CAMBRIDGE www.cambridge.org/9780521896856

This page intentionally left blank

KANT'S *CRITIQUE OF PRACTICAL REASON*

The *Critique of Practical Reason* is the second of Kant's three *Critiques* and his second work in moral theory after the *Groundwork of the Metaphysics of Morals*. Its systematic account of the authority of moral principles grounded in human autonomy unfolds Kant's considered views on morality and provides the keystone to his philosophical system.

These new essays shed light on the principal arguments of the second *Critique* and explore their relation to Kant's critical philosophy as a whole. They examine the genesis of the *Critique*, Kant's approach to the authority of the moral law given as a 'fact of reason', the metaphysics of free agency, the account of respect for morality as the moral motive, and questions raised by the 'primacy of practical reason' and the idea of the 'postulates of pure practical reason'. Engaging and critical, this volume will be invaluable to advanced students and scholars of Kant and to moral theorists alike.

ANDREWS REATH is Professor of Philosophy at the University of California, Riverside. He is the author of *Agency and Autonomy in Kant's Moral Theory* (2006) and the co-editor (with Barbara Herman and Christine M. Korsgaard) of *Reclaiming the History of Ethics: Essays on John Rawls* (Cambridge, 1997).

JENS TIMMERMANN is Senior Lecturer in Moral Philosophy at the University of St Andrews. He is the author of *Kant's 'Groundwork of the Metaphysics of Morals': A Commentary* (Cambridge, 2007) and *Sittengesetz und Freiheit* (2003) and the editor in German, of the *Critique of Pure Reason*, the *Groundwork* and, for the Berlin-Brandenburg Academy, the *Critique of Practical Reason*.

CAMBRIDGE CRITICAL GUIDES

Titles published in this series:

Hegel's *Phenomenology of Spirit*
EDITED BY DEAN MOYAR AND MICHAEL QUANTE

Mill's *On Liberty*
EDITED BY C. L. TEN

Kant's *Idea for a Universal History With a Cosmopolitan Aim*
EDITED BY AMÉLIE OKSENBERG RORTY AND JAMES SCHMIDT

Kant's *Groundwork of the Metaphysics of Morals*
EDITED BY JENS TIMMERMANN

Kant's *Critique of Practical Reason*
EDITED BY ANDREWS REATH AND JENS TIMMERMANN

KANT'S
Critique of Practical Reason
A Critical Guide

EDITED BY

ANDREWS REATH
University of California, Riverside

AND

JENS TIMMERMANN
University of St Andrews

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521896856

First published in print format 2010

ISBN-13 978-0-511-77035-7 eBook (NetLibrary)

ISBN-13 978-0-521-89685-6 Hardback

Contents

Preface

The idea for this project grew out of a conversation between Jens Timmermann, Steve Engstrom and Andrews Reath at a conference on Kant's moral philosophy hosted by Peking University in May 2004. We believe that Andy suggested that there was a need for a volume of new essays on the *Critique of Practical Reason*, and Jens suggested that we assemble a group that would read through the *Critique* together before embarking on the essays. This volume is thus part of a larger project that included a working group that met several times. Jens Timmermann kindly agreed to host meetings of the group at the University of St Andrews, and we met there twice – in July 2006 to read the *Critique* together and in late August–early September 2007 to discuss preliminary versions of the papers. Pauline Kleingeld invited us to meet at Leiden University, and we convened there a third time in June 2008 to discuss penultimate versions of the papers. We received financial support from several institutions along the way. We should like to thank Cambridge University Press, the University of St Andrews, the Paton Fund, the Scots Philosophical Club, the University of California at Riverside, the Philosophy Department of Leiden University and the Leiden University Fund. We should also like to thank Lucy Richmond (St Andrews) for compiling the index.

AR
JT

Contributors

STEFANO BACIN is currently Alexander von Humboldt research fellow at the Kant Research Centre at the University of Mainz (Germany). He has written on German philosophy, and his publications include a monograph on the development of Kant's practical philosophy between the 1760s and *The Metaphysics of Morals* (*Il senso dell'etica. Kant e la costruzione di una teoria morale*, Il Mulino, 2006) and *Fichte a Schulpforta* (Guerini, 2003; German edition Frommann-Holzboog, 2007).

STEPHEN ENGSTROM is Professor of Philosophy at the University of Pittsburgh (USA). He is the author of *The Form of Practical Knowledge* (Harvard University Press, 2009) and editor (with Jennifer Whiting) of *Aristotle, Kant, and the Stoics* (Cambridge University Press, 1996). He has also written several articles on Kant's practical and theoretical philosophy, including the Introduction to the Hackett Publishing edition of the *Critique of Practical Reason* (2002).

PIERRE KELLER is Professor of Philosophy at the University of California, Riverside (USA). He is the author of *Kant and the Demands of Self-Consciousness* (Cambridge University Press, 1998) and *Husserl and Heidegger on Human Experience* (Cambridge University Press, 1999) and of articles on Kant and post-Kantian German philosophy. He is currently working on a book on Kant's practical philosophy.

PAULINE KLEINGELD is Professor of Philosophy at Leiden University (Netherlands). She works on Kant's and Kantian philosophy, moral theory, political philosophy and feminism. She is the author of *Fortschritt und Vernunft: Zur Geschichtsphilosophie Kants* (Königshausen & Neumann, 1995), the editor of Immanuel Kant, *Toward Perpetual Peace and Other Texts on Politics, Peace, and History* (Yale University Press, 2006), and the author of *Kant and Cosmopolitanism: The Philosophical Ideal of World Citizenship* (Cambridge University Press, forthcoming).

HEINER F. KLEMME is Professor of Philosophy and head of the Kant Research Centre at the University of Mainz (Germany). He has written extensively on Kant's philosophy, the Scottish Enlightenment and issues in practical philosophy. His books include *Kants Philosophie des Subjekts* (Meiner, 1996), *Immanuel Kant* (Campus, 2004) and *David Hume* (Junius, 2007). He is co-editor of *Kant-Studien* and also wrote the Introduction to the 2003 Meiner edition of the *Kritik der praktischen Vernunft*.

ANDREWS REATH is Professor of Philosophy at the University of California, Riverside (USA). He has worked extensively on Kant's moral philosophy and recently published a collection of his essays, *Agency and Autonomy in Kant's Moral Theory* (Oxford University Press, 2006). He also wrote the Introduction to the Cambridge University Press edition of the *Critique of Practical Reason* (1997).

JENS TIMMERMANN is Senior Lecturer in Moral Philosophy at the University of St Andrews (UK) and author of *Sittengesetz und Freiheit* (de Gruyter, 2003) and *Kant's 'Groundwork of the Metaphysics of Morals': A Commentary* (Cambridge University Press, 2007). He has edited, in German, the *Critique of Pure Reason*, the *Groundwork* and, for the Berlin-Brandenburg Academy, the *Critique of Practical Reason*. Cambridge University Press is publishing his facing-page edition of Kant's *Groundwork* in 2010.

ERIC WATKINS is Professor of Philosophy at the University of California, San Diego (USA). He has written on Kant's theoretical philosophy, including his pre-critical period and philosophy of science, as well as on Kant's practical philosophy. He is the author of *Kant and the Metaphysics of Causality* (Cambridge University Press, 2005). He has edited *Kant and the Sciences* (Oxford University Press, 2001) and edited and translated *Kant's* Critique of Pure Reason: *Background Source Materials* (Cambridge University Press, 2009).

MARCUS WILLASCHEK is Professor of Philosophy at the University of Frankfurt (Germany). He has written widely on Kant's theoretical and practical philosophy, including several articles on the *Critique of Practical Reason*. His books include *Praktische Vernunft. Handlungstheorie und Moralbegründung bei Kant* (Metzler, 1992) and *Der mentale Zugang zur Welt: Realismus, Skeptizismus und Intentionalität* (Klostermann, 2003). He is editor (with Georg Mohr and Jürgen Stolzenberg) of the *Kant-Lexikon* (de Gruyter, forthcoming).

Translations and abbreviations

Citations of Kant's works refer to the volume and page number in the Academy Edition of Immanuel Kant, *Gesammelte Schriften* (Berlin: Walter de Gruyter and predecessors, 1900–). In some cases, the citation includes the line number in the Academy Edition. References to the *Critique of Pure Reason*, which cite the page numbers of the A and B editions, are the exception. Quotations use or have been adapted from the following translations:

Anth *Anthropology from a Pragmatic Point of View*, in Immanuel Kant, *Anthropology, History and Education*, ed. Robert B. Louden and Günter Zöller (Cambridge and New York: Cambridge University Press, 2006).

CpV *Critique of Practical Reason*, trans. Mary J. Gregor, in Immanuel Kant, *Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1996).

CrV *Critique of Pure Reason*, trans. Paul Guyer and Allen W. Wood (Cambridge and New York: Cambridge University Press, 1998).

CU *Critique of Judgment*, trans. Paul Guyer and Eric Matthews (Cambridge and New York: Cambridge University Press, 2000).

G *Groundwork of the Metaphysics of Morals*, trans. Mary J. Gregor, in Immanuel Kant, *Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1996).

Logic *The Jäsche Logic* ('Immanuel Kant's Logic', ed. Gottlob Benjamin Jäsche), in Immanuel Kant, *Lectures on Logic*, trans. and ed. Michael J. Young (Cambridge and New York: Cambridge University Press, 1992).

MdS *The Metaphysics of Morals*, trans. Mary J. Gregor, in Immanuel Kant, *Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1996).

P	*Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science*, trans. Gary Hatfield, in Immanuel Kant, *Theoretical Philosophy after 1781*, ed. Henry Allison and Peter Heath (Cambridge and New York: Cambridge University Press, 2002).
R	*Religion within the Boundaries of Mere Reason*, trans. George di Giovanni, in Immanuel Kant, *Religion and Rational Theology*, ed. Allen W. Wood and George di Giovanni (Cambridge and New York: Cambridge University Press, 1996).
TP	*On the Common Saying: That May Be Correct in Theory, But It Is of No Use in Practice*, trans. Mary J. Gregor, in Immanuel Kant, *Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1996).
ZeF	*Toward Perpetual Peace*, trans. Mary J. Gregor, in Immanuel Kant, *Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1996).

Citations have also been made to the following works of Kant: *Eine Vorlesung über Ethik*, ed. Gerd Gerhardt (Frankfurt am Main: Fischer Taschenbuch Verlag, 1990). *Lectures on Ethics*, ed. Peter Heath and J. B. Schneewind (Cambridge and New York: Cambridge University Press, 1997). *Lectures on Metaphysics*, ed. Karl Ameriks and Steve Naragon (Cambridge and New York: Cambridge University Press, 1997). *Notes and Fragments*, trans. Paul Guyer, Curtis Bowman and Frederick Rauscher (Cambridge and New York: Cambridge University Press, 2005). *Religion and Rational Theology*, ed. Allen W. Wood and George di Giovanni (Cambridge and New York: Cambridge University Press, 1996). *Theoretical Philosophy after 1781*, ed. Henry Allison and Peter Heath (Cambridge and New York: Cambridge University Press, 2002). *Theoretical Philosophy, 1755–1770*, trans. and ed. David Walford and Ralf Meerbote (Cambridge and New York: Cambridge University Press, 1992) *Vorlesung zur Moralphilosophie*, ed. W. Stark (Berlin and New York: de Gruyter, 2004).

Introduction

Andrews Reath

The *Critique of Practical Reason* is Kant's second foundational work in moral theory after the *Groundwork of the Metaphysics of Morals*. Its stated aim is 'merely to show that there is *pure practical reason*' (*CpV* 5:3). That is, it attempts to show that reason by itself yields an objective principle of conduct that applies independently of individuals' preferences and empirically given aims – a principle that Kant identifies with the fundamental principle of morality – and thus to show that practical reasoning is not limited to instrumental and prudential reasoning as the empiricist tradition holds. To accomplish this aim, Kant tries to document the origin of the fundamental principle that underwrites common moral thought in reason (in 'pure practical reason') and to establish its overriding authority.

A 'critique' is a critical examination of a cognitive faculty that sets out its powers and limits, and in particular establishes the legitimacy of any a priori concepts and principles that structure the relevant domain of cognitive activity. Kant's views about the need for and proper focus of a critique of practical reason changed over time. When he published the *Groundwork* in 1785, he intended it to take the place of a critique of pure practical reason. He writes that although 'there is really no other foundation for a metaphysics of morals than a critique of *pure practical reason*, just as that of metaphysics is the critique of pure speculative reason', the need for the former is less urgent. That is 'because in moral matters human reason can be brought to a high degree of correctness and accomplishment', while the theoretical use of reason tends to overstep its limits and to make illusory metaphysical claims. Further, a full-blown critique of practical reason would introduce complexities that are not strictly necessary to present and to ground the authority of the basic principle of morality (*G* 4:391). The third section of the *Groundwork* is entitled 'Transition from Metaphysics of Morals to the Critique of Pure Practical Reason', and initially Kant thought that the arguments in this section were sufficient title for the pure practical or moral use of reason.

After publishing the *Groundwork*, however, while Kant was preparing a revised edition of the *Critique of Pure Reason*, he decided to add a 'Critique of Pure Practical Reason' as an appendix to that work, in order to respond to various objections to the *Groundwork* and to complete his critical system. But this appendix was not part of the second edition of the *Critique of Pure Reason* when it appeared in 1787. It was published instead as a separate work in 1788 – only now with the title *Critique of Practical Reason*. Kant continued to think that while the existence of pure practical reason needs some vindication, it does not need critical limitation. Rather, once the existence of pure practical reason is established, only the 'empirically conditioned' use of practical reason needs a critique, in order to limit its presumption of supplying the only grounds of choice (*CpV* 5:16).

The organizational structure that Kant imposes on the *Critique of Practical Reason* is similar to those of the *Critique of Pure Reason* and the *Critique of Judgment*. The second *Critique* is divided into a 'Doctrine of Elements' that takes up the bulk of the work and a much shorter 'Doctrine of Method', and the Doctrine of Elements in turn has both an 'Analytic' and a 'Dialectic' of pure practical reason. The main concern of the opening chapter of the Analytic is to establish the authority of the moral law as the fundamental principle of pure practical reason. Chapter II addresses questions about Kant's concept of the good, and it defends his 'method' in moral theory of beginning with the concept of law rather than the concept of the good. Only by establishing that there is a practical law can one show that the necessity that is part of the common concept of duty is genuine. Chapter III is a detailed discussion of respect for the moral law as the moral motive, a topic that Kant had addressed only briefly in a footnote in the *Groundwork* (*G* 4:401fn.). This chapter explores the phenomenology of moral motivation and explains how the principle of morality functions as a motive, with an eye to substantiating the claim of Chapter I that pure reason is practical – that reason by itself yields practical requirements that can move the will. The Analytic concludes with a 'Critical Elucidation' that, among other things, explains its overall structure and sets out Kant's two-standpoint approach to the problem of free will.

The Dialectic introduces the idea of the highest good as the necessary final aim of moral conduct and argues that our interest in the highest good warrants assuming the existence of God and the immortality of the soul as 'postulates of pure practical reason'. The possibility of the highest good appears to generate an antinomy within practical reason that is resolved through these postulates. Without such beliefs we cannot conceive of the possibility of the highest good, and thus cannot rationally sustain the

commitment to its pursuit. In this way the second *Critique* contributes to and fills a lacuna in Kant's critical system, giving 'objective though only practical reality' to metaphysical objects that the *Critique of Pure Reason* had shown to be beyond the limits of human knowledge. The idea of a practical postulate is of interest as a form of rational belief whose warrant is not based on evidence, but rather on a set of practical interests that we are not free to abandon.[1]

The second *Critique* assumes the basic account of the fundamental principle of morality – the Categorical Imperative – given in the *Groundwork*, and many of the central ideas and theses of the *Groundwork* reappear in the Analytic of the second *Critique*, especially its first chapter. Both works assume that it is part of common moral thought that moral requirements carry unconditional authority, and their central arguments begin from an analysis of the concept of an unconditional requirement on conduct. To cite a few points of overlap, just as the *Groundwork* derives a statement of the Categorical Imperative from the concept of a categorical imperative (*G* 4:402–3, 420–1), the statement of the 'fundamental law of pure practical reason' in the second *Critique* (*CpV* 5:30) results from an analysis of the concept of a practical law (*CpV* 5:19). Both works argue that the fundamental principle of morality must, in some sense, be a 'formal' principle – a principle whose normative force depends on its form, and not on any end or purpose that constitutes the 'matter' of the principle (see *G* 4:400, 414, 415, 416 and *CpV* 5:27, 39–41). Moral theories that base their fundamental principle on an object given to the will – whether it be an object of the senses or one thought through reason – lead to 'heteronomy' and are unable to ground true categorical imperatives or practical laws (see *G* 4:441–4 and *CpV* 5:21–2, 39–41). Thus both works argue for the analytic claim that only a principle of autonomy – a principle based in the nature of rational volition or one that the will in some sense gives to itself – can ground the necessity that is part of the common concept of duty (see *G* 4:432–3, 440, 444 and *CpV* 5:33). Finally, both works argue for analytic connections between freedom and morality – that a will with the capacity to act from the formal principle of morality is free, and that the principle of

[1] For more complete overviews of the main themes of the second *Critique*, see the introductory essays by Stephen Engstrom, in Kant, *Critique of Practical Reason*, trans. Werner Pluhar (Indianapolis: Hackett Publishing Company, 2002), xv–liv; by Andrews Reath, in Kant, *Critique of Practical Reason*, trans. Mary J. Gregor (Cambridge and New York: Cambridge University Press, 1997), vii–xxxi; and by Heiner Klemme, in Kant, *Kritik der praktischen Vernunft*, ed. Horst D. Brandt and Heiner Klemme (Hamburg: Felix Meiner Verlag, 2003), ix–lxii.

morality is the fundamental principle that governs a free will (see *G* 4:446–7 and *CpV* 5:28–30).

There are also important points on which the works appear to differ, though how deep these differences run is a matter of controversy. In the third section of the *Groundwork* Kant attempts to establish the authority of morality through a 'deduction', and as part of the overall argument he appears to claim that we may ascribe a robust form of free agency to ourselves simply on the basis of general features of rationality, including theoretical rationality (*G* 4:448, 451–3). However, in the second *Critique*, Kant claims that a deduction of the moral law is neither possible nor necessary and that the authority of the moral law is instead given as a 'fact of reason' (*CpV* 5:31, 42, 46–7). Furthermore, he claims that only moral consciousness gives us grounds for ascribing free agency to ourselves; rather than seeking a deduction of the moral law, Kant now holds that the moral law is the basis of a deduction of the capacity of free agency (*CpV* 5:30, 47–8).

The composition of the second *Critique*, its main lines of argument and their relation to those of the *Groundwork* raise several questions. What led Kant to write a self-standing 'critique of practical reason'? What does Kant mean by a 'formal principle', and why does he think that the fundamental principle of morality must be a formal principle, or principle whose normative force comes from its form rather than its matter? What exactly is the 'fact of reason' and does it provide a satisfactory account of the authority of morality? How far-reaching are the differences between Kant's approaches to the authority of the moral law in the *Groundwork* and the second *Critique*? If he abandons the idea of a deduction of the moral law, where does he think that the earlier argument falls short? What is Kant's stand on the metaphysics of free agency in the second *Critique*, and how is it related to his views in other works, such as the first *Critique* and the *Groundwork*? The essays in this volume take up these and other questions that are central to understanding the aims of the *Critique of Practical Reason*, its central doctrines and contribution to moral theory and its role within Kant's critical system.

Heiner Klemme's essay, 'The origin and aim of Kant's *Critique of Practical Reason*', addresses the question why Kant decided to write a self-standing 'critique' of practical reason by turning to developments in Kant's critical philosophy in 1786–7. Klemme identifies two reasons for this change. First, in the spring of 1787, Kant came to believe that judgments of taste had an a priori character, and he planned a 'Groundwork of the Critique of Taste' that later became the *Critique of Judgment*. This

development led Kant to modify his overall critical plan and to divide the project of a single 'critique of pure reason' into three separate critiques related to the faculties of understanding, practical reason and judgment. The *one* critique of pure reason is now presented in three works dedicated to (a) the constitutive use of the understanding and the speculative use of pure reason, (b) the constitutive use of pure practical reason and (c) the regulative use of the power of judgment in determining the feeling of pleasure and displeasure. Second, Klemme looks for the genesis of the *Critique of Practical Reason* not just in its Analytic but in its Dialectic. He argues that Kant's discovery of an 'antinomy of practical reason' is a new development with no model in Kant's previous work, and that even with the 'fact of reason', the validity of the categorical imperative cannot be completely secure until this antinomy is resolved – requiring a separate 'critique' of practical reason.

