Press, 1997.
- Schiller, Friedrich. 'Über Anmut und Würde.' 1793. *Schillers sämtliche Werke*. Vol. 11. Ed. Eduard der Hellen. Anmerkungen von Oskar Walzel. Stuttgart: J. G. Cotta'sche Buchhandlung, 1904. 180-245.
- . *Werke*. Vol. I. Ed. Julius Peterson and Friedrich Beißner. Weimar: Hermann Böhlaus Nachfolger, 1943.
- Schneewind, J. B. 'Natural Law, Scepticism, and Methods of Ethics.' *Journal of the History of Ideas*. 52 (1991): 289-308.
- . *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge: Cambridge University Press, 1998.
- Shell, Susan Meld. *Kant and the Limits of Autonomy*. Cambridge, MA: Harvard University Press, 2009.
- Sidgwick, Henry. *The Methods of Ethics*. 1874. 7th ed. 1907. Indianapolis, IN: Hackett Publishing Company, 1981.
- Skorupski, John. 'Rescuing Obligation.' *European Journal of Philosophy*. 6 (1998): 335-355.
- Timmermann, Jens. 'Kantian Duties to the Self: Explained and Defended.' *Philosophy* 81(2006): 505-530.
- . *Kant's Groundwork of the Metaphysics of Morals: A Commentary*. Cambridge: Cambridge University Press, 2007.
- . 'Acting from Duty: Inclination, Reason and Moral Worth.' *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*. Ed. Jens Timmermann. Cambridge: Cambridge University Press, forthcoming.
- Watkins, Eric and William Fitzpatrick. 'O'Neill and Korsgaard on the Construction of Normativity.' *The Journal of Value Inquiry* 36 (2002): 349-67.

- Williams, Bernard A. O. and Thomas Nagel. 'Moral Luck.' *Proceedings of the Aristotelian Society*, Supplementary Volume 50 (1976): 115-135, 137-51.
- Wolff, Robert Paul. *The Autonomy of Reason: A Commentary on Kant's Groundwork of the Metaphysics of Morals*. Gloucester, MA: Peter Smith, 1986.
- Wood, Allen W. *Kant's Moral Religion*. Ithaca, NY: Cornell University Press, 1970.
- . 'Kant's Compatibilism.' *Self and Nature in Kant's Philosophy*. Ed. Allan Wood. Ithaca: Cornell University Press, 1984. 73-101.
- . 'The Emptiness of the Moral Will.' *The Monist* 72 (1989): 457.
- . *Kant's Ethical Thought*. Cambridge: Cambridge University Press, 1999.
- . *Kantian Ethics*. Cambridge: Cambridge University Press, 2008.