Andrews Reath asks how to understand Kant's notion of a 'formal principle' in his essay, 'Formal principles and the form of a law'. Early in Chapter I of the Analytic, Kant claims that a practical law must be a formal principle – that is, a practical principle that provides a ground of choice through its form, rather than its matter. Further he suggests that fundamental normative principles must be formal principles. Kant explicitly argues that practical principles that provide a ground of choice through their 'matter' cannot provide laws or apply with true normative necessity, but he does not clearly explain why formal principles (in his sense), and only formal principles, do apply with necessity. Reath argues that Kant's conception of a formal principle is best understood not simply as a principle that involves some abstraction from content, but as a principle that is constitutive of some domain of cognitive activity – a principle that defines and makes possible and tacitly guides all instances of that kind of cognitive activity. So the formal principle of the rational or pure will would be its internal constitutive principle. Reath argues that understanding formal principles as constitutive principles establishes the connection that Kant sees between form, or formal principles, and normative necessity: a constitutive principle necessarily governs the relevant domain of cognitive activity and is not coherently rejected by anyone engaged in that activity. Reath develops this notion of formal principle, then considers how it figures in various central Kantian arguments – for example, that 'material principles of morality' cannot ground genuine laws and that a will governed by the formal principle of volition is a free will.

In 'Moral consciousness and the "fact of reason"', Pauline Kleingeld sheds new light on the significance of Kant's claim that moral consciousness

can be called a 'fact of reason' (*Factum der Vernunft*). Many commentators take this claim to indicate that Kant gives up his earlier aspiration of justifying the principle of morality, and instead adopts an unarguable point of departure. This reading often leads to the complaint that Kant now leaves his moral theory without a proper foundation, and that even if moral consciousness is universal, morality might still be an illusion. Kleingeld argues that the hermeneutical key to a better interpretation lies in the proper understanding of the meaning of '*Factum*'. Current interpretations of '*Factum*' as 'fact' fail to consider the etymological background. In Kant's times, the first meaning of the term was 'deed', and its second meaning was 'fact', understood as the result of activity. Against recent proposals to read it as 'deed' or as a technical term indicating a moment in Kant's proof structure, she argues that '*Factum*', as used by Kant, is best understood as a fact that is the result of an activity. Furthermore, Kleingeld argues that the consciousness at issue in the fact of reason should be understood fundamentally as the consciousness of a *rational* principle, namely, of the law of pure practical reason. The argument of the first chapter of the second *Critique* proceeds almost entirely in terms of practical reason and the fundamental practical law, not morality. Of course Kant identifies this fundamental law with the moral law, but highlighting his focus on a rational principle helps avoid the misunderstanding that he is simply presupposing a particular conception of morality. This reading of Kant's claims about the fact of reason makes it possible to answer the main criticisms voiced in the literature.

Jens Timmermann's chapter, 'Reversal or retreat? Kant's deductions of freedom and morality', concerns the differences between the *Critique of Practical Reason* and the *Groundwork of the Metaphysics of Morals*, published three years earlier – in particular their differences on the question whether the principle of morality stands in need of philosophical justification, as do the categories in Kant's theoretical philosophy. The third section of the *Groundwork* purports to contain such a 'deduction' and thus the core of a 'critique of pure practical reason'. In the second *Critique*, Kant explicitly rejects his previous conception of a 'critique of pure practical reason', comes close to admitting that his search for a 'deduction of the moral principle' was unsuccessful and deems explanation of the possibility of moral commands both impossible and unnecessary. Criticizing the practical faculty as a whole now reveals that the moral law is given as a 'fact of reason' and leads to a 'deduction of freedom' on moral grounds. Timmermann argues that the *Critique of Practical Reason* marks not just a strategic reversal, but also a retreat. The reason is that although many of the familiar doctrines and

arguments of *Groundwork* III reappear in the later work – the reciprocity thesis, the notion of 'transferral into an intelligible order', the idea that we conceive of ourselves as members of a normative realm – Kant is no longer willing to employ these elements in a formal deduction of the categorical imperative. A deduction would have to rely on the intuition of ourselves as members of a realm of autonomy, and according to Kant's own epistemic restrictions, no such intuition is possible. Thus Kant now realizes that the justification of morality cannot be assimilated to the transcendental deduction of the concepts of the pure understanding. Furthermore, the second *Critique* represents a retreat because what in the *Groundwork* is merely meant to *confirm* the correctness of the moral deduction now takes its place: since ordinary moral consciousness establishes that morality is real, the question of its possibility can safely be disregarded.

In 'The *Triebfeder* of pure practical reason', Stephen Engstrom explores Kant's account, in Chapter III of the Analytic, of how pure reason moves us to act through the moral law that it legislates, or how it is practical in human beings, in whom choice is subject to sensible impulses that may conflict with that law. Since the practicality at issue belongs to reason, it cannot depend on any special feature of our sensible nature that distinguishes humans from other sensible but rational beings. So the challenge is to show how the moral law can exercise an effect on the capacity to feel, and thereby become a spring of action, but without assuming any special capacity of moral feeling. In response, Kant describes certain natural attitudes that we have towards ourselves, claiming that our sensible nature includes a propensity of self-love, which also involves a propensity to self-conceit – an attitude of esteeming oneself in comparison with others. Since these propensities to love and to esteem oneself belong to us as sensible but rational beings, they take the form of tendencies to advance certain claims in which we deem ourselves worthy of the love and the esteem of others. Each person, however, being implicitly conscious of the moral law as the standard of validity for all such claims, recognizes that although love directed to oneself can be valid when broadened to include others, self-conceit, being essentially exclusive on account of its comparative nature, is inherently invalid and therefore to be wholly rejected. This recognition is in the first instance humiliating – a negative feeling expressing the passive side of the moral law's striking down self-conceit. But complementary to this humiliation is a feeling of respect for the moral law, a feeling that is positive in so far as the law is recognized as integral to one's constitution as a rational being. Observing that the diminution of self-conceit through this humiliation constitutes a furtherance of the moral law's efficacy, Kant points to

this distinctive feeling of respect as the moral law's effect on the capacity to feel, or pure reason's practicality as it operates in rational beings with a sensible nature.

Pierre Keller's contribution, 'Two conceptions of compatibilism in the Critical Elucidation', takes on the conception of free will in the second *Critique*. Kant rejects a standard form of compatibilism, according to which psychological freedom is compatible with causal determinism, because that conception of freedom is insufficient to ground moral responsibility. Instead he develops a more complex form of compatibilism aimed at reconciling a stronger 'transcendental' (incompatibilist) conception of freedom with causal determinism through a two-standpoint approach to action. Keller's essay explores the contours of Kant's more complex form of compatibilism. Keller discusses Kant's view that the ideal of complete explanation generated by theoretical reason requires the real possibility of uncaused causes, but that attempts by theoretical reason to make sense of this idea lead to antinomy. Only practical reason can give content to this idea, through the commitments of ordinary moral thought. Common-sense morality is committed to the existence of categorical obligations that apply unconditionally, and thus to the possibility of being motivated solely by the intrinsic reasonableness of an action, independently of antecedent considerations tied to one's spatio-temporal position. In this way moral agency supports a notion of 'absolute spontaneity' or transcendental freedom. Keller explores the limits that Kant sees to empirical or psychological accounts of free agency. He then shows how the two-standpoint approach that emerges from Kant's transcendental idealism attempts to reconcile a more robust notion of free agency with causal determinism. If the notion of a complete set of inquiry-independent causal conditions is a necessary illusion of inquiry, then the claim that unconditioned agency and causal determinism stand in antinomial relation to each other is itself illusory. Since the transcendental idealist views the idea of complete causal explanation as a regulative ideal that can never be fully carried out, empirical causal explanations are never complete. That leaves us free in principle to regard our actions in terms of normative reasons, that is, to adopt a different standpoint towards our action from that of deterministic causal explanation. Furthermore, the commitments of common-sense morality to normative reasons based on unconditional obligations require that we take ourselves to be able to act on such reasons. In this manner, common-sense morality's commitment to the moral law gives both warrant and content to unconditional reasons as causes of action.

In his discussion of the Antinomy of Practical Reason, 'The Antinomy of Practical Reason: reason, the unconditioned and the highest good', Eric Watkins draws on certain features of the Antinomy of Pure Reason in the *Critique of Pure Reason* to reconstruct Kant's argument and articulate its resolution. He then poses a number of basic questions about the concept of the highest good and develops detailed answers. The questions are: must there must be an object of pure practical reason at all?; must there be a *single* object of pure practical reason?; must the object of pure practical reason be the *highest good*?; could the highest good not be simply the supreme good rather than the complete good?; why must virtue and happiness be related by means of a one-way causal relation in the complete good? The resource that Watkins finds most helpful in addressing these questions is Kant's conception of reason, according to which reason searches for the totality of conditions, and thus the unconditioned, for any conditioned object.

Marcus Willaschek's essay is on 'The primacy of practical reason and the idea of a practical postulate'. In several places in Kant's work, most prominently in the Dialectic of the second *Critique*, Kant denies the unrestricted validity of the principle that rational belief requires evidence in favour of its truth. Rather, we can have rational warrant for a belief even in the complete absence of evidence for it, subject to two conditions. First, the belief must be 'theoretically undecidable': there can be no possible empirical evidence nor a conclusive a priori argument for or against the belief in question. Second, the belief must be 'practically necessary': someone who acknowledges the moral law as binding must, by a kind of subjective but still rational necessity, hold the belief in question. Kant calls a theoretical proposition that is both theoretically undecidable and practically necessary a 'postulate of pure practical reason', and he argues that there are exactly three such postulates – the existence of God, our own transcendental freedom and our immortality. With respect to these postulates, Kant holds that it is rational for us to *believe* in their truth even though they lie beyond the reach of human knowledge. In his essay, Willaschek concentrates on the general idea of a postulate of pure practical reason as a form of belief that is rationally held, though not based on evidence. He begins by laying out the special logical structure of the argument for the possibility of a postulate that Kant gives in the section 'On the Primacy of Practical Reason', then critically discusses and ultimately defends that argument. He considers the general idea of a postulate of pure practical reason and the epistemological status that Kant assigns to it. Finally, he suggests that the main idea behind Kant's argument does not depend on Kant's own, very demanding conception of morality.

Stefano Bacin's essay, 'The meaning of the *Critique of Practical Reason* for moral beings: the Doctrine of Method of Pure Practical Reason', addresses the philosophical role of the Doctrine of Method, the brief Part II of the second *Critique*. In order to underscore the functional connection of this part with the whole, Bacin begins by discussing the general meaning of 'a doctrine of method' in Kant's work, as well as the specific goals of the Doctrine of Method of the second *Critique*. The central section of the chapter focuses on the notion of 'receptivity to morality', which here has a central role and a quite distinct meaning. Bacin argues that Kant's main point in his account of how to 'make objective practical reason *subjectively* practical' (*CpV* 5:151) is that one ought to lead the individual agent to become aware of his own humanity, or fundamental dignity as a moral being, through an understanding of the basic concepts of the Doctrine of Elements. Awareness of one's humanity is the proper basis for conscious moral life. In Kant's view, recognition of this point is relevant to the overall aim of the second *Critique* – to show that pure reason is practical – and of moral theory itself. Kant believes that he has supplied a theory of moral agency that for the first time allows agents to understand their status as moral beings. The task of the Doctrine of Method is to show how it is possible to make agents aware of their basic moral capacities, and through that awareness to instil genuine moral dispositions. Accordingly, the Doctrine of Method is the completion of the *Critique*, confirming the conclusions of the Analytic through the common use of pure practical reason and connecting them with the experience of every moral agent. Bacin closes by discussing how Kant proposes to accomplish this task through exercises and examples intended to make agents aware of different features of their receptivity to morality. The 'science', or moral theory, developed in the first part of the *Critique* is thus connected with its final goal of a 'doctrine of wisdom'.

CHAPTER 1

The origin and aim of Kant's Critique of Practical Reason

Heiner F. Klemme

Why did Kant feel the need to write a *Critique of Practical Reason* after he had published the *Groundwork of the Metaphysics of Morals*? Is it because he acknowledged, with the 'fact of pure practical reason', that his attempt to deduce the moral law in the *Groundwork* had failed? Or did he have other reasons to write a second *Critique* before publishing his *Metaphysics of Morals*? Obviously enough, Kant seeks to establish in the *Critique of Practical Reason* 'that there is pure practical reason'.[1] Kant does so with his critique of the 'entire practical faculty' that underlies the merely pure practical and the empirical practical application of reason. These two forms of the application of practical reason (roughly) correspond to the distinction between the 'analytic of pure practical reason' and the 'dialectic of pure practical reason'.

In the Analytic, Kant shows 'that pure reason can be practical – that is, can of itself, independently of anything empirical, determine the will – and it does so by a fact in which pure reason in us proves itself actually practical, namely autonomy in the principle of morality by which reason determines the will to deeds' (*CpV* 5:42). In the Dialectic in contrast, Kant maintains that pure reason in its practical application gets into conflict with itself while determining the highest good, which runs counter to the facticity of pure practical reason and is in danger of undercutting the validity of the moral law: 'If, therefore, the highest good is impossible in accordance with practical rules, then the moral law, which commands us to promote it, must be fantastic and directed to empty imaginary ends and must therefore in itself be false' (*CpV* 5:114). If there were an antinomy in the concept of a highest good, the doctrine of a 'fact of pure reason' could not be true. In

[1] Earlier versions of this text were presented at conferences in St Andrews, Leiden, Pisa and São Paulo. For further suggestions and criticism, I am especially thankful to Stefano Bacin, Reinhard Brandt, Bernd Ludwig, Andrews Reath and Jens Timmermann. I also thank Falk Wunderlich for the English translation.

order to secure the validity of the moral law, Kant thus has to show *why* pure reason gets into dialectical troubles, and *how* this dialectic can be resolved. The key to understanding this part of the argument is empirical practical reason, to which the empiricist philosophers refer as the chief witness in their sceptical attacks against the morality of the categorical imperative. According to Kant, empirical practical reason threatens the reality of the moral law, and at the same time, criticizing empirical practical reason secures its reality.

If the 'fact of pure practical reason' can be regarded as secure only when the 'antinomy of pure practical reason' has been formulated and also resolved, clearly the key to understanding the origin and aim of the second *Critique* lies in the Dialectic, and not (just) in the Analytic. With the doctrine of the 'fact of pure practical reason', Kant may or may not have abandoned the deductive strategy of the *Groundwork*; but without taking the 'antinomy of pure practical reason' into account, we will not understand why Kant composed the second *Critique* in the way he did. There is no model for this antinomy in Kant's previous writings.

But what kind of considerations might have lead Kant to the 'discovery' (or construction[2]) of an 'antinomy of pure practical reason'? Why does he publish the *Critique of Practical Reason* – and not just an appendix to the second edition of the *Critique of Pure Reason* entitled 'Critique of Pure Practical Reason', as he intended to do in late 1786? These questions are answered best when we turn to the history of the development of Kant's philosophy and look for the motives that guided his reasoning between November 1786 and April 1787.

1. CRITICAL MORAL PHILOSOPHY AND ITS CRITICS

In spring 1786, Kant prepares a second edition of the *Critique of Pure Reason* at his publisher's suggestion, where he wishes to react to the objections raised against the first edition from 1781. According to Kant, all these objections rest upon a 'mistaken interpretation'[3] of his main work. In the advertisement for the new edition, published on 21 November 1786 in the

[2] For more details on the 'antinomy of practical reason' that turns out not in fact to exist, see the instructive study by Bernhard Milz, *Der gesuchte Widerstreit. Die Antinomie in Kants 'Kritik der praktischen Vernunft'* (Berlin and New York: de Gruyter, 2002).

[3] See Kant's letter to L. H. Jakob of 26 May 1786: 'I am just now occupied with a second edition of the *Critique*, at the request of my publisher, and with it I shall clarify certain parts of the work whose misunderstanding has occasioned all the objections so far brought against it … their strength will subside of itself once the pretext of their mistaken interpretation is removed' (10:450 f.).

Allgemeine Literatur-Zeitung (Jena), he also refers to a 'Critique of Pure Practical Reason' extending the *Critique of Pure Reason*, where he intends to rebut 'objections', especially against the 'principle of morality', and to conclude 'the whole of the critical investigations':

> there will also be a *Critique of Pure Practical Reason* as an addition to the *Critique of Pure Speculative Reason* in the first edition, which will, on the one hand, secure the principle of morality against the objections that have been raised or will be raised in the future, and on the other hand, is capable of accomplishing the whole of the critical investigations that precede the system of the philosophy of pure reason.[4]

However, Kant completed the second edition of the *Critique of Pure Reason* in April 1787 without the 'Critique of Pure Practical Reason' that he had announced. In the meantime, he has changed his plans: he abandons the 'Critique of Pure Practical Reason' as an appendix to the '*Critique of Pure Speculative Reason*' (as he dubs the *Critique of Pure Reason* in 1786), and instead, he prepares for publication a monograph with the title *Critique of Practical Reason*.[5] As early as 25 June 1787, Kant reports the imminent completion of his work on the new opus:

> I am so far along with my *Critique of Practical Reason* that I intend to send it to Halle for printing next week. This work will better demonstrate and make comprehensible the possibility of supplementing, by pure practical reason, that which I denied to speculative reason – better than all the controversies with Feder and Abel[6] (of whom the first maintains that there is no a priori cognition at all while the other maintains that there is some sort of cognition halfway between the empirical and the a priori). For this is really the stumbling block that made these men prefer to take the most impossible, yes, absurd path, in order to extend the speculative faculty to the supersensible, rather than submit to what they felt to be the wholly desolate verdict of the *Critique*.[7]

On 11 September, Kant writes to Ludwig Heinrich Jakob that the *Critique of Practical Reason* 'contains many things that will serve to correct

[4] Quoted from Albert Landau (ed.), *Rezensionen zur Kantischen Philosophie*, vol. I (Bebra: Landau, 1990), 471–2. Kant had already mentioned this endeavour in a letter to Friedrich Gottlob Born of 24 September 1786 that has been lost. Born replies to Kant as follows: 'I am looking forward immensely, by the by, to the important addition of a critique of practical reason, with which you will embellish your splendid work [i.e. the *CrV*] still more' (10:470–2, here 471).

[5] Against this, Beck argues with regard to *CrV* Bxliii that Kant had conceived such a plan 'only after April 1787' (Lewis White Beck, *Kants 'Kritik der praktischen Vernunft'. Ein Kommentar* (Munich: W. Fink, 1974) 26; see also 24, 27–8; similarly Giovanni B. Sala, *Kants 'Kritik der praktischen Vernunft'. Ein Kommentar* (Darmstadt: Wissenschaftliche Buchgesellschaft, 2004), 53).

[6] See Jacob Friedrich Abel, *Versuch über die Natur der speculativen Vernunft zur Prüfung des Kantischen Systems* (Frankfurt and Leipzig, 1787).

[7] 10:490.

the misunderstandings of the [Critique of] theoretical [reason]. I shall now turn at once to the *Critique of Taste*, with which I shall have finished my critical work, so that I can proceed to the dogmatic part. I think it will appear before Easter' (10:494). In his letter to Carl Leonhard Reinhold, dated 28 and 31 December 1787, Kant once more expresses the aim of the *Critique of Practical Reason*: 'This little book will sufficiently resolve the many contradictions that the followers of the old-guard philosophy imagine they see in my *Critique* [*of Pure Reason*], and at the same time the contradictions in which they themselves are unavoidably caught up if they refuse to abandon their botched job are made perspicuous.'[8]

The *Critique of Practical Reason* appears (with the year 1788 on the title page) in December 1787.[9] In the preface, Kant comments on the title of the work:

> Why this *Critique* is not entitled a *Critique of Pure Practical Reason* but simply a *Critique of Practical Reason* generally, although its parallelism with the speculative seems to require the first, is sufficiently explained in this treatise. It has merely to show *that there is pure practical reason*, and for this purpose it criticizes reason's entire *practical faculty*. If it succeeds in this it has no need to criticize the *pure faculty itself* in order to see whether reason is merely making a claim in which it presumptuously *oversteps* itself (as does happen with speculative reason).[10]

The aim of the *Critique of Practical Reason*, thus, is inverse compared with the aim of the *Critique of Pure Reason*: Whereas the first *Critique* is supposed to show that we cannot apply pure reason constitutively in a speculative sense, the second *Critique* is supposed to substantiate the view that objections against the practical constitutive application of pure reason put forward by the empiricists and sceptics are vacuous. Although we cannot understand the possibility of a pure application of practical reason a priori, this application must be possible, since the moral law is given to us in consciousness as 'the sole fact of pure reason' (*CpV* 5:31).

The objections against the possibility and the reality of the pure practical use of reason dissolve to nothing once we clearly distinguish between the

[8] 10:513–16, here: 514.

[9] See Kant's letter to Marcus Herz of 24 December 1787 (10:512) and Friedrich August Grunert's letter to Kant of December 1787 (10:506).

[10] *CpV* 5:3. With pure practical reason, 'transcendental *freedom* is also established, taken indeed in that absolute sense in which speculative reason needed it, in its use of the concept of causality, in order to rescue itself from the antinomy into which it unavoidably falls when it wants to think the *unconditioned* in the series of causal connection' (*CpV* 5:3). For the significance of the *Critique of Pure Reason* for the positive concept of practical freedom in the *Critique of Practical Reason* see *CrV* Bxxv and *CrV* Bxxviii–xxix.

pure practical and the empirical practical application of (pure) reason. This point is stressed by Kant in one of his reflections:

> The critique of practical reason has as its basis the differentiation of empirically conditioned practical reason from the pure and yet practical reason and asks whether there is such a thing as the latter. The critique cannot have insight into this possibility *a priori* because it concerns the relation of a real ground to a consequence, thus something must be given which can arise from it alone; and from reality possibility can be inferred. The moral laws are of this sort, and this must be proven in the same way we proved the representations of space and time as *a priori* representations, only with the difference that the latter concern intuitions but the former mere concepts of reason. The only difference is that in theoretical knowledge the concepts have no meaning and the principles no use except with regard to objects of experience, whereas in the practical, by contrast, they have much wider use, namely they apply to all rational beings in general and are independent of all empirical determining grounds, and that indeed, even if no object of experience corresponds to them, the mere manner of thinking and the disposition in accordance with principles already suffice.[11]

Because the second *Critique* is intended to rebut the attacks against the practical application of pure reason, it cannot be a 'Critique of Pure Practical Reason'. The problem is not pure practical reason, but empirical practical reason, and thereby those philosophers who intend to undermine the absolute validity of the moral law with reference to it.

In the preface to the *Critique of Practical Reason*, Kant comments on a few objections that have been raised against his critical philosophy in general, as well as against his foundation of moral philosophy in particular, since 1781. The most important of them pertain to (a) the relationship of freedom and the moral law, (b) the possibility of expanding our knowledge by the practical application of reason beyond the boundaries of experience and (c) the relationship of moral law and highest good.

(a) Johann Friedrich Flatt,[12] in his review of the *Groundwork*, had objected that it is inconsistent to 'call freedom the condition of the moral law' on

[11] Reflection 7201, 19:275–6 (in *Notes and Fragments*, trans. Paul Guyer, Curtins Bowman and Frederick Rauscher (Cambridge and New York: Cambridge University Press, 2005), 464).

[12] 'Were this writing in front of us a few centuries older: we would hardly be able to resist the temptation to doubt the authenticity of one of the passages mentioned. But since this doubt is completely unnecessary in the present case, what remains is only to deem the inconsistencies and paralogisms we encountered *antinomies* of *Kant's* reason.' *Tübingische gelehrte Anzeigen*, fourteenth piece, 16 February 1786, 105–12, here 108–9; reprinted in Landau, *Rezensionen*, 277–83, here 281. For the significance of this review see also Johann Georg Hamann's letter to Friedrich Heinrich Jacobi of 13 May 1786, in Johann Georg Hamann, *Briefwechsel*, vol. VI, ed. A. Henkel (Frankfurt am Main: Insel, 1975), 389–90. References to further reviews and critical reactions are to be found in Heiner F. Klemme,

the one hand, and on the other hand to claim 'that the moral law is the condition under which we can first *become aware* of freedom' (*CpV* 5:4fn.). Kant addresses this objection with his famous distinction between freedom as *ratio essendi* of the moral law, and the moral law as *ratio cognoscendi* of freedom: 'Lest anyone suppose that he finds an *inconsistency* when I now call freedom the condition of the moral law and afterwards, in the treatise, maintain that the moral law is the condition under which we can first *become aware* of freedom, I want only to remark that whereas freedom is indeed the *ratio essendi* of the moral law, the moral law is the *ratio cognoscendi* of freedom' (*CpV* 5:4fn.).

(b) According to Kant,

the most considerable objections to the *Critique* that have so far come to my attention turn about just these two points: namely, *on the one side* the objective reality of the categories applied to noumena, denied in theoretical cognition[13] and affirmed in practical, and *on the other side* the paradoxical requirement to make oneself as subject of freedom a noumenon but at the same, with regard to nature, a phenomenon in one's own empirical consciousness[14]

These objections, raised by Johann Georg Heinrich Feder[15] and Hermann August Pistorius,[16] for example, can only be answered by 'a detailed *Critique of Practical Reason*' (*CpV* 5:6–7), where the 'concepts of morality and freedom' (*CpV* 5:6) are clearly determined. Hence, the *Critique of Practical*

'Einleitung' and 'Sachanmerkungen', in Immanuel Kant, *Kritik der praktischen Vernunft*, ed. H. D. Brandt and H. F. Klemme (Hamburg: Meiner, 2003) and in Heiner F. Klemme, 'Praktische Gründe und moralische Motivation. Eine deontologische Perspektive', in H. F. Klemme, M. Kühn and D. Schönecker (eds.), *Moralische Motivation. Kant und die Alternativen* (Hamburg: Meiner, 2006), 113–53.

[13] See *CrV* A378–9 and B429–30. [14] *CpV* 5:6.

[15] *Zugabe zu den Göttingischen Anzeigen von gelehrten Sachen*, third piece, January 19, 1782, 40–8, here 46; reprinted in Landau, *Rezensionen*, 10–17, here 16.

[16] Hermann Andreas Pistorius comments on this point in his anonymous review of Johann Schultz's *Erläuterungen über des Herrn Professor Kant Critik der reinen Vernunft* (Königsberg, 1784) as follows: 'I further ask how it is possible, if man's entire soul, his entire faculty of representation with all its impacts is regarded as appearance (how it has to be done according to the principles the author established and to his notion of space and time, as I understand it), to interpret a part of this soul – and what else is reason – as a noumenon, or a thing in itself! How do we know, given we are completely ignorant of the intelligible world and the *things in themselves*, that something belonging to man's subjective and apparent power of thought, i.e. his reason, and thus he himself as well, insofar he is provided with reason, a part of the world of the understanding, *is a thing in itself*!' (*Allgemeine deutsche Bibliothek*, 1786, 111; Landau, *Rezensionen*, 342; for Pistorius see also Bernward Gesang (ed.), *Kants vergessener Rezensent. Die Kritik der theoretischen und praktischen Philosophie Kants in fünf Rezensionen von Hermann Andreas Pistorius* (Hamburg: Meiner, 2007). Kant also refers to this review in a preparatory work for the preface to the second *Critique*, see 21:416.

Reason has to show how the concept of causality through freedom obtains a non-empirical meaning.

(c) With the second chapter of the Analytic ('Of the Concept of an Object of Pure Practical Reason'), Kant reacts to an objection raised by Pistorius.[17] According to Pistorius, Kant mistakenly did not establish the concept of the good '*before the moral principle*' (*CpV* 5:9). In this chapter, Kant tries to justify 'the paradox of method', according to which '*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 (as was done here) after it and by means of it*.'[18]

Without these objections,[19] Kant probably would not have seen a reason to enlarge his *Critique of Pure Reason* with a 'Critique of Pure Practical Reason'. Actually, in none of the other critical writings does he deal with his critics to such an extent as in the *Critique of Practical Reason*. But what exactly explains the shift from the 'Critique of Pure Practical Reason' (as an appendix to the *Critique of Pure Reason*) to the *Critique of Practical Reason* (as a monograph) that occurred between November 1786 and April 1787? Was this shift due to pragmatic or merely technical reasons, as some commentators believe?[20] Or did Kant, in the context of his engagement with his critics, arrive at new insights into the nature of pure reason that forced him to publish a *Critique of Practical Reason*? Generally, two scenarios can be distinguished. According to the *first* scenario, Kant initially intended to reject the arguments against his doctrine of the categorical imperative through a line of thought belonging to

[17] In his review of the *Groundwork*, Pistorius writes: 'First, the author remarks that only a good will can be regarded as good without qualification, and that this will is good only by the act of willing, i.e. it is good in itself, and not by what it causes and achieves or by its capability of achieving some predetermined end … The author admits that there is something disconcerting in this principle for the estimation of the value of the will, although it is supposed to be accepted also by common reason; therefore, he considers it necessary to further examine it once again. I primarily wish the author had discussed the general concept of *good*, and specified what he understands with it, because obviously, we first would have to agree on this before we were able to work out anything concerning the absolute value of a good will. Therefore, I am entitled to ask: what is good in general, and what is a good will in particular?' (*Allgemeine deutsche Bibliothek*, 1786, 448–9; Landau, *Rezensionen*, 354–5).

[18] *CpV* 5:62–3. Regarding the 'paradox of method', see Reinhard Brandt, *Die Bestimmung des Menschen bei Kant* (Hamburg: Meiner, 2007), 361–70.

[19] Further remarks are to be found in Klemme, 'Einleitung' and 'Sachanmerkungen', and Allen W. Wood, 'Preface and Introduction (3–16)', in O. Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft* (Berlin: Akademie Verlag, 2002), 25–41.

[20] This is Allen W. Wood's thesis: 'But obviously in the course of preparing a second edition of the *Critique of Pure Reason*, it became evident to Kant that neither the time allotted for completion of his revisions nor reasonable boundaries of length for what was already a very long book would permit him to do things this way' (Wood, 'Preface and Introduction', 26; see also Beck, *Kants 'Kritik der praktischen Vernunft'*, 25).

the subject areas of Section III of the *Groundwork*, but then 'discovered' the 'fact of pure reason' that motivated him to write the *Critique of Practical Reason*. According to the *second* scenario, the crucial step towards the second *Critique* could not have occurred before the 'antinomy of pure practical reason' had been discovered, whether or not Kant has developed the doctrine of the fact of pure reason prior to November 1786. Or put differently: by discovering the antinomy, Kant accomplishes the transition from the 'Critique of Pure Practical Reason' to the *Critique of Practical Reason*.

In what follows, I would like to side with the second scenario. I want to make the case that after November 1786, Kant redetermined the relationship of empirical practical and pure practical reason. Whereas in the *Groundwork*, he speaks of a '*natural dialectic*' (*G* 4:405) into which common practical reason inevitably falls when reasoning about the categorical obligation of the moral law, in the *Critique of Practical Reason* Kant undermines any basis for doubt in the reality of the moral law with his doctrine of the 'fact of pure reason'. Instead, Kant 'discovers' that when reason is practically applied to determine the concept of the highest good, an (apparent) antinomy results. This 'dialectic of pure practical reason' (*CpV* 5:107) (which had a place neither in the *Critique of Pure Reason* nor in the *Groundwork of the Metaphysics of Morals*, and is without a historical model[21]) explains, by its sheer existence, why there has to be a separate *Critique of Practical Reason* (see section III).

This discovery is accompanied by a second one regarding the realm of our a priori knowledge. Before 1787, Kant holds that also the feeling of pleasure and displeasure can be evaluated a priori, which would be part of a *Groundwork of the Critique of Taste* or the *Critique of Judgment*, respectively. This is accompanied by a redetermination of critical philosophy: the *one* critique of pure reason[22] is realized in three separate critiques, dedicated to the understanding (*Critique of Pure Reason* or 'critique of speculative reason'), to the power of judgment ('foundation of the critique of taste' or *Critique of Judgment*) and to reason (*Critique of Practical Reason*) (see section II).

[21] See Michael Albrecht, *Kants Antinomie der praktischen Vernunft* (Hildesheim: G. Olms Verlag, 1978), 15, 133 and Milz, *Der gesuchte Widerstreit*, 265.

[22] In the so-called 'First Introduction to the Critique of Judgment', Kant speaks of the 'critique of pure reason (taken in the most general sense)' (20:241).

II. THE TRIAD OF UPPER FACULTIES AND THE 'CRITIQUE OF TASTE'

In the note published in *Allgemeine Literatur-Zeitung* in November 1786 quoted above, Kant points to the significance of the 'Critique of Pure Practical Reason' for accomplishing the 'critical investigations'. Here, a first reason for the separate publication of the *Critique of Practical Reason* is hinted at. In order to understand it, we have to turn towards an aspect of the history of Kant's development in the 1780s, although there is not a single word on it in the second *Critique*, namely, the discovery that even our judgment of taste is a priori.[23] If taste can be given and needs a critique, it seems natural to devote a separate critique to each of the three faculties of cognition, feeling and volition. With the decision to write a 'Groundwork of the Critique of Taste' – leading to the *Critique of Judgment* – it no longer makes sense to publish the 'Critique of Pure Practical Reason' as an appendix to the *Critique of Pure Reason*.[24]

When exactly does Kant decide to write a *Critique of Taste*? The announcement of a 'Groundwork of the Critique of Taste' (*Grundlegung zur Critik des Geschmacks*) that Kant's publisher Johann Friedrich Hartknoch in Riga arranged for the Leipzig exhibition catalogue[25] for spring 1787 is particularly telling. Johann Bering refers to it in a letter to Kant of 28 May 1787: 'As much as I was delighted when reading in the Leipzig exhibition catalogue that you [Ew. Wohlgeb.] endow us again with a foundation of the critique of taste, in addition to the new edition of the Critique; so much was I saddened, since I did not find what I have so long desired, namely, the system of pure speculative and practical philosophy. If only you would take delight in endowing us with it soon' (10:488). Bering could not know that Kant was already working on the fulfilment of his wish. In the preface to the *Critique of Practical Reason* Kant emphasizes – also as a

[23] See Piero Giordanetti, 'Kants Entdeckung der Apriorität des Geschmacksurteils. Zur Genese der Kritik der Urteilskraft', in H. F. Klemme *et al.* (eds.), *Aufklärung und Interpretation* (Würzburg: Königshausen & Neumann, 1999), 171–96.

[24] Förster argues in a similar way: 'The critique of taste would thus have to be integrated into the new edition of the first Critique, together with its analogue in moral philosophy. Therefore, it can hardly be surprising that Kant decided in favour of three separate Critiques.' (Eckart Förster, 'Kant und Strawson über ästhetische Urteile', in J. Stolzenberg (ed.), *Kant in der Gegenwart* (Berlin and New York: de Gruyter, 2007), 269–89, here 276.) However, Förster does not consider that Kant initially intended to write a 'Critique of Pure Practical Reason', and accordingly, he does not discuss possible reasons to explain the turn towards the *Critique of Practical Reason*.

[25] 'Dessen Grundlegung zur Critik des Geschmacks. gr. 8. Ebenda [= Riga: Hartknoch]', Weidmann, *Leipziger Meßkatalog Ostern 1787* (see 5:515).

reaction to Pistorius' criticism – the relevance of this work to the systematicity of his philosophy:

> Now, the concept of freedom, insofar as its reality is proved by an apodictic law of practical reason, constitutes the *keystone* of the whole structure of a system of pure reason, even of speculative reason; and all other concepts (those of God and immortality), which as mere ideas remain without support in the latter, now attach themselves to this concept and with it and by means of it get stability and objective reality, that is, their *possibility* is *proved* by this: that freedom is real, for this idea reveals itself through the moral law.[26]

The *Critique of Practical Reason* rebuts the objections against the *Critique of Pure Reason* by providing 'objective reality' to its ideas of God and immortality with the concept of practical freedom.[27]

But what could be the function of a 'Groundwork of the Critique of Taste' within the critical architectonic, if the 'keystone' of the system has already been set? In June 1787, Kant is still announcing a 'Groundwork of the Critique of Taste',[28] but changes his plans in December 1787 at the latest: the foundation *of*, or rather *for*, the critique of taste turns into *the* 'Critique of Taste'. In his letter to Reinhold of December 1787, Kant comments on the genesis of his third *Critique* as follows:

> I am now at work on the critique of taste, and I have discovered a new sort of a priori principle, different from those heretofore observed. For there are three faculties of the mind: the faculty of cognition, the faculty of feeling pleasure and displeasure, and the faculty of desire. In the Critique *of Pure* (theoretical) *Reason*, I found a priori principles for the first of these, and in the Critique *of Practical Reason*, a priori principles for the third. I tried to find them for the second as well, and though I thought it impossible to find such principles, the analysis of the previously mentioned faculties of the human mind allowed me to discover a systematicity, giving me ample material at which to marvel and if possible to explore, material sufficient to last for the rest of my life. This systematicity put me on the path to recognizing the three parts of philosophy, each of which has its a priori principles, which can be enumerated and for which one can delimit precisely the knowledge that may be based on them; theoretical philosophy, teleology, and practical philosophy, of which the second is, to be sure, the least rich in a priori grounds of determination.[29]

[26] *CpV* 5:3–4.

[27] Kant emphasizes this also in the preface to the second edition of the *Critique of Pure Reason*, see *CrV* Bxxix–xxx and B395fn.

[28] See his letter to Christian Gottfried Schütz of 25 June 1787 (10:490). The Academy Edition mistakenly reads '*Grundlage*' (10:490) instead of 'Grundleg.'. For the corrected text, see Werner Stark, *Nachforschungen zu Briefen und Handschriften Immanuel Kants* (Berlin: Akademie Verlag, 1993), 228.

[29] 10:514–15.

So, Kant had been planning to work on a monograph entitled 'Groundwork of the Critique of Taste' since spring 1787 at the latest, which he refers to as 'Critique of Taste' in a letter to Reinhold in December of the same year. The decision to publish a *Critique of Practical Reason*, already established in April 1787, and its actual publication in December 1787 are therefore closely connected to the decision initially to publish a 'Groundwork of', and subsequently a 'Critique of Taste'.

Kant's discovery of the a priori character of the judgment of taste provides him with an excellent reason to divide the project of the *one* 'critique of pure reason' into three separate critiques, which are yet materially related with regard to our three higher faculties (understanding, judgment and reason).[30] It should not come as a surprise here that in the *Critique of Practical Reason*, Kant does not assign the function of demonstrating the *unity* of the three critical writings to the 'Critique of Taste'. This function is assigned to it only in the *Critique of Judgment*. Because the 'Critique of Taste' is not a critique of either pure speculative or practical reason, but rather dedicates itself to a critical examination of the reflective *power of judgment*, Kant was able to declare 'the structure of a system of pure reason' (*CpV* 5:3) completed in 1787, without thereby completing his 'critical enterprise' (10:494) as well, as he had announced in November 1786. He completed this 'enterprise' by his own account only with and in the *Critique of Judgment*.[31] As he puts it in the 'First Introduction to the *Critique of Judgment*': 'Philosophy, as a doctrinal system of the cognition of nature as well as freedom' does not acquire 'a new part' by the critical examination of the power of judgment (20:205), but would remain incomplete as the system of critical philosophy.

The relevance of the triad of higher faculties to Kant's decision to write a separate *Critique of Practical Reason* is sustained by formulations in the prefaces to the second edition of the *Critique of Pure Reason*,[32] to the *Critique of Practical Reason*[33] and to the *Critique of Judgment*. To give but one example, Kant begins the preface to the third *Critique* as follows:

[30] See *CrV* A130/B169.

[31] 'Thus with this I bring my entire critical enterprise to an end' (*CU* 5:170).

[32] 'Since during these labors I have come to be rather advanced in age (this month I will attain my sixty-fourth year), I must proceed frugally with my time if I am to carry out my plan of providing the metaphysics both of nature and of morals, as confirmation of the correctness of the critique both of theoretical and practical reason' (*CrV* Bxliii).

[33] 'It may be observed throughout the course of the critical philosophy (of theoretical as well as practical reason)' (*CpV* 5:9fn.).

> The faculty of cognition from a priori principles can be called *pure reason*, and the investigation of its possibility and boundaries in general can be called the critique of pure reason; although by this faculty only reason in its theoretical use is understood, as was also the case in the first work under this title, without bringing into the investigation its capacity as practical reason, in accordance with its special principles.[34]

Kant instantly correlates these remarks to the three faculties: the *Critique of Pure Reason* is correlated with the understanding (in a constitutive respect) and the *Critique of Practical Reason* with reason (in a constitutive respect). The latter 'contains constitutive principles a priori … with regard to the *faculty of desire*'. The *one* project of a critique of pure reason is realized in three steps, according to Kant's remarks from 1790: 'the critique of the pure understanding, of the pure power of judgment, and of pure reason, which faculties are called pure because they are legislative a priori'.[35] This new determination of the relationship between critique and faculty explains why Kant abandons the 'Critique of Pure Practical Reason' as an appendix to the *Critique of Pure Reason*.

III. FROM THE *GROUNDWORK OF THE METAPHYSICS OF MORALS* TO THE *CRITIQUE OF PRACTICAL REASON*

Certainly, discovering a priori principles of the feeling of pleasure and displeasure was an important reason for Kant's decision to write a second *Critique*, but it was certainly not the only one. Since he mentions the 'Critique of Pure Practical Reason' (*G* 4:391; cf. *G* 4:446 ff.) in the *Groundwork*, it can be assumed that a comparison between the *Groundwork* and the *Critique of Practical Reason* provides us with another clue why Kant published the *Critique of Practical Reason* in late 1787 instead of a 'Critique of Pure Practical Reason'.

In the *Groundwork*, Kant distinguishes between the 'critique of pure speculative reason' and the 'Critique of Pure Practical Reason'.[36] In 1785, however, he does not yet intend to devote a monograph to pure practical reason that will go beyond the discussion of the third section of the *Groundwork* ('Transition from Metaphysics of Morals to the Critique of Pure Practical Reason') *before* writing the *Metaphysics of Morals*. Rather, Kant's opinion in 1785 seems to be that the *Groundwork* is entirely sufficient as a 'preliminary work' (*G* 4:391) to

[34] *CU* 5:167.
[35] *CU* 5:179. For the suggestion of a fourth *Critique*, apart from the three *Critiques* from 1781 (1787), 1787/8 and 1790, see Brandt, *Die Bestimmung des Menschen*, 497 ff.
[36] *G* 4:391; cf. *CrV* A841/B869.

the *Metaphysics of Morals*. A critique of pure practical reason is dispensable not least because, contrary to the theoretical application of our pure reason ('it is wholly dialectical'), in its moral application 'human reason can easily be brought to a high degree of correctness and accomplishment, even in the most common understanding' (*G* 4:391).

This statement is not completely unequivocal, however. Does Kant claim that the practical application of pure reason is *not* dialectical, contrary to its theoretical application? Or does he want to point out that 'in moral matters human reason' is involved with a certain dialectic, but one that can be resolved 'easily', contrary to the dialectic in the *Critique of Pure Reason*? Kant takes up the issue of the dialectic at the end of the first section of the *Groundwork* and interprets it in a way that supports the second alternative. After he has found the principle of morals in 'common rational moral cognition' (*G* 4:393 ff.),[37] he states:

> Yet we cannot consider without admiration how great an advantage the practical faculty of appraising has over the theoretical in common human understanding. In the latter, if common reason ventures to depart from laws of experience and perceptions of the senses it falls into sheer incomprehensibilities and self-contradictions, at least into a chaos of uncertainty, obscurity, and instability. But in practical matters, it is just when common understanding excludes all sensible incentives from practical laws that its faculty of appraising first begins to show itself to advantage.[38]

Although the practical application of pure reason does *not* lead to a dialectic (based on an antinomy), we get entangled in a dialectical process when making (general) practical use of our reason that enforces the transition from common to philosophical rational cognition concerning morals. This process is explained by our sensible nature: as beings that strive for the satisfaction of their inclinations, human beings are motivated to reason *against* the strict demands of pure reason as soon as they interfere with their pursuit of happiness:

> Now reason issues its precepts unremittingly, without thereby promising anything to the inclinations, and so, as it were, with disregard and contempt for those claims, which are so impetuous and besides so apparently equitable (and refuse to be neutralized by any command). But from this there arises a *natural dialectic*, that is, a propensity to rationalize against those strict laws of duty and to cast doubt upon their validity, or at least upon their purity and strictness, and where possible, to make them better suited to our wishes and inclinations.[39]

When we want to provide 'access and durability' (*G* 4:405) to the moral law that is based on the application of pure reason, we have to resolve this – (as

[37] See *G* 4:403. [38] *G* 4:404. [39] *G* 4:405.

we might call it) weak – dialectic located between this rational insight and our desire for happiness. This can only happen within the frame of a metaphysic of morals that repudiates any doubt about the 'purity and strictness' (*G* 4:405) of the moral law. The *Groundwork* is necessary for purely practical reasons, because

> *common human reason* is impelled, not by some need of speculation (which never touches it as long as it is content to be mere sound reason), but on practical grounds themselves, to go out of its sphere and to take a step into the field of *practical philosophy*, in order to obtain there information and distinct instruction regarding the source of its principle and the correct determination of this principle in comparison with maxims based on need and inclination, so that it may escape from its predicament about claims from both sides and not run the risk of being deprived of all genuine moral principles through the ambiguity into which it easily falls. So there develops unnoticed in common practical reason as well, when it cultivates itself, a *dialectic* that constrains it to seek help in philosophy, just as happens in its theoretical use; and the first will, accordingly, find no more rest than the other except in a complete critique of our reason.[40]

Because the third section of the *Groundwork* announces the 'transition from metaphysics of morals to the critique of pure practical reason', and all dialectic requires a critique, clearly one should search for the resolution of the 'natural dialectic' in this section. However, Kant's '*natural dialectic*' does not establish a direct transition from 'Metaphysics of Morals' to 'Critique of Pure Practical Reason'. Rather, the attempt to show '*on what grounds the moral law is binding*',[41] announced at the end of the first section, and to resolve the 'natural dialectic' in doing so, leads into 'a kind of circle' that requires critique in turn:

> It must be freely admitted that a kind of circle comes to light here from which, as it seems, there is no way to escape. We take ourselves as free in the order of efficient causes in order to think ourselves under moral laws in the order of ends; and we afterwards think ourselves as subject to these laws because we have ascribed to ourselves freedom of will: for, freedom and the will's own lawgiving are both autonomy and hence reciprocal concepts.[42]

[40] *G* 4:405. 'But such a completely isolated metaphysics of morals … is not only an indispensable substratum of all theoretical and surely determined cognition of duties; it is also a desideratum of utmost importance to the actual fulfilment of their precepts' (*G* 4:410; cf. *G* 4:389–90, 411–12).

[41] *G* 4:450.

[42] *G* 4:450. For a discussion of the suspected circle see, among others, Reinhard Brandt, 'Der Zirkel im dritten Abschnitt von Kants "Grundlegung zur Metaphysik der Sitten"', in H. Oberer and G. Seel (eds.), *Kant. Analysen – Probleme – Kritik* (Würzburg: Königshausen & Neumann, 1988), 169–91; Dieter Schönecker, *Kant: Grundlegung III. Die Deduktion des kategorischen Imperativs* (Freiburg and Munich: Alber, 1999); Marcel Quarfood, 'The Circle and the Two Standpoints (GMS III, 3)', in C. Horn and D. Schönecker (eds.), *Groundwork for the Metaphysics of Morals* (Berlin and New York: de Gruyter, 2006), 285–300; and Jens Timmermann, *Kant's 'Groundwork of the Metaphysics of Morals'. A Commentary* (Cambridge: Cambridge University Press, 2007), 131–4.

The suspicion that there might be a circle is cleared by answering the question about the 'Whence'[43] of our obligation. In doing so, pure practical reason appeals to the *Critique of Pure Reason*, which, with its distinction between things in themselves and appearances, indicates the point from which the concept of obligation originates, and the suspicion of a circle is eliminated. We have an obligation because we exist in the *mundus sensibilis* and the *mundus intelligibilis* at the same time: '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 we regard ourselves as belonging to the world of sense and yet at the same time to the world of understanding' (*G* 4:453; cf. *G* 4:450). The suspicion of a circle is resolved together with the '*natural dialectic*', since the critique of speculative reason takes care 'that practical reason may have tranquillity and security from the external attacks that could make the land on which it wants to build a matter of dispute' (*G* 4:457). These 'external attacks' are stated in the guise of the question 'why the universal validity of our maxim as a law must be the limiting condition of our actions' (*G* 4:449). The '*natural dialectic*' is based on precisely this question – and with its resolution, the *Groundwork* has accomplished its practical end.[44]

When we correlate our interpretation of the '*natural dialectic*' with the question why Kant gave up his project of a 'Critique of Pure Practical Reason', we first have to search for the position of this dialectic in the *Critique of Practical Reason*. As a matter of fact, there is no space for the 'weak' '*natural dialectic*' in the second *Critique* anymore.[45] It is not the question about the 'Whence' of our moral obligation, which has been settled with the 'fact of pure reason' and the feeling of respect for the moral law, but rather the *defence* of this 'fact'. Kant has to show that happiness does not indirectly gain authority over our volition via the highest good, and that it does not undermine the 'purity' of the moral law. In its search for the

[43] Timmermann, *Kant's 'Groundwork of the Metaphysics of Morals'*, 131fn. 25, rightly points to the fact that 'whence' is the correct English translation of '*woher*', and not 'on what grounds' or 'how'. However, the German '*woher*' has both a spatial *and* a normative meaning in the context Kant uses it.

[44] See *G* 4:406, 410.

[45] So also Milz, *Der gesuchte Widerstreit*, 202, 277–8, and 316–17, and Eckart Förster, *Kant's Final Synthesis. An Essay on the* Opus postumum (Cambridge, MA: Harvard University Press, 2000), 197. I do not think that Greenberg's opinion is convincing, according to which the differences between the 'natural dialectic' of the *Groundwork*, the antinomy in the second *Critique* and the chapter on the canon of pure reason in the *Critique of Pure Reason* are due to the different intentions of these works (Sean Greenberg, 'From Canon to Dialectic to Antinomy: Giving Inclinations Their Due', *Inquiry* 48 (2005), 232–48). This interpretation ignores significant modifications of Kant's ethics in the 1780s (see, among others, Klemme, 'Praktische Gründe und moralische Motivation').

highest and last object of its volition, pure reason engenders a conflict, namely, whether our subjective desire for happiness motivates us to virtue, or whether virtue causes our happiness (see *CpV* 5:113). Thus, the '*natural dialectic*' of 1785, based on our sensible propensity to 'reason' against the validity of the moral law, turns into a dialectic in the concept of the highest good in 1787. Therefore, Kant changes the object of doubt (that proves unfounded in the end): whereas in 1785, doubt is cast upon the categorical validity of the moral law (hence upon freedom and our concurrent citizenship in both *mundus intelligibilis* and *mundus sensibilis*), in the *Critique of Practical Reason* doubt arises from the object[46] of our volition with the doctrine of the highest good: Do I have a motive to act virtuously only when I can hope that my virtuous aspiration causes my happiness? Or is it my desire for happiness that induces me to act virtuously?

Indeed, Kant's plan to begin the second *Critique* with the doctrine of the fact of pure reason, which supersedes the systematic place 'of this vainly sought deduction of the moral principle' (*CpV* 5:47) of the *Groundwork*, has to be considered ingenious (even if perhaps not convincing on a systematical level). The validity of the categorical imperative is given relief from the '*natural dialectic*', and the doubt of the critiques is explained by the 'Antinomy of Pure Practical Reason' (*CpV* 5:113–14), which they cannot resolve because of their lack of the proper (critical) instruments. Along these lines, Kant had mentioned to Reinhold in December 1787 that the *Critique of Practical Reason* demonstrates the contradictions in which those who 'refuse to abandon their botched job' are 'unavoidably caught up'. Those who criticize the morality of the categorical imperative are confronted with a contradiction that is inevitable on the one hand, and whose resolution, on the other hand, at the same time proves the failure of the eudaimonist conception of morality.[47]

With the doctrine of a consciousness of the moral law that is immune to doubt – the '*data*' of the 'original practical principles lying in our reason' (*CrV* Bxxviii) – Kant rejects, in the manner of a *coup de main*, two objections: first the objection that he did not prove the moral law, and second, the objection that he did not give a meaning to the concept of freedom that transcends the state of discussion of the *Critique of Pure Reason*, in both cases because man's transcendental freedom is established with the practical reality thought in the moral law, whose possibility had to

[46] See Eckart Förster, 'Die Dialektik der reinen praktischen Vernunft (107–121)', in Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft*, 173–86, here 174.

[47] Similarly Albrecht, *Kants Antinomie der praktischen Vernunft*, 133.

be left open in the first *Critique*. Already in its preface, Kant determines the relation between freedom and moral law as follows: 'For, had not the moral law *already* been distinctly thought in our reason, we should never consider ourselves justified in *assuming* such a thing as freedom (even though it is not self-contradictory). But were there no freedom, the moral law would *not be encountered* at all in ourselves' (*CpV* 5:4). Pure practical reason fills a gap in the system of human reason. With it and with the concept of freedom 'the enigma of the critical philosophy is first explained: how one can *deny* objective *reality* to the supersensible *use of the categories* in speculation and yet *grant* them this *reality* with respect to the objects of pure practical reason; for this must previously have seemed *inconsistent*, as long as such a practical use is known only by name' (*CpV* 5:5).

Moreover, the 'dialectic of pure practical reason' reveals a parallel between the speculative and the practical application of pure reason. In both cases, we want to know the unconditioned: 'Pure reason always has its dialectic, whether it is considered in its speculative or in its practical use; for it requires the absolute totality of conditions for a given conditioned, and this can be found only in things in themselves' (*CpV* 5:107).

It seems that with the publication of the *Critique of Practical Reason* – as regards its doctrinal aspect – the harmony between the pure speculative and the practical application of reason has been perfected: The *Critique of Practical Reason* confirms the reality of our moral consciousness with the concept of practical freedom, which is secured, for its part, by the transcendental idealism of the *Critique of Pure Reason*. With the doctrine of the 'fact of pure reason', the primacy of the moral law over the highest good of our volition is certain – against the objection by Pistorius. The validity of the moral law is not challenged by the highest good as the final end of our volition. Our respect for the moral law and our hope for happiness are two sides of one and the same practical application of reason. Put according to a formulation in the *Moral-Kaehler* from the mid-1770s: 'Every upright person has this belief, it is impossible that he can be upright without at the same time hoping, according to the analogy to the physical world, that such behaviour really has to be rewarded. He believes in the reward for the same reason he believes in virtue.'[48] Because there is only one pure reason,

[48] Immanuel Kant, *Vorlesung zur Moralphilosophie*, ed. W. Stark (Berlin and New York: de Gruyter, 2004), 82. Accordingly, Kant writes in section 87 in the *Critique of Judgment*: 'The moral law, as the formal rational condition of the use of our freedom, obligates us by itself alone, without depending on any sort of end as a material condition; yet it also determines for us, and indeed does so a priori, a final end, to strive after which it makes obligatory for us, and this is the *highest good in the world* possible through freedom' (*CU* 5:450).

the question about the determining ground (what should I do?) cannot be separated from the question about the final end of our volition (what may I hope?).

IV. CONCLUSION

The second *Critique* originates out of an expansion of a 'critique of pure practical reason' that Kant initially intends to provide for the *Critique of Pure Reason*, where he wants to remove objections raised against his conception of the practical application of reason and its compatibility with the pure speculative application of reason. The transition from the 'Critique of Pure Practical Reason' to the *Critique of Practical Reason* has two main explanations:

(1) With the discovery of the a priori character of the judgment of taste, and the plan to write a 'Critique of Taste', Kant modifies his overall critical plan: The *one* critique of pure reason is now presented in three works that are dedicated to (a) the constitutive application of the understanding and the speculative application of pure reason, (b) the constitutive application of pure practical reason and (c) the regulative application of the power of judgment in the determination of the feeling of pleasure and displeasure.

(2) With the doctrine of the 'fact of pure reason' and the 'dialectic of pure practical reason', Kant goes beyond the theoretical resources of the *Groundwork* with the intention of rebutting important objections raised by his critics. With the doctrine of the 'fact of pure reason', pure practical reason makes itself immune to the objection that it is not the determining ground of our volition. And with the transformation of the '*natural dialectic*' of the *Groundwork*, those who are sceptical about the unconditioned validity of the moral law are referred to a contradiction in the concept of the highest good that exposes the 'rational' core of their criticism of Kant's philosophy of pure reason, and, at the same time, Kant rebuts this criticism by resolving the 'antinomy of practical reason'.

With the resolution of the 'antinomy of practical reason', aided by the *Critique of Pure Reason*, Kant has demonstrated, according to his own intention, (a) that the determination of the highest good of our volition has to follow *after* the moral law has been determined, (b) that pure practical reason does not require common sense to dispense with happiness, but shows that we become worthy of happiness as a result of our virtuous striving and (c) that pure speculative reason secures the concept of freedom

of pure practical reason and thereby, in turn, provides 'objective reality' to the ideas of speculative reason (God, freedom, immortality). In doing so, the '[c]ritical resolution of the antinomy of practical reason' establishes the 'primacy of pure practical reason in its connection with speculative reason' (*CpV* 5:119).

Although both *Critiques* together lay the foundations 'for a scientific system of philosophy, both theoretical and practical' (*CpV* 5:12), their argumentative aims are inverted. The first *Critique* rejects pure speculative reason's request to consider the conditions of thinking sufficient conditions of theoretical knowledge, and the second *Critique* rejects empirical practical reason's demand to be the only practical instance as dogmatic presumption.[49]

More pointedly, the transition from the 'Critique of Pure Practical Reason' to the *Critique of Practical Reason* is explained by Kant's philosophical insights regarding the nature of pure reason. As pure practical reason, it manifests its reality in the guise of our consciousness of the moral law, but not without becoming entangled in a dialectical contradiction, just as pure speculative reason does. For the resolution of this contradiction, it depends on the *Critique of Pure Reason* as much as the latter benefits from the *Critique of Practical Reason* providing its speculative ideas with 'objective reality'. From a philosophical point of view, the *Critique of Practical Reason* is the consequence of the emancipation of the practical application of reason from the theoretical one, which is accompanied by a reorganization of the 'critical business' in the guise of three complementary critiques.

Has the *Critique of Practical Reason* been successful in the end? We have good reason to answer this question in the negative. With the analytic and dialectic, Kant creates a tension that threatens the unity of his critical work of 1787/8. Whereas he claims, on the one hand, that the moral law is given to us as a fact, he emphasizes, on the other hand, that the 'antinomy of practical reason' puts just this fact in question. However, if the fact is given, how can the suspicion that pure reason contradicts itself in its practical application come up at all? This suspicion of an antinomy obviously suits Kant, because only thus can he assign the criticism of the validity of the categorical imperative a place within the system of pure reason, and at the same time prove that this criticism is unfounded. As he had already done

[49] See Dieter Henrich, 'Ethik der Autonomie', in D. Henrich, *Selbstverhältnisse. Gedanken und Auslegungen zu den Grundlagen der klassischen deutschen Philosophie* (Stuttgart: Reclam, 1982), 11–13. Translated as 'Ethics of Autonomy', trans. Manfred Kuehn, in Dieter Henrich, *The Unity of Reason: Essays on Kant's Philosophy*, ed. Richard Velkley (Cambridge, MA: Harvard University Press, 1994), 89–123, here 92–4.

in the *Critique of Pure Reason*, Kant searches for the 'rational' occasion of our doubts about the reality of pure reason in the *Critique of Practical Reason* as well. Whether he was able to remove it remains debatable in view of the tension discussed.

In the end Kant might have felt that something went wrong with his *Critique of Practical Reason*, because in the *Critique of Judgment*, he seems to renounce his earlier view of the relationship between the moral law and the highest good. The moral proof for the existence of God

> is not meant to say that it is just as necessary to assume the existence of God as it is to acknowledge the validity of the moral law, hence that whoever cannot convince himself of the former can judge himself to be free from the obligations of the latter. No! All that would have to be surrendered in that case would be the *aim* of realizing the final end in the world (a happiness of rational beings harmoniously coinciding with conformity to the moral law, as the highest and best thing in the world) by conformity to the moral law. Every rational being would still have to recognize himself as forever strictly bound to the precept of morals; for its laws are formal and command unconditionally, without regard to ends (as the matter of the will).[50]

[50] *CU* 5:451.

CHAPTER 2

Formal principles and the form of a law

Andrews Reath

1. INTRODUCTION

One aim of the *Critique of Practical Reason* is to establish that reason alone can determine the will. To show that it can, it suffices to show that there are practical principles given by reason alone – what Kant terms 'practical laws', or (roughly) requirements of reason on action. Chapter I of the Analytic accomplishes this aim by arguing that the moral law is an authoritative practical principle given as a 'fact of reason'. The chapter begins in section 1 with a 'Definition' (*Erklärung*) of a practical law as a practical principle that holds necessarily for every rational being (as a principle that 'determines the will simply as will' without presupposing any subjective or rationally contingent interests). From this definition Kant uses conceptual arguments to derive three 'Theorems' and various 'Corollaries' that state conditions that any practical law must satisfy. The principal condition, stated in Theorem III, is that a practical law determines choice through its form – through 'the mere form of giving universal law' – rather than through its matter (*CpV* 5:27). The arguments culminate in section 7 with the statement of the 'Fundamental Law of Pure Practical Reason', which Kant identifies with the moral law: 'So act that the maxim of your will could at the same time always hold as a principle in a giving of universal law' (*CpV* 5:30). Section 7 includes a formulation of the fundamental law that follows from the Definition and Theorems – just as in the *Groundwork* a formula of the Categorical Imperative is derived first from the analysis of the concept of duty (*G* 4:402), then from the concept of a categorical imperative (*G* 4:421). But since the moral law is put forward here, in imperative form, as an authoritative practical law, section 7 is the synthetic claim that *Groundwork* III attempts to establish. Thus in short order, these pages of the second *Critique* cover the terrain of the extended argument of the *Groundwork*.

The analytic portions of this chapter contain several striking claims about features of practical laws and fundamental practical principles that centre on

the notion of 'form'. Because Kant's foundational works aim to establish that there are moral requirements with genuine rational necessity, the second *Critique* begins with a definition of a practical law. Kant then claims that a fundamental principle of morality is in some sense a formal principle (Theorem II, Corollary, *CpV* 5:22) and that a practical law provides a ground of choice through its form rather than its matter (Theorem III). Moreover, he claims that only a formal principle can be the basis of categorical imperatives that apply with necessity.[1] He also argues for analytical connections between freedom of the will and the form of a law: that a will 'for which the mere lawgiving form of a maxim can alone serve as a law is a free will' and that if a will is free, the lawgiving form of a maxim is 'the only thing that can constitute a determining ground of the will' (Problems I and II, *CpV* 5:28–9). To assess these claims, we need a handle on certain concepts: what does Kant mean by the form of universal lawgiving, or the form of a practical law? What does it mean to think of the lawgiving form of a practical principle as the determining ground of choice? What does he mean by a 'formal principle' and why does he think that fundamental normative principles must be 'formal'? In particular, why does he think that a practical principle is normatively necessary only if it is a formal principle in his sense (to be discussed in section III below) and provides a ground of choice through its form?

The difficulty in unpacking these claims is to some extent a matter of their familiarity. The arguments found in this chapter of the second *Critique* are compact versions of arguments given in the *Groundwork* that Kant, perhaps, took greater pains to develop in the earlier work. These claims, as they appear in the second *Critique*, rely on various distinctions that Kant now takes for granted. For example, Theorem III simply assumes that there is an exhaustive distinction between the form and the matter of a principle, and that a practical principle provides a ground of choice either through its matter (through a contingent interest in its matter) or through its form. Given the claim that a principle that provides a ground of choice through its matter cannot serve as a practical law (convincingly argued in Theorem I), it readily follows that a practical law provides a ground of choice through its form. The arguments of Problems I and II identify the

[1] '[S]ince material principles are quite unfit to be the supreme moral law (as has been proved), the *formal practical principle* of pure reason (in accordance with which the mere form of a possible giving of universal law through our maxims must constitute the supreme and immediate determining ground of the will) is the *sole* principle that can *possibly* be fit for categorical imperatives, that is, practical laws' (*CpV* 5:41). See also *CpV* 5:64: 'Only a formal law, that is, one that prescribes to reason nothing more than the form of its universal lawgiving as the supreme condition of maxims, can be a priori a determining ground of practical reason.'

matter of a principle with items given in appearance and assume that lawgiving form is not found among appearances. Kant accordingly concludes that the form of a law is a ground of choice that is independent of the causal laws that govern appearances, thus that a will that can be determined by the form of law is transcendentally free. Because arguments like these are so easily made within Kant's philosophical framework, it is hard to get an independent perspective on the basic ideas at issue.

The form of a practical law (the form of universal lawgiving) is commonly understood as the formal feature of necessary and universal applicability – the universal validity of some practical principle for all rational beings, or the fact that a principle makes a necessary demand on all rational agents that excludes the force of competing reasons for action. Stephen Engstrom has pointed to Kant's distinction between 'objective universal validity' and 'subjective universal validity' to suggest a gloss on this notion. Objective universal validity is the applicability of a principle to all objects that fall under the principle. The objective universal validity of a practical principle is the fact that it applies to or governs the actions of all agents in the circumstances covered by the principle. Subjective universal validity is the validity of a principle for all knowing subjects – the fact that it governs judgment and determines how all subjects should think about the matter, 'so that all such subjects could agree in the matter and share the same judgment'.[2] Engstrom observes that the two kinds of universal validity coincide in the case of practical principles:

> The principle applies to the will of every rational being, and every such being can recognize this universal applicability. This is as much as to say that principles of practical cognition are necessarily such that every subject can agree to every subject's acting on them, as would actually happen if all subjects were jointly to legislate this principle for themselves.[3]

According to this suggestion, the form of a practical law is the fact that all subjects can agree to any subject's acting on the principle. To say that the form of a law is a ground of choice is to say that the necessity and universality of a principle (the fact that it is suited to hold as universal law, or makes a necessary demand) is a sufficient justifying reason to adopt or comply with the principle. The lawgiving form of the principle is the ground of a particular agent's choice (i.e. is that agent's reason for acting) when the agent complies simply because the principle meets these criteria and makes a

[2] Stephen Engstrom, 'Introduction', in Kant, *Critique of Practical Reason*, trans. Werner Pluhar (Indianapolis: Hackett Publishing Company, 2002), xxxiv–xxxv. See also his *The Form of Practical Knowledge* (Cambridge, MA: Harvard University Press, 2009), 115–17, 122–7.

[3] Engstrom, 'Introduction', xxxv.

necessary demand on any rational agent. A formal practical principle – or better, since Kant appears to think that there is only one, *the* formal principle of volition – is presumably the directive to act only from maxims that have the form of a law and to regard this demand as a sufficient reason for choice. This principle is seen in the universal law version of the Categorical Imperative, which Kant refers to, variously, as 'the formal principle of volition' and 'the formal principle of pure practical reason.'[4] I accept these basic understandings, but by raising various questions, I will try to fill out what they amount to.

In the next section, I attempt to say something further about Kant's notion of the form of a law. In section III I develop an understanding of what Kant means by a 'formal principle' that explains why formal principles are uniquely suited to apply with normative necessity. In the concluding section I shall use these points to give readings of the arguments for Kant's Theorems I and III and Problem I.

II. THE FORM OF A LAW

Let me rehearse several familiar points that provide some parameters for understanding the idea of the form of a practical law. First, Kant thinks that the form of a law is common to both natural laws and practical laws. Second, some practical principles have the form of a law and some do not. Third, the form of a practical principle inheres in some substance or content that is the 'matter' of the principle. The last two points raise the questions how a principle can provide a ground of choice in virtue of having the form of a law, and whether it is only principles with the form of a law that provide a ground of choice through their form.

On the first point, textual support that Kant thought that the form of a law is common to both natural causal laws and practical laws, should one need it, is found in the 'Typic of Pure Practical Reason'. This brief but obscure section of the second *Critique* gives the theoretical underpinning for the law of nature variant of the Categorical Imperative found in the *Groundwork* (*G* 4:421). Kant wishes to explain how we can judge whether a maxim specifying a proposed choice that can be carried out in the sensible world (e.g. an action to be done for some end or reason, all specifiable in experiential terms) satisfies the criterion of morally good choice – that it has the form of universal law. Practical (moral) judgment is 'subject to special difficulties': since the criterion of morally good choice is a purely rational

[4] *G* 4:400 and *CpV* 5:41.

requirement, it cannot be represented in experiential terms in a way that can be applied directly to a maxim – that is, to subjective principles that can be implemented in the sensible world – to determine whether the maxim 'is a case of' or satisfies the criterion.[5] It is not completely clear why (or whether) this is a problem. But perhaps the idea is, first, that whether a maxim has the form of a law cannot be read directly off the matter of the maxim; and second, that if we could determine whether a criterion is satisfied simply by considering the matter of the maxim, that criterion would employ a 'material principle of morality'.[6] However, in lieu of applying the fundamental criterion of good choice directly to maxims of action, Kant writes that we can use a law of nature as 'a type for the appraisal of maxims in accordance with moral principles. If the maxim is not so constituted so that it can stand the test as to the form of a law of nature in general, then it is morally impermissible' (*CpV* 5:70–1). In other words, we determine whether a maxim has or is consistent with the form of a practical law (satisfies the fundamental criterion of pure practical reason) by determining whether it could hold as a universal law of nature. The fact that we can assess the legislative form of a maxim in this way indicates that 'the form of lawfulness in general' (*CpV* 5:70) is shared between natural laws and practical laws.[7]

[5] As Kant describes the problem: 'it seems absurd to want to find in the sensible world a case which, though as such it stands only under a law of nature, yet admits of the application to it of a law of freedom and to which there could be applied the supersensible idea of the morally good, which is to be exhibited *in concreto*'. The 'special difficulty' is that 'a law of freedom is to be applied to actions as events that take place in the sensible world and so, to that extent, belong to nature'(*CpV* 5:68). The issue is how laws of freedom and the rational idea of the morally good can be applied to actions that are events in the natural world, given the fact that moral principles and ideas of the good cannot be instantiated in experience.

[6] To elaborate the second part of this suggestion, the idea is that any criterion of moral goodness that could apply directly to maxims of action would assess maxims through their matter, and that any such method of assessment would employ a 'material principle of morality'. To see why, assume that the criterion of moral goodness were such that we could determine whether it is satisfied by examining the content of the maxim, or some properties of the maxim accessible through observation and sensible experience (as Kant understands it). That is, assume that we could read off from the content of the maxim whether or not it is morally good. For that to be possible, the criterion of moral goodness would have to be specifiable in experiential terms. One possible criterion of this sort might assess maxims in terms of their consequences – e.g. whether adoption of the maxim promotes some value in which we as a matter of psychological fact take an interest. Another might assess maxims in terms of our disinterested responses to them – e.g. whether disinterested reflection on the maxim would lead to feelings of pleasure or satisfaction (of approval). There may be other possibilities as well, but the thought is that any criterion of goodness specifiable in experiential terms would end up employing what Kant calls a 'material principle of morality'. But no such principle, Kant argues, can have the requisite normative necessity. I'll say more about this point in section IV (a).

[7] Here Kant writes: 'It is permitted to use the nature of the sensible world as the type of an intelligible world, provided that I do not carry over into the latter intuitions and what depends upon them but refer only to the form of lawfulness in general [*die Form der Gesetzmaßigkeit überhaupt*] … For to this extent laws are the same, no matter from what they derive their determining ground' (CpV 5:70).

Regarding the second point, it seems equally clear that some maxims have (or are consistent with) the form of a law and that some do not; otherwise lawgiving form would not provide a criterion of morally good choice. Kant thinks that 'the most common understanding can distinguish without instruction what form of a maxim makes it fit for a giving of universal law and what does not' (*CpV* 5:27).[8] How then might we characterize the form of practical principles that are *not* suited for a giving of universal law? One suggestion is that what Kant calls the 'principle of happiness or self-love' captures the form of a practical principle that is to be contrasted with the form of a law.[9]

The principle of happiness is the principle of making 'a rational being's consciousness of the agreeableness of life uninterruptedly accompanying his whole existence … the supreme determining ground of choice', or the principle of choice that places 'the determining ground of choice in the pleasure or displeasure to be felt in the reality of some object' (*CpV* 5:22). In other words, it is the fundamental principle of finding reasons for action in the fact that an object of choice offers prospective satisfaction. An agent who makes happiness his principle will act so as to secure or increase overall satisfaction of his desires, or will do what he desires most strongly on balance. As I understand Kant's principle of happiness, it sets no limits on the objects in which one takes satisfaction. For example, the objects of desire need not be self-regarding, as we learn from Kant's friend of humanity in *Groundwork* I.[10] (In this respect it is somewhat misleading of Kant to refer to it as the principle of 'self-love'.) The principle of happiness is thus a 'formal' principle in the contemporary sense – it directs action to ensure

[8] Some reason to think otherwise might come from remarks such as the following: 'Now, all that remains of a law if one separates from it everything material, that is, every object of the will (as its determining ground) is the mere *form* of giving universal law' (*CpV* 5: 27). One might take that to mean that once we abstract away the content of any maxim or practical principle (or set aside reasons for adopting the principle based in a contingent interest in the content), what is left is the form of a law – in which case the form of a law would be the general form of any practical principle. But that reading would imply that all practical principles have the form of a law, which is implausible. In this remark, however, Kant is talking about practical laws (*Gesetze*), not principles (*Prinzipien*). That said, it may still be Kant's view that *taking* an action to be objectively good, or *taking* a maxim to be universally valid, is a formal element of all free choice. For discussion of this point, see Reath, 'Autonomy, Practical Law, and Taking One's Choices to Be Good: Replies to Critics', *Philosophical Books* 49:2 (April 2008), 132–4.

[9] I made this suggestion in an earlier essay, 'Hedonism, Heteronomy, and Kant's Principle of Happiness', reprinted in *Agency and Autonomy in Kant's Moral Theory* (Oxford and New York: Oxford University Press, 2006). See sections III–IV and the Appendix.

[10] See Reath, *Agency and Autonomy*, 39–40. I also argue in this essay that Kant does not regard pleasure or satisfaction as the object of all choice done from the principle of happiness, thus that he does not accept a simple form of psychological hedonism for action on this principle.

satisfaction of your desires whatever their objects may be – although abstraction from the objects of volition is not enough to make it a formal principle in Kant's specialized sense. What comes close to making it a formal principle in Kant's sense is that the principle of happiness captures the defining feature or form of one model of human choice – that feature being that the reason for choice lies in expected satisfaction or the ways in which the object of choice answers to antecedently given object-dependent desires and dispositions that are part of an individual's subjective motivational set.[11] This is the structural feature shared by what Kant calls 'material practical principles' – practical principles that have normative force only on the condition that one has an object-dependent desire for, or is antecedently disposed to take satisfaction in, their object.[12] The significance of the principle of happiness for Kant is that it is the basic principle underlying a certain form of practical principle or choice that may be contrasted with choice guided by moral principle.

The third point is that (with the exceptions noted below) every practical principle has a form and a matter, and the form inheres in some matter or content and is the form of some substantive principle. So a practical principle with the form of a law will be a substantive maxim or principle that has the form of a practical law, for example, a substantive principle that makes a necessary demand on anyone.[13] Likewise what has the form captured by the principle of happiness are specific material practical principles. The exceptions are the most general principles, such as the formal principle of morality or the principle of happiness, that abstract from the content that differentiates one specific principle (practical law or material practical principle) from another. I presume that the formal principle of

[11] See Williams, 'Internal and External Reasons', in *Moral Luck* (Cambridge and New York: Cambridge University Press, 1981), 101–13. For the idea of 'object-dependent desires', see Rawls, *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, MA: Harvard University Press, 2000), 46, 151–2. Rawls characterizes them as desires for objects or states of affairs 'that can be described without the use of any moral conceptions, or reasonable and rational principles'. That makes them desires that arise independently of practical reasoning and judgments about value – desires that do not presuppose or are not generated by the acceptance of moral conceptions or the application of principles of reason.

[12] Such principles presuppose 'an object (matter) of the faculty of desire as the determining ground of the will' where 'desire for this object precedes the practical rule and is the condition of its becoming a principle' (*CpV* 5:21). For further discussion of the concept of a material practical principle, see Reath, 'Introduction', in Kant, *Critique of Practical Reason*, trans. and ed. Mary J. Gregor (Cambridge and New York: Cambridge University Press, 1997), xx.

[13] Substantive principles with the form of a law are 'formal principles' in Kant's sense, though they need to be distinguished from *the* formal principle of morality, e.g. the Formula of Universal Law. In discussing formal principles in section III, I will have the latter in mind.

morality, in addition to expressing the requirement that one's maxims have the form of a law, itself has the form of a law.

When Kant holds that a practical law provides a ground of choice through its form, he is thinking in the first instance of substantive principles and maxims that involve either requirements or prohibitions (a principle of providing aid, of fulfilling a voluntarily undertaken obligation, of avoiding dishonesty and so on). In his discussion of the duty to promote the happiness of others, the determining ground of choice is the 'mere lawful form' of this particular principle.[14] So his view is best stated as follows: whether there is sufficient justification to act on some substantive principle turns on the form instantiated in the matter of that principle; the ground of choice is the fact that a certain substantive principle has (or is consistent with) the form of a practical law. The fact that a certain principle, such as a principle of mutual aid, has the form that makes it a requirement is sufficient (decisive) and overriding reason to adopt that principle, whether or not one is inclined or would profit from doing so. The fact that a certain principle does not have or is not consistent with the form of a practical law (cannot be made a universal law) is a sufficient reason to abandon a maxim that otherwise holds considerable interest. In the case of both requirements and prohibitions, the form of a practical law is a ground of choice in a meaningful sense, first, because the fact that a principle has a certain form is a reason for choice (either to adopt or to abandon the principle), and second, because that reason is sufficient to override or exclude competing material reasons based in desire or interest – e.g. that helping would be burdensome or inconvenient, that the deception would further an important personal aim, and so forth.

Permissible maxims introduce complications, but I assume that one can say that they have the form of practical law in an extended sense marked by the phrase 'consistent with the form of practical law'. A maxim contains a proposed action and some set of considerations regarded as good reasons for performing it (such as an end that the action will further, one's circumstances of action and so forth). When a maxim can be made universal law without inconsistency or irrationality, all can agree to anyone's adopting the maxim (to regarding those considerations as good reasons) and no one may

[14] Kant writes: 'Thus the matter of the maxim … must not be the condition of the maxim since the maxim would then not be fit for a law. Hence the mere form of a law, which limits the matter, must at the same time be a ground for adding this matter to the will …' (*CpV* 5:34). Note also that Theorem III is concerned with how an agent can 'think of his maxims as practical universal laws', and that the following Remark illustrates with the example of a deposit whose owner has died; the issue is whether this maxim has the form of lawgiving.

complain of an agent who acts on those reasons. Here, the reasons stated in the agent's maxim do justify the action: they can be regarded by anyone as good reasons for that agent to perform the action. The fact that a maxim is in this way consistent with the form of a practical law can be an agent's reasons for action, again in an extended sense, when the agent recognizes the permissibility of the maxim as a limiting condition on its adoption – that is, when the agent would have rejected the maxim had the proposed reasons not met this standard of permissibility.[15]

Let me now raise two questions: first, can the form of a practical law provide a ground of choice by itself? Second, can Kant claim (as he wishes to) that only a practical law can provide a ground of choice through its form? The first question arises because the form of a law by itself cannot point the will in any specific direction. The higher-order commitment to act only from principles that have the form of a law is a fundamental practical orientation that determines which substantive considerations an agent treats as good reasons. But the substantive considerations that provide the matter of choice specify what is to be done. The higher order commitment cannot lead to action unless it is applied to some concrete circumstances. Nonetheless, Kant can hold that the lawgiving form of a maxim can provide a ground of choice by itself in the case of requirements and prohibitions (though not permissions): the fact that a maxim or substantive practical principle has the form of a law (the form of a requirement) is a sufficient reason to adopt it and the fact that it does not is a sufficient reason to abandon it. The matter provides direction for the will, but the ground of choice is the fact that the matter has (or lacks) the form of a practical law.

Regarding the second question, I have suggested that the principle of happiness is a kind of formal principle because it captures the form of choice on a material practical principle. The fact that an object of choice offers satisfaction or answers to one's desires and dispositions sounds like a reason of a very general sort – i.e. something that can be said on behalf of that choice; or at least it seems to be the form of a (kind of) reason. Someone who adopts and acts on a material practical principle is taking the fact that the object of choice answers to his independently given desires and dispositions as his reason for action at the most general level. This may seem a case where the form of a principle – here the form of a material practical principle – provides a ground of choice (a motivating reason, even if not a sufficient justifying

[15] The idea that the motive of duty, thus the permissibility of a maxim, can function as a 'limiting condition' on choice was introduced by Barbara Herman; see *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993), chapter 1.

reason).[16] Hence my question: does a practical law determine choice in virtue of its form in a way that a material practical principle does not? For the following reason I would say, 'yes'. A material practical principle (such as the principle of seeking all the comforts of life, or the principle of working and saving in one's youth to provide for one's old age) gets its normative force for an agent through its relation to the agent's interests – i.e. from the fact that its object answers to an agent's antecedent desires and dispositions. Since the reasons given by material practical principles have subjective conditions, any specification of those reasons must be completed by a reference to those interests. Thus one cannot say that the form of a principle by itself provides a reason for choice. By contrast, in the case of a practical law no reference to the interests of the agent is needed to substantiate the reason for adopting the principle. The ground of choice is the fact that the matter of the principle has the form of a law.[17]

Let me close this section with another way of characterizing the form of a practical law that fits well with the three familiar points just outlined. Laws of any kind state necessary connections that hold between some ground and some consequence. One might think of the form of a law as the structural relationship that holds between ground and consequence, the formal relationship characterized by the notion of sufficient reason.[18] Natural (causal)

[16] To illustrate, take the example of someone who seeks all the comforts of life and never willingly forgoes any comfort. The substantive considerations that he treats as good reasons include the appealing features of good meals, well-tailored clothes, well-designed furniture and cars, the avoidance of stress and so on. He need not consciously have adopted the principle of seeking all comforts, but this principle unifies and explains his choices. His substantive reason for choice in a given instance will be an appealing feature of some object – of this meal, this article of clothing, this relaxing activity. But at the most abstract level, his reasons derive from the fact that this overall end (or this method of determining what count as reasons) appeals to him, and that these specific objects of choice fit together to further this end. The fact that he finds these objects of choice agreeable, given his desires and tastes, is the common structural feature (form) of his choices, and it seems to me that with a bit of abstraction, he can cite this fact as a reason for choice. Is it then fair to say that this formal feature is the ground of choice and that a material practical principle can determine choice through its form?

[17] I have been discussing substantive practical laws and have argued that a substantive practical law can determine choice through its form alone, while a material practical principle cannot. Can the Formula of Universal Law (FUL), in contrast to the principle of happiness (PH), determine choice through its form? Both FUL and PH can be the basis of higher-order commitments (basic practical orientations) to treat certain substantive considerations as reasons in certain ways, and neither can lead to action without application to some concrete circumstances. In the case of PH, choice requires further input from an agent's interests (so it cannot determine choice beyond being the basis of the aforementioned commitment). But as I argue in the next section, FUL can determine the will in virtue of its form: the formal principle or morality is a ground of choice in virtue of the fact that it is the formal principle of pure (i.e. free) volition, or is the form of autonomous willing.

[18] Here I draw on Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge and New York: Cambridge University Press, 1996), 102, and *The Sources of Normativity* (Cambridge and New York: Cambridge University Press, 1996), 107–8. I discuss this idea further in *Agency and Autonomy*, 133–4.

laws are concerned with necessary connections that obtain between antecedent conditions and some effect when those conditions are sufficient to produce the effect in all cases. A natural law states that whenever certain conditions obtain, a specified event follows as their effect. Likewise practical laws are concerned with the relationship between considerations that provide practical reasons and choices when those considerations constitute sufficient justification for the choice. A practical law states, in effect, that certain considerations or facts about the circumstances (that an individual needs life-saving aid that one can provide, that one has promised to repay) are in every case sufficient justification for some action (providing the aid, repaying).[19] This structural relationship – necessary connection between ground and consequence, or sufficient reason – is common to both natural laws and practical laws. It is also a relationship that in a relatively clear sense 'inheres in the matter' of a maxim or principle. A maxim has the form of a law when a certain relationship holds between the components of the maxim – when the considerations cited by the maxim are indeed a sufficient justification for the action. And this relation of sufficient reason will obtain in some practical principles but not others. Finally, practical laws will have both the objective universal validity and subjective universal validity to which Engstrom points: a practical law holds that a certain consideration is a sufficient justification for any agent in the situation covered by the principle; and it governs how all subjects should think about the issue – it is a reason that anyone can agree to or regard as sufficient.

III. FORMAL PRINCIPLES (THE FORMAL PRINCIPLE OF VOLITION)

In this section I discuss one way of understanding Kant's conception of a 'formal principle'. As far as I know, Kant does not explicitly characterize formal principles in the way that I shall propose (though there is some textual evidence for attributing this view to him). But this account offers a natural explanation of why Kant thought that the fundamental principles

[19] Again, the structural relationship between the considerations that provide reasons and the choice, and thus the notion of sufficient reason, are slightly different in requirements, prohibitions and permissions. In the case of a requirement, the relevant considerations – the fact that an agent needs life-saving aid that one is in a position to provide – are in every case decisive and overriding reason for the action (providing the aid). In the case of permissions, the relevant considerations provide sufficient justification. When a maxim of self-interest is universalizable, the agent's reasons (e.g. the fact that an action will further my end) are sufficient justification for the action which anyone must recognize as such (permitting but not requiring adoption of the maxim). In the case of a prohibition, they never provide justification: the fact that dishonesty will further my self-interest never justifies dishonesty.

of certain domains of rational activity are 'formal', and why formal principles, so understood, are uniquely suited to apply with normative necessity to some domain of rational activity. We may think of the *form* of some rational activity or object of cognition as the constitutive or defining features of an activity or entity of that kind – the features that an activity or entity must possess to count as an instance of that kind. The form of some rational activity or object of cognition will be associated with a *formal principle* that is constitutive of that rational activity or object of cognition. The formal principle of some rational activity would be the guiding internal or constitutive norm that a subject must follow in order to engage in that activity. By specifying the form of that activity, it provides a norm that anyone engaged in that activity must satisfy and that in some sense does guide any instance of the activity (even if defectively).[20] The formal principle of a kind of object of cognition would be a norm that sets necessary conditions on thought about such objects. The form of a law would be the defining features that a principle must have in order to qualify as a practical law. Since practical laws regulate volition, or more specifically pure volition, the associated formal principle will be a principle of pure volition – the internal guiding norm of pure willing, or the principle that one must follow in order to guide one's will by reason alone (ironic aside: if that is important to you …).

[20] Let me mention a problem that I am unable to deal with adequately here: if a formal principle is constitutive of an activity, it must be implicated in all instances of the activity. The problem is how to characterize mistakes or defective instances of an activity. As an example, if syntactical rules are constitutive linguistic rules, they are rules that one must follow in order to produce speech. This holds for grammatically well-formed statements as well as those containing grammatical errors. Syntactical rules in some sense guide even ungrammatical speech – if they did not, the utterance would not count as an instance of speech, in which case it could not be the kind of activity that contains a grammatical error or is subject to assessment by grammatical rules. Likewise, someone who makes a mistake in addition in some sense follows the rules of addition, while at the same time violating them. In such cases, we want to say that the subject is guided by or implicitly follows the constitutive rules, even in instances where she violates them. The question of how defective instances of an activity (those that fail to conform to the rules) are guided by constitutive rules is a problem that this account of formal rules needs to address.

Barbara Herman makes a promising suggestion about this issue. One might think that all rational or cognitive activity is guided by a *representation* of various formal or constitutive principles, and that in defective activity, the relevant principle is *mis*represented. Herman specifically develops this idea to explain how Kant can hold both that the moral law is a law of freedom and that morally bad action is freely willed. Her idea is that all free action is derived from a representation of the formal principle of the will (the moral law), but that in bad or faulty action this principle is misrepresented. Bad action is free because derived from a representation of the principle of the will, but faulty because that principle is misrepresented. See Herman, *Moral Literacy* (Cambridge, MA: Harvard University Press, 2007), 171–2, 245–6.

When formal principles are understood as constitutive principles, certain features stand out. Formal principles are, first, *enabling principles* or conditions. By defining what counts as a certain kind of activity they make it possible to engage in that activity. Second, by constituting or defining an activity, they are *internal* to the activity. Since a formal principle specifies the form of an activity and serves as a norm for any subject engaged in the activity, it is a principle that arises from the nature of that activity – as one might say, a principle that the relevant rational activity 'gives to itself'. Third, and this is the point to which I am leading, formal principles, understood as constitutive principles, *apply with necessity* to anyone engaged in the activity, and their constitutive role explains their necessity. Since they specify how one must proceed in order to engage in the activity (for what one does to count as an instance of the activity), they are not coherently rejected by anyone engaged in that activity. It is for this reason that formal principles are uniquely suited to govern an activity with unconditional necessity. It is somewhat beyond my competence to pursue this idea elsewhere in Kant's work, but I presume that this understanding of a formal principle and of the connection between formal principles and necessity can be seen throughout his system. The principles of formal logic specify the form of thinking in general, and can be viewed as enabling conditions of thought. Their necessity lies in the fact that unless one is governed by these principles, one is not thinking (and, one should add, thinking is a necessary activity for rational creatures like us). The pure categories and principles of the understanding (principles of transcendental logic) specify the form of and enable thinking about objects of experience, thus necessarily govern all thought about objects of experience. And so on.[21]

A formal principle involves some abstraction from content: the form is what remains when one sets aside those features that differentiate one instance of an activity from any other (e.g. that distinguish one principle of volition from another). But focusing on abstraction from specific differences in content only characterizes formal principles negatively. It does not bring out the foundational role of such principles and the point of introducing them. For example, simply pointing to the fact that a formal

[21] For the most part, the formal or constitutive principles of various domains of cognitive activity that one finds in Kant's thought are synthetic a priori principles, such as the Categorical Imperative or the a priori principles of the understanding. But the principles or formal logic are analytic. Can formal or constitutive principles include both analytic principles and synthetic a priori principles? (Can the analytic status of formal logic be due to the fact that it sets constraints on thinking about any kinds of object whatsoever?) This is one of several issues that the understanding of formal principles proposed here needs to address in greater depth. I am grateful to Pierre Keller for raising this question.

principle abstracts from ends (or does not get its authority from any interest in an end) provides almost no insight as to why Kant thinks that fundamental principles must be formal and where their normative authority comes from. This is one reason why the argument for Theorem III – that a practical law provides a ground of choice through its form – appears too quick. The arguments here and elsewhere explain why material practical principles, understood as principles whose normative force presupposes an empirically given interest, lack the normative necessity required of a law. But they do not give a clear positive argument as to why formal principles and only formal principles (or why principles that determine choice through the form of a law and only such principles) do have this necessity. Understanding formal principles as constitutive principles fills this lacuna by establishing a connection between form and necessity. For that reason, when I refer to formal principles, such as the formal principle of morality, I mean to draw attention to their role as constitutive, and therefore necessary, principles in some domain of rational activity, and not to the fact that they involve a certain abstraction from content.

To see what this conception of a formal principle might do for Kant, it may help to consider how the connection between form and necessity figures in the more ambitious argument of the *Groundwork*. The overall task of the *Groundwork* is to show (through a 'deduction') that the special authority that ordinary moral thought assigns to moral requirements is genuine – as Kant puts it, to show how the necessity thought in a categorical imperative is possible (*G* 4:417). Kant's resolution of this problem, in a word, is to argue that the moral law is the constitutive principle of a necessary self-conception. First, as rational agents we necessarily think of ourselves as free, various features of our capacity for theoretical reason confirm ascribing freedom to ourselves, and we identify with our capacity for free agency.[22] And second, the moral law is the formal or constitutive principle of free volition. In other words, he identifies the formal principle of morality with the formal principle of our agency. Kant launches his argument by deriving a statement of the Categorical Imperative – the Formula of Universal Law (FUL) – from the concept of a categorical imperative. This derivation presents the FUL as the formal principle of conformity to moral requirement – the principle one must follow in order

[22] The claim that we necessarily act under the Idea of Freedom comes, of course, at *G* 4:448. That our theoretical capacities confirm ascribing freedom to ourselves (and that we necessarily act under the Idea of Freedom) comes at *G* 4:451–3. The claim that we identify with our capacities for free agency is seen in the references to the 'proper self' at *G* 4:457 and 461. I believe that this idea is behind the claim at *G* 4:453, that the intelligible world is the ground of the sensible world.

to conform to moral law. While this principle tells us what morality demands of us, understanding the principle as constitutive of conformity to moral requirement does not speak to the question of its authority. That is the function of the sequence of formulas that sets up the arguments of Section III. The introduction of autonomy (at *G* 4:431) shows that the form of a law is also the form of lawgiving – that is to say, the formal principle of conformity to universal law is the principle that is constitutive of the autonomy, or lawgiving capacity, of the will; it is the principle that you must follow to give law through your will. And the argument in *Groundwork* III that autonomy specifies the positive concept of freedom leads to the idea that the form of lawgiving is the general form of free volition. (Free agency is the capacity to act from principles that one gives to oneself through the will's own principle, independently of determination by external conditions.)[23] The formal principle of morality, in other words, is at the same time the constitutive principle of free volition: it is the principle one must follow to exercise one's free agency and the capacity to follow this principle makes one a free agent. On the assumption that volition is free volition, the sequence of formulas permits Kant to claim that the FUL 'contains merely the form of volition as such' (*G* 4:444). If it is the capacity to follow this principle that makes us agents, we cannot coherently reject its authority and continue to think of ourselves as agents. Thus, given the claim that we necessarily think of ourselves as free and identify with our agency, the necessity of morality is vindicated by showing that the fundamental (formal) principle of morality is the formal principle of our agency.

Although the main reasons for thinking of formal principles as constitutive principles are philosophical, there is textual evidence in the second *Critique* that Kant does understand formal principles in this way. I'll cite two passages. The first is the Remark following the Definition of a practical law, in which Kant says that a practical law 'determines the will as will'. In this paragraph Kant draws a contrast between, on the one hand, principles that determine the 'conditions of causality of a rational being as an efficient cause merely with respect to the effect' and, on the other hand, principles 'that determine only the will, whether or not it is sufficient for the effect', 'principles that determine the will as will even before I ask whether I have the ability required for a desired effect or what I am to do in order to produce it', or principles that have to do 'only with [one's] will, regardless of whether the purposes the human being may have can thereby be attained'.

[23] For further discussion, see Reath, *Agency and Autonomy*, 154–5 and Reath, 'Autonomy, Practical Law, and Taking One's Choices to Be Good', 130–4.

In the latter case (the case of a law), the principle determines 'the mere volition' (*das bloße Wollen*) and 'refer[s] only to the will, without regard to what is attained by its causality' (5:20–1). I am not sure that I fully understand certain details of this contrast, but this much is clear.[24] Practical laws do not prescribe means to desired ends or give directives that are conditional on subjective interests, but rather determine what it is necessary for any rational agent to will. That they 'determine the will as will' suggests that practical laws are internal to (constitutive of) purely rational volition – the kind of principle that one must follow in order to will from reason alone. They apply in virtue of defining features of pure volition that hold for any agent with that capacity. The higher-order principle of conformity to practical law would be the principle that defines the capacity for purely rational volition, and therefore is not coherently rejected by any agent with this capacity. The formal principle of morality, in other words, is the formal principle of pure volition.

Second, in Remark I following Theorem IV, Kant writes: 'the necessity that the law [of the pure will] expresses, since it is not to be a natural necessity, can therefore consist only in the formal conditions of the possibility of law as such' (*CpV* 5:34). This Remark follows the claim that the moral law is the principle of the autonomy of the will, but it makes the point that the moral law states the constitutive or defining features that a principle must have to qualify as a practical law. That makes it the constitutive principle of conformity to moral requirement, which we know is the formal principle of pure volition.

On one influential reading of Kant's Formula of Humanity, it can also be understood as a 'formal principle' in the sense explained here: it is the formal principle of finding sufficient reason to adopt some end and taking its achievement to be worthwhile or objectively good. According to this reading of the value of humanity or rational nature, the exercise of rational choice confers objective value on its objects. Rational choice is directed at ends taken to be objectively good, and since the value of these ends comes from the fact that they are rationally chosen by some person, rational nature is implicitly valued in all rational choice. One component of this view is that ends are adopted for the sake of some person (often myself, but possibly

[24] Kant says that laws 'determine only the will, whether or not it is sufficient to the effect' and '… even before I ask whether I have the ability required for a desired effect'. See also, *CpV* 5:45, 66. But why should practical laws set aside one's ability to achieve some end? After all, pursuing an end that one knows to be beyond one's capacity involves a kind of irrationality. Presumably his idea is that laws concern what we are to will for its own sake in some absolute sense, and what we are to will in that sense should no more depend on our practical abilities than on our subjective interests.

others as well), and the value of the person confers value on the achievement of the end. What makes the achievement of some (rationally chosen) end worthwhile is that it is adopted for the sake of an end in itself, and ends in themselves are persons. For example, in finding sufficient reason to pursue an end in the fact that it interests me and, having made it my end, thinking that its achievement matters, I value myself as an end in itself. So it is by adopting an end for the sake of some person and valuing that person (in this case, myself) as an end in itself that one finds the achievement of the end to be objectively worthwhile. And whenever one takes the achievement of an end to matter in this way, it is adopted for the sake of some person who is valued as an end in itself. In this way, one can argue that valuing humanity as an end in itself is the formal principle of finding objective value in the achievement of rationally chosen ends.[25]

In this section, I have claimed that since formal principles are constitutive of some domain of rational activity, they apply with normative necessity to anyone engaged in that activity. Whether such principles are unconditionally necessary, of course, depends on whether that form of rational activity is necessary, or one to which we are unavoidably committed. Here it is important to note that early in the first chapter of the second *Critique*, Kant asserts up front that the formal principle of morality is the formal principle of *pure* volition. The *Groundwork*, by contrast, aims at the more ambitious claim that the formal principle of morality is the formal principle of *volition as such* – that is, that it is constitutive of the brand of free volition that underlies and is exercised in *anything* recognizable as rational choice. If the *Groundwork* argument succeeds, it would show that the basic principle of morality is not coherently rejected by anyone in the business of choosing (which we are in most of the time). But it is hard to see how this argument could succeed. To make it work, one needs to assume that a strong and controversial conception of transcendental freedom underlies all forms of rational agency; but there are certainly coherent conceptions of agency that do not meet this standard (and one may indeed be true of us). In the second *Critique*, Kant pulls back from this claim, perhaps because he came to see the difficulties of showing that a rational agent necessarily acts under the idea of *transcendental* freedom, but certainly because of the obstacles to ascribing transcendental freedom to ourselves on grounds that are independent of moral consciousness. He takes the more modest line that the

[25] For examples of this reading, see Korsgaard, 'Kant's Formula of Humanity', in *Creating the Kingdom of Ends*, 114–24, and *The Sources of Normativity*, 120–6; and Wood, *Kant's Ethical Thought* (Cambridge and New York: Cambridge University Press, 1999), 125–32.

formal principle of morality is the formal principle of pure volition – an assertion with which one could hardly take issue.[26] The question to ask here is whether we must be in the business of pure willing. The fact of reason asserts, among other things, that we are in this business. The fact of reason, as I understand it, is our recognition of the authority of moral concerns in everyday moral thought, judgment and feeling, and the subsequent awareness of our capacity to act from moral concerns, even in the face of powerful countervailing reasons.[27] The fundamental (formal) principle of pure volition presents itself to us, synthetically, with an authority 'not to be outweighed by any sensible conditions and indeed quite independent of them' (*CpV* 5:29–30) that, on due reflection, we cannot reject without loss, and our recognition of its authority reveals the capacity for purely rational volition. We do, indeed, take the standards of the pure will to apply to us with normative necessity, though a deduction that establishes this necessity is neither possible nor required (*CpV* 5:47).[28]

IV. INTERPRETATIONS

In this section, I'll use some of the points made above to reconstruct the arguments of Theorems I and III and Problem I.

(a) Theorems I and III

Theorems I and III claim, first, that no material practical principle can furnish a practical law, and second, that if a principle is to serve as a law, it must provide a ground of choice through the fact that it has the form of a

[26] The pullback in the second *Critique* is not complete. Throughout Chapter I, Kant is explicit that the moral law is the formal principle of pure volition, and I take that to differ from asserting that it is the formal principle of volition as such. But at *CpV* 5:32 he does appear to make the stronger claim:

> Now this principle of morality, just on account of the universality of the lawgiving that makes it the formal supreme determining ground of the will regardless of all subjective differences, is declared by reason to be at the same time *a law for all rational beings insofar as they have a will*, that is, the ability to determine their causality by the representation of rules, hence insofar as they are capable of actions in accordance with principles and consequently also in accordance with a priori practical principles. [italics added]

Here Kant claims that the moral law is 'a law for all rational beings insofar as they have a will' and goes on to define the will in the standard way. But an agent whose principles were based on desire and contingent interests would still have a will according to this conception. If so, he cannot claim that the moral law is the formal principle of volition as such.

[27] I take this understanding of the fact of reason from Rawls. See 'Kant: Lecture X', in *Lectures on the History of Moral Philosophy*, especially 255, 260.

[28] For further discussion of the fact of reason and the changes in argument between *Groundwork* III and the second *Critique*, see the essays by Pauline Kleingeld and Jens Timmermann in this volume.

law. These Theorems are the basis of Kant's claim that the fundamental principle of morality cannot be a material practical principle, and that only a formal practical principle can apply with the necessity implicit in the ordinary concept of duty (*CpV* 5:41, 64–5). To substantiate this claim, Kant must first show that there is an exhaustive distinction between practical principles that provide a ground of choice through their matter and those that do so through their having the form of a law. That distinction will be exhaustive only if Kant can make the case that his category of material practical principles is quite inclusive – that is, that all principles other than those that are formal in his sense 'presuppose an *object* (matter) of the faculty of desire as the determining ground of the will' (*CpV* 5:21). Second, he needs a positive account of why only formal principles (or principles that provide reasons through their having the form of a law) are normatively necessary.

A material practical principle gets its normative force from an antecedently given interest in some object, broadly construed to include a wide range of ends and values. Kant explains why all such principles are empirical and unable to ground laws in the following passage:

> For, the determining ground of choice is then the representation of an object and that relation of the representation to the subject by which the faculty of desire is determined to realize the object. Such a relation to the subject, however, is called *pleasure* in the reality of an object. This would have to be presupposed as a condition of the possibility of the determination of choice … [I]n such a case the determining ground of choice must always be empirical, and so too must be the practical material principle that presupposes it as a condition. (*CpV* 5:21)

The passage describes one way in which a practical principle can provide reasons – the form of choice, as noted in section II above, captured by the 'principle of happiness'. But it also describes the method of moral philosophy employed by theorists who 'begin from a concept of the good in order to derive from it laws of the will' (*CpV* 5:63). Kant rejects this method in Chapter II because a theory that proceeds in this way cannot produce genuine practical laws, for the reasons given here. Such theorists present an object of choice as intrinsically good or choiceworthy – e.g. Epicurean happiness, agreement with Hutchesonian moral feeling, Wolffian perfection, and so forth – and from this object derive rules of conduct (say, to act in ways conducive to or in agreement with this object). Here Kant claims that the reasons for accepting such a principle come from (first) the representation of the object and (second) the fact that the representation of the object has a certain relation to an agent through which the agent is

moved to take an interest in the object, a relation that Kant terms 'pleasure in the reality of an object'. I take that to mean that features of the object provide reasons because when represented to an agent, they produce interest – they elicit responses or interest or answer to antecedent dispositions that agents are assumed to have. These responses of interest are 'based on the receptivity of the subject' and belong to 'feeling' (*CpV* 5:22). They include any interests that an agent may have, and indeed could share with most other members of the species, as a matter of empirical fact, but that are not essential to rational capacity. In this way, any such principles are 'subject to an empirical condition' and thus not laws (*CpV* 5:27).

This argument shows that no principle of pursuing an obviously desire-based end can be a practical law. More importantly, it shows that empiricist moral theories, such as an Epicurean principle of happiness or Hutcheson's or Hume's moral sense theory, cannot ground the necessity that is part of the ordinary concept of duty (not that this fact would disturb them). But Kant needs to address this objection: why should one think that all principles based on an end or substantive value are 'subject to an empirical condition' in this way? Some theorists will resist assimilating their substantive first principles to Kant's understanding of a 'material practical principle' (where the reason-giving force presupposes rationally contingent responses of interests). We can readily imagine principles based on ends or substantive values that are taken to be immediately or intrinsically good and to make claims on the will of any agent with adequate powers of appreciation – for example, Wolffian perfection, Cumberland's greatest happiness of all rationals or Clarke's eternal fitnesses. If such ends and intrinsic values make necessary claims on the will, as these theorists might hold, they would be the basis of substantive practical principles that hold for any rational agent, and thus hold as laws.

Kant's reply to this objection will have different facets. First, he can counter that many of the objects proposed as the basis of fundamental principles are not in fact necessary objects of volition; they can be rejected without obvious irrationality.[29] A second and more direct response is that any theory that 'begins with the concept of the good' does employ the method that he rejects, and that the representation of an end or value – as immediately or intrinsically good – is not enough to establish a necessary claim on the will.[30] How would such a theory establish that some end

[29] For discussion of this point, see Thomas E. Hill, Jr, *Dignity and Practical Reason* (Ithaca: Cornell University Press, 1991), 128–31.

[30] Here again I have learned from Engstrom, 'Introduction', xxxvi–xxxix.

(e.g. perfection) or a set of substantive values (e.g. relations between agents that ground the moral fitness of certain acts) is intrinsically choiceworthy and can ground practical principles? Presumably, all that theory can do is to present the end or value for our consideration – as good. If we respond to the representation of the end or value with interest, the associated principle will be reason-giving. But no necessary claim has been established unless it can be shown that one ought to respond with interest (regardless of antecedent dispositions), and a theory that begins by presenting what is good in this way lacks the resources to establish this claim. This is the methodological point of Chapter II: a theory that begins by proposing some end as good cannot establish true practical laws, because the representation of the end can provide reasons only in relation to antecedently given dispositions and interests that are not part of rational capacity, and are in that sense subjective.[31] To show that an end is necessary, one must show that it is required by a principle to which any rational agent is committed; that is, one must begin from a practical law.

Along these lines one can refine Kant's argument that any principle based on the prior presentation of an end or value as intrinsically good is 'subject to an empirical condition'. Kant's positive account of why the necessity of a practical law is conferred by its form, very briefly, has to do with constitutive role of the moral law in pure willing. When one takes the lawgiving form of one's maxims as a sufficient reason for action, one acts from the formal principle of morality, and that is what it is to exercise one's pure will. The formal principle of morality is the constitutive principle of pure volition, thus necessary for anyone engaged in that activity.

(b) Problem I

Problems I and II argue that 'freedom and unconditional practical law reciprocally imply each other' (*CpV* 5:29). I understand Kant to be claiming that the formal principle of pure practical reason is the formal (constitutive) principle of transcendentally free agency and that the capacity to act from this principle confers transcendental freedom. Let's see how far Kant can get in solving 'Problem I', which reads as follows:

[31] See *CpV* 5: 63: 'Suppose that we wanted to begin with the concept of the good in order to derive from it laws of the will … Now, since this concept had no practical a priori law for its standard, the criterion of good and evil could be placed in nothing other than the agreement of the object with our feeling of pleasure and displeasure.'

Supposing that the mere lawgiving form of maxims is the only sufficient determining ground of the will: to find the constitution of a will that is determinable by it alone. (*CpV* 5:28)

To begin with a question, can this mystery will take its reasons from considerations other than the fact that its maxims have the form of a law? Certainly the lawgiving form of its maxims is the only sufficient *justifying* ground of choice. But that point aside, could this volitional capacity choose to act from some material principle? The way Kant poses the problem, clearly not: this will is 'determinable alone' through the lawgiving form of its maxims and no ground of choice other than the 'mere universal lawgiving form can serve as a law for it' (*CpV* 5:28, 29). It is then a pure will – a volitional capacity to act exclusively on purely rational considerations; and the FUL is its formal principle. Given this stipulation, what can we conclude about such a will?

We know that a will that bases its reasons exclusively on the lawgiving form of its maxims will not act from practical principles whose normative force depends upon empirically given interests. Interests of this sort arise through the experience of finding an object or activity agreeable in some way, as a result of an individual's aptitudes and susceptibilities. Such interests 'belong among appearances' and can be given naturalistic causal explanations in terms of psychological principles. But the stipulation that the mystery will acts independently of such interests does not secure the independence from natural causation that Kant is after without some argument. Kant's conception of rational agency commits him to the view that even interests of this sort influence choice in some sense independently of natural causal laws. The interests from which a material practical principle gets its normative force are brought to practical consciousness as rational dispositions – as acceptance of principles and values, or the tendency to view certain considerations as reasons, and so on – and they govern choice normatively, not causally. Let me explain.

Suppose that one of my ends is to increase my knowledge of nature and that I am attached to this end because I find it stimulating to learn about nature. It agrees with me. So I am imagining that this end is the basis of a material practical principle. There is a naturalistic causal story to be told about how this interest arises that will cite my aptitudes and dispositions, various experiences that I have had, and so on. However, although there is a causal explanation of the origin of this interest, it is taken up into practical consciousness in a certain form (and I use that term intentionally) – e.g. as a disposition to judge that increasing my understanding of nature is a worthwhile activity and to see reasons for action in opportunities to add to my

knowledge. Further, this interest influences my choices through my self-conscious application of various normative considerations and rational principles – for example, through judgments that I have good reason to pursue some opportunity to increase my understanding of nature, given that it is one of my ends. And as a rational agent, my judgments never 'consciously receive direction from any other quarter' (*G* 4:448). They are never directed by anything other than my own grasp of various reasons and my application of relevant rational norms to my circumstances. There is then a substantive sense in which rational choice on a material practical principle is independent of determination by causal laws: since it is governed by judgments about reasons and normative principles, it is governed other than by causal laws. Though the interest from which the practical principle gets its normative force has a causal explanation, choice on that principle is still normatively, rather than causally governed. It thus displays a not uninteresting form of freedom.

In the case of my decent though not pure willing, the rationality of judgment and choice blocks naturalistic causal explanation of the choice, but not of the interests on which the reasons for choice are based. But pure volition is not based on any such interests and takes its reasons only from the fact that a certain substantive principle instantiates the form of a law. What Kant seems to be getting at in Problem I is that for pure volition, all causal explanation is blocked, because the reasons for choice are not based on any empirically given interest (something given in appearances) for which a causal explanation can be given. When the ground of choice is the legislative form of a maxim, 'the determining ground of the will is distinct from all determining grounds of events in nature in accordance with the law of causality, because in their case the determining grounds must themselves be appearances' (*CpV* 5:28–9). So now suppose that I bring my activity under the duty of natural perfection and increase my understanding of nature out of a commitment to make good use of my natural faculties. Since my reasons for choice are now based only on the lawgiving form of my maxim, they are not based on any interest in the matter of choice that admits of the kind of causal explanation sketched above. To put the point another way, since I act on what is necessarily a reason for any rational agent (the lawgiving form of my maxim), my reasons are not taken from interests due to influences specifically on me. In this way the activity of a pure will (of an agent who exercises the capacity for purely rational volition) 'must be thought as *altogether* independent of the natural law of appearances, namely the law of causality' (*CpV* 5:29; italics added). It is governed by a fundamental normative principle that operates independently of any interests that

admit of naturalistic causal explanation. It seems to follow that the pure will satisfies the definition of transcendental freedom. It is governed by its own constitutive principle – a principle that it gives to itself in the sense that it arises from the nature of pure volition. And the capacity to follow this principle makes such a will 'altogether independent' of natural causality, since its reasons are not based on any interest that can be given a naturalistic causal explanation.

Though I cannot say for sure, this argument seems intriguingly close to succeeding.

CHAPTER 3

Moral consciousness and the 'fact of reason'

Pauline Kleingeld

1. INTRODUCTION

At the very heart of the argument of the *Critique of Practical Reason*, one finds Kant's puzzling discussion of the 'fact of reason'. Kant introduces the notion in the course of arguing that pure reason is practical, which is the main task of the first chapter of the Analytic. Having claimed that 'we' have a 'consciousness of the moral law', and that this leads us to the concept of the freedom of the will (*CpV* 5:29–30), Kant argues that the consciousness of the moral law can be called a 'fact of reason'. This passage is found in the brief remark between the two most important conclusions of the second *Critique*. It is located between the formulation of the 'Fundamental Law of Pure Practical Reason', 'so act that the maxim of your will could always hold at the same time as a principle in a giving of universal law' (*CpV* 5:30), and the 'Conclusion' that 'Pure reason is practical of itself alone and provides (the human being) with a universal law which we call the *moral law*' (*CpV* 5:31). Even if the exact role of the notion of a 'fact of reason' is not immediately clear, there is no doubt that the argument in which it plays a role is central to Kant's moral theory.

Ever since it saw the light of day, however, Kant's argument regarding the fact of reason has met with strong criticism, although the critics disagree fundamentally as to what exactly is wrong with it. Not surprisingly, commentators also have wildly divergent views on the relationship between the second *Critique* and the third section of the *Groundwork of the Metaphysics of Morals*, but most seem to hold that although the account in the *Groundwork* is very problematic, the argument in the second *Critique* is even worse.

Kant's assertions certainly raise many difficult questions. When he introduces the 'fact of reason', he refers to an alleged consciousness of the fundamental law of pure practical reason, which he also calls consciousness

of the moral law. Does this mean he simply regards it as a fact that there is consciousness of the categorical imperative? This might seem odd, because Kant regards himself as the first person to formulate this moral principle clearly, so he cannot be assuming that the articulate consciousness of it is widespread. Or does Kant mean the fact that there is consciousness of moral obligation in general, even if its specific content is undetermined, vague or misconceived? But then, what is this exactly, and, at least as important: how does Kant know? Is his claim an empirical thesis that can be tested? If it is, does it rely on empirical (sociological or psychological) assumptions about the universal acknowledgment of moral obligation, and, if so, could Kant's claim still be defended in an era in which it is recognized that there is widespread disagreement about moral demands? Moreover: even if we granted, for the sake of argument, that there is a universally shared consciousness of moral obligation, how can this help to justify the assumption of freedom? After all, people's belief that they have moral obligations could be illusory. If, on the other hand, Kant should not be understood as making an empirical claim, in what sense is he then speaking of a 'fact', and which fact is he talking about?

There is, then, an entire nest of related questions that emerge on the basis of Kant's equation of a 'consciousness of the moral law' with a 'fact of reason'. Kant further adds to the difficulties by designating as a 'fact of reason' not only the alleged consciousness of this law, but also 'the moral law' itself (*CpV* 5:47, 91), and 'autonomy in the principle of morality' (*CpV* 5:42), and by claiming that the fact of reason is 'identical with consciousness of freedom of the will' (*CpV* 5:42).

In this essay, I attempt to shed some new light on the meaning and the importance of the 'fact of reason' in the second *Critique*. I clarify the meaning of the term '*Factum*' itself, situate the maligned passages within their argumentative context and argue that Kant's argument can be given a consistent reading on the basis of which the main questions and criticisms can be answered. While this does not amount to a full validation of Kant's justification of freedom and morality, it should lay the foundation for one, and it should prompt a re-evaluation of the strength of this argument as compared to the argument found in the third section of the *Groundwork*.

I first situate the occurrence of the terminology of a *Factum der Vernunft* in its argumentative context and provide a first interpretation. I then look at the meaning of the term more closely and clarify the argument further in a second round, also discussing the main standard objections.

II. KANT'S INTRODUCTION OF THE 'FACT OF REASON'

It is crucial to see that the passage in which Kant introduces the 'fact of reason' continues his discussion of the consciousness of the moral law in the remark at the end of section 6. This link between the remarks of sections 6 and 7 often goes unnoticed, and that makes it harder to make proper sense of Kant's talk of the *Factum*.

In the first six sections of the *Critique of Practical Reason*, Kant establishes the reciprocal relationship between freedom of the will and unconditional practical law: if a will is determined by the principle of the 'mere lawgiving form of maxims' (instead of by their content) this will is free, and if a will is free, only the lawgiving form of the maxim can be its sufficient determining ground. Hence, Kant says: 'Freedom and unconditional practical law reciprocally refer to one another' (*CpV* 5:29). This reciprocity thesis, however, is a conceptual thesis and does not yet show that we have grounds for assuming that our will is *actually* free or that we are *indeed* bound by the law of pure practical reason.

In the Remark in section 6, Kant moves from the conceptual level to the question of whether freedom or the moral law comes first in the order of cognition. He asks 'from whence our *cognitive awareness*[1] of the unconditionally practical *starts*, whether from freedom or from the practical law' (*CpV* 5:29; Kant's emphasis).

Kant's brief answer is that the concept of freedom is not primary. This is, first, because we do not have immediate consciousness of freedom. Our initial concept of freedom is merely negative, Kant asserts, viz., independence from determination by the causality of nature. The positive concept of freedom is mediate (as Kant claims to have shown through his earlier discussions in sections 5 and 6). Second, we cannot infer the freedom of the will on the basis of experience. Experience, after all, presents us with the opposite of freedom, namely, determination by mechanical natural causes (*CpV* 5:29).

Therefore, Kant says, our consciousness of the moral law (or more precisely, as Rawls has pointed out, our consciousness of the moral law as authoritative)[2] is what comes first: consciousness of moral obligation discloses our freedom. It is 'the *moral law*, of which we become immediately

[1] *Erkenntnis*, see *CrV* A320/B376–7.
[2] John Rawls, *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, MA: Harvard University Press, 2000), 255.

conscious (as soon as we construct maxims of the will for ourselves), which *first* presents itself to us' and which 'leads to the concept of freedom' (*CpV* 5:30). Consciousness of the moral law leads to the awareness of freedom, Kant argues, because reason presents this law to us as a determining ground in us that is entirely independent of sensible conditions (*CpV* 5:29–30).

It is not immediately clear *how* exactly this happens, and to clarify it further Kant gives the famous false testimony example. He asks his readers to imagine a man whose prince orders him to do something which the man recognizes as immoral (to give false testimony against an honest man), and who is threatened with execution should he fail to obey. Such a man, Kant claims, must admit that it would be *possible* for him to overcome his love of life (and any other inclination he may have) and refuse to commit the immoral act, even if he cannot be certain whether he *would actually* act accordingly in the situation. He *judges* that he *can* act in a way that goes against his inclinations, even against his love of life, in light of his consciousness that he *ought* to. In this way, moral consciousness shows him his freedom (*CpV* 5:30).

Immediately following this argument, Kant formulates the 'Fundamental Law of Pure Practical Reason'. Kant formulates this fundamental law as a categorical imperative: 'So act that the maxim of your will could always hold at the same time as a principle in a universal law-giving' (*CpV* 5:30). The very formulation of the principle (as an imperative) indicates that Kant is here assuming that he has now shown not only what this law entails (which he had already shown in section 4), but also that we stand under the obligation to act in accordance with it. Thus, he assumes he has taken an important step beyond the formulation of the reciprocity thesis. The reciprocity thesis is merely conceptual, and does not entail that we are indeed obligated to act in accordance with the law of pure practical reason. This obligation is established only after the reciprocity thesis is combined with a proof of freedom of the will. And this is what Kant does in the Remark in section 6, illustrated with the false testimony example. Despite being presented as merely a 'remark', then, the justification of the assumption of freedom of the will here, on the basis of moral consciousness, fulfils a crucial role in Kant's argument for the validity of the 'fundamental law', i.e. for its validity, for us humans, as a categorical imperative.

It might seem at first as if Kant argues from the Fundamental Law of Pure Practical Reason to freedom and from freedom to the fundamental law, and hence it might seem as if his argument is blatantly circular. But this is not an accurate description of the structure of the argument. The argument starts

from the *consciousness* of the fundamental moral law as an imperative and justifies the assumption of freedom of the will on this basis. Subsequently, the assumption of the freedom of the will serves to underwrite the actual *validity* of this law for us. So the grounding relationship is different in each direction. Moreover, there is only one starting point: the consciousness of the fundamental law (or, as Kant here also puts it, the moral law). At the beginning of the argument Kant has not yet established that this *belief* that one is obligated is not illusory. In the first step of the argument, the belief that one is obligated is shown to disclose our freedom. On the basis of the reciprocity thesis, this then yields the conclusion that the moral law is indeed valid for us as an imperative.

Kant introduces the terminology of a 'fact of reason' (*Factum der Vernunft*) after the formulation of the fundamental law in section 7.[3] He says that 'this consciousness of the fundamental law' (which he introduced in section 6) 'can be called a fact of reason'. Kant mentions two grounds for his terminology. First, one *cannot infer* this consciousness on the basis of antecedent data of reason. It is not a matter of logical entailment but a claim about existence. The fundamental law is 'given': this law 'forces itself upon us' as a synthetical judgment a priori (*CpV* 5:31). This motivates Kant's choice of the term '*Factum*': this law is *not derived but given* in consciousness. Second, the consciousness in question is *not grounded in any intuition*, pure or empirical. This, together with the unconditionality of the law in question, motivates his calling the consciousness of this law a fact '*of reason*'. He emphasizes that although the law is given to consciousness, it is not given empirically: consciousness of the law is a fact of pure reason, and it is the only such fact (*CpV* 5:31).

This passage, wedged between the formulation of the 'fundamental law' and the conclusion that pure reason is practical, is the core passage on the fact of reason. The fact of reason is again mentioned and discussed later in the Analytic, in the section 'On the Deduction of the Principles of Pure Practical Reason' (*CpV* 5:42, 47), as well as in the discussion about the relationship between theoretical and practical reason (*CpV* 5:55), and then again in the Critical Elucidation (*CpV* 5:91,104). All these occurrences, as well as the mention in the preface (*CpV* 5:6), refer to Kant's argument in sections 6 and 7 of the second *Critique*. There is only one fact of reason in the overall argument of the *Critique of Practical Reason*.

In his later discussions, however, Kant occasionally uses strikingly different wording. Instead of speaking of the 'consciousness' of the 'fundamental

[3] The 'fact' is also mentioned in the preface but not explained there.

law' he speaks of the 'moral law' as the fact, and he repeatedly adds the qualification 'as it were' (*gleichsam*), saying the moral law is 'given as it were as a fact of pure reason' (*CpV* 5:47, 91).

The designation of the law itself (instead of the consciousness of it) as a fact (or at least as given to us as a fact) is probably best explained by pointing out that in so far as the law is *given to us* it is of course given in the form of our consciousness of it. In this sense, Kant can also say that the moral law 'provides' the fact (*CpV* 5:43).

Why Kant modifies his talk of a fact by repeatedly adding 'as it were' is less clear. He often emphasizes the difference between the fact of reason and all other facts, viz., that the first is non-empirical and that it is the only fact that has its origin in pure practical reason. Therefore, it is likely that the qualification is meant to underscore this difference and to avoid a misreading of the fact of reason as empirical.

III. THE *FACTUM* OF REASON: INTERPRETATIONS AND PROBLEMS

It has seemed to many commentators that by appealing to a 'fact' of reason Kant merely bluntly stipulates what he should have carefully argued for. Yet there are profound disagreements among interpreters regarding the way in which the expression is to be understood, and in order to assess these criticisms, it is important to look at these different interpretations of the meaning of '*Factum*' in *Factum der Vernunft*.

Most commentators read the term 'fact' in the (currently) common sense of the word as 'something that has actual existence' (*Tatsache* in German). This has led to a number of questions and objections. In the eyes of many, Kant simply seems to assume the existence and validity of a certain kind of moral experience. Occasionally, a commentator commends Kant for doing so;[4] most interpreters, however, see it as a ground for criticism, because they take Kant to be starting his argument from an 'ultimately un-argued-for premise of the validity of morality'.[5] The problem they see is that the existence of moral experience could be challenged on empirical grounds.

[4] Dieter Henrich, 'The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason', trans. Manfred Kuehn, in Dieter Henrich, *The Unity of Reason: Essays on Kant's Philosophy*, ed. Richard Velkley (Cambridge, MA: Harvard University Press, 1994), 55–87, here 83. Originally published as 'Der Begriff der sittlichen Einsicht und Kants Lehre vom Faktum der Vernunft', in Dieter Henrich (ed.), *Die Gegenwart der Griechen im neueren Denken: Festschrift für Hans-Georg Gadamer zum 60. Geburtstag* (Tübingen: JCB Mohr, 1960), 77–115.

[5] Karl Ameriks, *Interpreting Kant's Critiques* (Oxford: Clarendon Press, 2003), 176, cf. 254.

Moreover, it looks as if Kant starts from an intuitionist claim to moral insight that is not properly available to him in his own terms. Thus, the appeal to a fact of reason looks to some like a form of 'footstamping'[6] or 'moralistic bluster' that makes Kant's argument in the *Critique of Practical Reason* 'significantly weaker' than the argument of the *Groundwork*.[7] Milder assessments, too, see the 'fact of reason' argument as resting on substantive presuppositions that many readers won't accept.[8]

Some commentators deal with the perceived problems by downplaying the importance of the fact of reason. Onora O'Neill and Paweł Łuków regard Kant's argument as a kind of encore, as an explication of how practical reason registers in the 'ordinary lives of ordinary people',[9] without it being part of the main argument of the *Critique of Practical Reason*. Kant's argument concerning the fact of reason doesn't constitute a *failed* justification, on this view, because he does not intend it to be a justification at all. Łuków writes, referring to *CpV* 5:42, that the fact of reason 'is introduced for the first time as a Remark *after* Kant has shown how pure reason could be practical', and that 'this suggests that the doctrine of the fact of reason supplements rather than constitutes the main argument of the second *Critique*'.[10] A fatal problem for this line of interpretation is that the fact of reason is actually introduced *before* the claim that pure reason can be practical (in *CpV* 5:31), and this latter claim is introduced as a 'Conclusion' immediately following the introduction of the fact of reason.

Two other interpretative proposals take a different approach, both focusing on the term 'fact' and breaking with the standard interpretation. One proposal, given its strongest and most detailed defence by Marcus Willaschek, is to read the term '*Factum*' not as *Tatsache* (fact) but as *Tat* (deed). The *Factum* of reason, on this approach, is a 'deed of reason', viz., reason's activity in producing the consciousness of moral obligation.[11]

[6] Paul Guyer, 'Naturalistic and Transcendental Moments in Kant's Moral Philosophy', *Inquiry* 50 (2007), 444–64, here 462.

[7] Allen W. Wood, *Kantian Ethics* (Cambridge: Cambridge University Press, 2008), 135.

[8] Henry E. Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), 230–49.

[9] Paweł Łuków, 'The Fact of Reason: Kant's Passage to Ordinary Moral Knowledge', *Kant-Studien* 84 (1993), 203–21, here 210; Onora O'Neill, 'Autonomy and the Fact of Reason in the *Kritik der praktischen Vernunft* (§§ 7–8, 30–41)', in *Immanuel Kant: Kritik der praktischen Vernunft*, ed. Otfried Höffe (Berlin: Akademie Verlag, 2002), 81–97, here 83, cf. 89.

[10] Łuków, 'The Fact of Reason', 210, referring to *CpV* 5:42. O'Neill speaks of a 'complementary' role of Kant's comments on the fact of reason: O'Neill, 'Autonomy and the Fact of Reason', 83, also 88–9.

[11] Marcus Willaschek, 'Die Tat der Vernunft: Zur Bedeutung der Kantischen These vom "Factum der Vernunft"', in Gerhard Funke (ed.), *Akten des Siebenten Internationalen Kant-Kongresses* (Bonn: Bouvier, 1991), 455–66; Marcus Willaschek, *Praktische Vernunft. Handlungstheorie und Moralbegründung bei Kant* (Stuttgart: Verlag J. B. Metzler, 1992), 174–93; along similar lines see Paul W. Franks, *All or Nothing. Systematicity, Transcendental Arguments, and Skepticism in German*

Linguistically speaking this interpretation is certainly plausible. The first meaning of '*Factum*' in eighteenth-century German was 'deed', and it is beyond doubt that Kant uses the term in this sense in other passages. Moreover, as I shall argue below, interpreting '*Factum*' in *Factum der Vernunft* as 'deed' is appealing for philosophical reasons and sheds interesting light on Kant's argument, although – as I also show – it does not mesh equally well with all the relevant passages.

Another, third approach, defended by Ian Proops, is to read '*Factum*' as a technical term that designates a particular moment in Kant's proof structure.[12] Dieter Henrich has drawn attention to the important role of legal metaphors in Kant's critical project.[13] On Henrich's view, the term 'deduction' should be read in the sense of the eighteenth-century *Deduktionsschriften* that provided legal justifications for territorial claims. According to Henrich, these deductions would typically involve tracing the claim back to a legal action or fact in which the claim originates, and this origin was called a '*factum*'. In other words, one would establish a legal entitlement by proving a '*factum*'. Proops argues that the 'fact of reason' in the *Critique of Practical Reason* should be read in this sense. On his interpretation, 'the *factum* of the Deduction of Freedom consists in the fact that the moral law has a pure origin', i.e. an origin in pure practical reason.[14] Proops sees this as incompatible with the second reading, and he even claims, with reference to the fact of reason in the second *Critique*, that 'in the process of transplanting the [legal] metaphor, the idea of the *factum* as an act recedes from view'.[15]

IV. *FACTUM*: FACT, DEED OR A TECHNICAL TERM?

Should '*Factum*' be understood as fact, as deed, or as a technical term? The Latin verb '*facere*' means 'to do' or 'to make'. The perfect participle '*factum*' can refer both to that which was done (the deed) and to that which was made (the product). In Zedler's *Universallexikon* (1732–54), '*That*' (deed) is the very first meaning of '*Factum*', followed by '*das geschehene Ding*' (the

Idealism (Cambridge, MA: Harvard University Press, 2005), 278; most recently also espoused by David Sussman, 'From Deduction to Deed: Kant's Grounding of the Moral Law', *Kantian Review* 13 (2008), 52–81, here 68.

[12] Ian Proops, 'Kant's Legal Metaphor and the Nature of a Deduction', *Journal of the History of Philosophy* 41 (2003), 209–29.

[13] Dieter Henrich, 'Kant's Notion of a Deduction and the Methodological Background of the First *Critique*', in Eckhart Förster (ed.), *Kant's Transcendental Deductions* (Stanford: Stanford University Press, 1989), 29–46.

[14] Proops, 'Kant's Legal Metaphor and the Nature of a Deduction', 228.

[15] Ibid., 215, and 215n. 19.

thing that happened) and other ways of referring to the *product* of the deed.[16] The German word '*Thatsache*' (fact, matter of fact) did not exist until the second half of the eighteenth century, when it was a neologism devised to translate '*res facti*' (matter of fact).[17] Zedler's lexicon (which predates '*Thatsache*') defines '*res facti*' as 'that which actually exists as a result of the activity of humans or of nature'. '*Thatsache*' came to serve as a translation not only of '*res facti*' but also of '*factum*' in its meaning of 'product'. By the end of the eighteenth century, then, '*factum*' could mean either 'deed' (*That*) or 'fact' (*Thatsache*).

Kant uses '*Factum*' in both senses, and he translates the Latin *factum* into German either as *That* (deed) or as *Thatsache* (fact), depending on which meaning is intended. He often uses it in the sense of 'fact'.[18] He uses *That* (deed) when the term refers to an imputable act, as, for example, in the following passage: '*Imputation* (*imputatio*) in the moral sense is the *judgment* by which someone is regarded as the author (*causa libera*) of an action, which is then called a *deed* (*factum*) and stands under laws' (*MdS* 6:227).

If the consciousness of the fundamental practical law is *produced* by reason, the term 'fact of reason' can naturally be understood to refer to a *deed* of reason. This reading would block interpretations according to which the fact of reason is a brute fact that reason simply has to swallow, a state of affairs that impinges on reason from the outside. Instead, this reading emphasizes that the consciousness of the law originates in reason's own activity.

The 'deed of reason' reading has obvious philosophical advantages over the older reading of the fact of reason as a brute fact. Instead of reading the *Factum* as a state of affairs which reason simply happens to find itself confronted with, on this interpretation it is a deed of reason itself.

In support of reading *Factum* as 'deed', the reading defended by Willaschek, one can point to passages in which the fact of reason is mentioned in a context in which reason is portrayed as active. Kant claims, for instance, that 'in' the *Factum*, pure reason 'proves itself indeed practical' or 'practical in the deed' (*in der That praktisch*, *CpV* 5:42). In the preface Kant writes that 'if it [viz., reason] as pure reason is actually practical, it proves its reality and that of its concepts through the deed, and all subtle reasoning against the possibility of its being practical is futile' (*CpV* 5:3).

[16] Johann Heinrich Zedler, *Grosses vollständiges Universallexikon* (Leipzig: Johann H. Zedler), 1732–54.
[17] Jacob Grimm and Wilhelm Grimm, *Deutsches Wörterbuch*, 16 vols. (Leipzig: Hirzel, 1854–1960), under '*Thatsache*'. See also Kant's translation of *res facti* as *Thatsache* in *CU* 5: 468.
[18] E.g. *CrV* A84/B116; *MdS* 6:371.

Willaschek rightly points out that Kant uses '*Factum*' in both senses (deed and fact). The question is, however, whether '*Factum*' in the core '*Factum der Vernunft*' passages in the second *Critique* (especially *CpV* 5:31) is best read as 'deed', as he believes that it is.[19] A first indication that this might not be so is the circumstance that this interpretation makes it difficult to make sense of comments such as Kant's statement that the 'moral law provides a *Factum*' (*CpV* 5:43) (*das moralische Gesetz gibt ein … Factum an die Hand*). The moral law does not provide a deed – it would be hard to know what that would mean. As I explain below, I believe that 'fact' is the better reading of the crucial passage in section 7, but with the understanding that this fact is the result of a deed of reason.

The textual difficulties with reading '*Factum*' as a technical term for a specific moment in a proof, as proposed by Proops, are considerable. At first sight, this proposal looks promising. As mentioned, Kant draws heavily on juridical terminology at many crucial junctures in his *Critiques*, which motivates Proops's proposal to read '*Factum*' as a technical juridical term. His view is that despite the contrast Kant draws in the *Critique of Pure Reason* (*CrV* A84/B116) between questions of right and questions of fact, his 'deductions' involve *both* and have a two-step structure, starting with a proof of a particular action or fact in which the claim at issue originates.[20] In the second *Critique*'s deduction of freedom, so Proops argues, this '*Factum*' is the fact of reason.[21]

One problem for this reading, however, is that Kant's own explanation for his choice of the term '*Factum*' has nothing to do with the proof structure of a deduction but rather with the consciousness of the fundamental law of pure practical reason as 'given'. Furthermore, there are no passages elsewhere in the *Critique of Practical Reason* where '*Factum*' is used in an unambiguously technical sense and which could directly corroborate this reading.

Another disadvantage of Proops's reading is that it downplays the active role of reason. As mentioned above, he even claims that the active meaning of *Factum* recedes from view in the second *Critique*. Given that many of the 'fact of reason' passages portray reason as active, it would seem that something important is left out if this aspect cannot be captured. This is not to

[19] Willaschek, *Praktische Vernunft*, 184–93 ('Die Tat der Vernunft und die Begründung der Moral').

[20] Proops, 'Kant's Legal Metaphor and the Nature of a Deduction', 211. Zedler's *Universallexikon* includes a separate entry for '*Factum*' in a technical legal sense of a '*casus*', as the case 'without precise insight [into which] the law cannot be applied'; Zedler, *Grosses vollständiges Universallexikon*.

[21] Proops, 'Kant's Legal Metaphor and the Nature of a Deduction', 228–9. It is somewhat puzzling that Proops formulates the first step in different ways: either as the proof that the moral law has a pure origin, or as the proof 'that the obligations we take ourselves to be under are genuinely *categorical* obligations, as opposed to hypothetical principles of prudence, or whatnot' (228–9).

deny that the deduction of freedom could be fruitfully reconstructed as an argument that starts with the fact of reason, but such a reconstruction does not depend on reading '*Factum*' in the technical sense. Therefore, it is unclear what novel hermeneutical insights are disclosed by this reading, whereas it is clear that this reading has significant drawbacks.

I believe that a better interpretation can be developed by modifying the first reading (*Factum* as fact), making use of the insights of the second reading (*Factum* as deed). This is to read '*Factum*' as a fact, that is, to take moral consciousness as something which exists – but not as an alien fact that reason happens to be confronted with, such as the existence of a contingent set of parochial values. Rather, it is to take moral consciousness as a fact that is the result of reason's activity. This is to read the '*Factum*' as a fact (*Tatsache*) that is generated by a deed (*Tat*) of pure practical reason itself, namely as the consciousness of the fundamental law (of pure practical reason), a consciousness that reason produces in rational agents. The fact is then a fact 'of reason' just as a decision can be 'a decision of the king', or a painting 'a painting of Rembrandt'. Read in this way, the fact of the consciousness of the fundamental law of practical reason is the result or product of a deed of reason. '*Factum*' is therefore best translated as 'fact', not as 'deed'; but what exactly the fact *is* is to be understood by taking into account that *Factum* is the perfect participle of the verb 'to do, to make', and that the fact at issue is the *product* of reason's own activity.

V. MORALITY AND PRACTICAL REASON

The proper way of *reading the expression* 'fact of reason' does not yet tell us how successful Kant's use of it is in his argument. Much more needs to be said in order to address the many questions and objections mentioned at the outset. Kant claims that consciousness of the moral law (as binding) exists, and a recurring complaint in the literature is that he fails to question critically the validity of the belief that one is morally obligated. Therefore, it is important to examine what exactly the consciousness *is* that Kant claims exists, and what his reasons are for claiming that it exists.

Virtually all authors who discuss the fact of reason do so in terms of morality. They introduce the 'consciousness' that Kant calls a fact of reason as the 'consciousness of the *moral* law' or '*moral* consciousness'. This is certainly right, because Kant does. But it can also be misleading. Most of the worries that I mentioned at the outset can be traced back to a failure to interpret properly Kant's equation of the fact of reason with moral consciousness.

It is striking and significant that the argument of the first chapter of the *Critique of Practical Reason* proceeds not in terms of morality and the moral law, but in terms of practical reason and the fundamental practical law. Indeed, the term 'moral' and related terms (*moralisch*, *Moral*, *sittlich*, *Sittengesetz*, and so forth) do not occur until *CpV* 5:29, just before the formulation of the Fundamental Law of Pure Practical Reason of which Kant then adds: 'which we call the moral law' (*CpV* 5:31). We find a similar locution in the *Metaphysics of Morals:* Kant says that the positive concept of freedom grounds 'unconditional practical laws, which are called *moral*' (*MdS* 6:221).

The consciousness of the fundamental law of pure practical reason is indeed most fundamentally the consciousness of a rational principle. Kant also calls this law the moral law. That should not be surprising: morality concerns the most fundamental principles for action, and so if a *rational principle* has been identified as the *most fundamental* principle of action, this rational principle can be called the *moral law* (see also *ZeF* 8:370). Once Kant has explicitly equated the two, he refers to the fundamental law of pure practical reason more frequently as the moral law. In essence, however, it is the consciousness of the fundamental law of pure practical reason that is called a fact of reason. This law is subsequently called the moral law – not the other way around.

Paying attention to this feature of Kant's argument helps to clarify some of the common questions and objections that are raised in response to Kant's discussion of the fact of reason. As I mentioned, one often hears the objection that Kant should not have assumed that there is widespread agreement about moral obligation. But it should now be clear that this objection attaches itself to the wrong object. Discussing the fact of reason exclusively in terms of a moral consciousness, instead of consciousness of the fundamental law of pure practical reason, creates the risk that 'morality' is read in terms of readers' own 'material' conception of morality instead of in terms of Kant's formal concept of it. The proper object of criticism should instead be Kant's claim about the alleged consciousness of the law of practical reason. For *that* is the alleged fact; and Kant calls it 'moral' consciousness simply because it is our consciousness of the most fundamental normative principle guiding human action.

This is not to answer the objection, of course, but merely to shift it to the proper location. The next question, then, is whether Kant's fact of reason argument fares any better as a result. Sceptics, psychological egoists and many other critics are likely to object to Kant's claim in its refined form, too, as the claim that there is consciousness of the law of pure practical reason which is produced by reason.

VI. KANT'S ARGUMENT REVISITED

Many critics see the major weakness of Kant's account in the second *Critique* as lying in his claim that moral consciousness is a fact and the basis for our assumption that we are free. Some challenge Kant's assertion regarding the very existence of the consciousness at issue. Others fault Kant for a bootstrapping problem in the argument. What is missing in the fact of reason argument, they say, is a convincing account of how *normativity* emerges from the mere existence of a particular consciousness or belief. In their eyes, Kant fails to address the problem that morality might be a 'figment of the brain', as he himself called it in the *Groundwork* (*G* 4:407, 445), where he did acknowledge this objection.

In order to assess whether these criticisms can be answered, we need to look more closely at what exactly the 'fact' is, and how Kant argues for its existence. The weight of the argument is on the assumption of the existence of this consciousness – both on the positing of its universal existence and on the articulation of what it consists in, and it is crucial to assess Kant's argument for both parts of the claim.

Now some commentators have suggested that we should not raise the bar too high with regard to the first part. Proving the *universal* existence of moral consciousness of one kind may not be possible, but it is nevertheless philosophically worth while to capture the principle underlying a form of moral consciousness that appears to be widespread.[22] Only very few people would deny or doubt the claim that it is wrong to torture babies for fun (to take just one example from the literature), and thus only very few people would deny or doubt that there are normative constraints on action at all. So most people, they argue, can be said to have a consciousness of moral obligation, even though there are of course vast differences when it comes to specific conceptions of how one is morally required to act. The substance of Kant's argument, on this reading, concerns the precise articulation of the principle of normative constraint (and the defence of this articulation), and this argument should speak to the vast majority of people.

The ambition of both Kant and most Kantians, however, is to accomplish much more, namely, to ground the consciousness of moral obligation in the structure of practical reason, so as to reach conclusions that apply universally. Kantian theorists from Habermas to Korsgaard have developed arguments to reply to the objections at issue. I would here like to examine, however, whether one can make good sense of Kant's own second *Critique*

[22] See, for example, Ameriks, *Interpreting Kant's Critiques*, 260.

argument regarding the consciousness of the fundamental law of practical reason in a way that answers the criticisms.

Kant claims that the fundamental law of practical reason immediately presents itself as soon as an agent constructs maxims for his will (*CpV* 5:29). Maxims are, on Kant's theory of action, the principles an agent adopts as rules for action. Mere behaviour on blind impulse does not involve the use of maxims. Maxim construction presupposes an agent who deliberates about possible courses of action in light of a normative standard. Kant's claim is that this standard is provided by practical reason. He claims that the very moment agents consciously reflect on possible maxims of action, they 'immediately' become conscious of the fundamental law of pure practical reason (*CpV* 5:29).

This claim requires further elaboration, but it is important to note at the outset that Kant's argument starts by positing the existence of agents who face decisions and who deliberate about maxims. That seems not too much to presuppose as a starting point: Kant's critics, too, will have to grant that agents, from the agents' own perspective at least, face decisions and deliberate about possible rules of action. Some may wish to add that the agents' sense of freedom in such situations is a mere illusion or that deliberation can concern only means–ends relationships, but that is not of concern until the next step in the argument.

What does it mean for Kant to say that agents become conscious of the fundamental law of pure practical reason the moment they start constructing and evaluating possible maxims of action? Agents who regard themselves as facing a decision and as deliberating about possible maxims of action ascribe to themselves (implicitly or explicitly) the capacity to guide their actions on the basis of reasons. They ask themselves how to act and why. Their questions do not confine themselves to questions of expediency and prudence, because if and when they seriously ask themselves *why* they would act in a certain way, their reflection naturally leads to questions regarding ends and general principles of acting. Such agents ask themselves not merely whether certain actions are instrumental in achieving a given end, but also whether the end itself is worth pursuing (and why). And this means that they can (and will, if they are serious and push this process far enough) ask themselves whether to act on their inclinations at all.

This very question, however, indicates that their deliberation about maxims presupposes a normative principle, and that this normative principle is conceived as independent from the inclinations. That is, when they conceive of themselves as acting on reasons, they conceive of their will as a causality on the basis of reason and 'independent from empirical conditions' (*CpV* 5:31). 'Independent' here does not mean that inclinations play no role

in practical deliberation, but rather, that the will is conceived as *not dependent on* inclinations, in the sense that agents judge that they can choose to act in a way that runs counter to all of them if they see reasons to do so. According to Kant's analysis, as we saw, this comes down to saying that they judge that their will is free. A free will conceived as a causality on the basis of reason, however, is, as Kant has argued earlier in the Analytic, a pure will, 'determined *by the mere form of the law*' (*CpV* 5:31). Therefore, from the standpoint of the deliberating agent, the 'fundamental law' is regarded as the supreme condition of all choice of maxims (*CpV* 5:31). Kant writes,

> The fact mentioned above is undeniable. One need only analyze the judgment that people pass on the lawfulness of their actions: one will always find that their reason, whatever [their] inclination may say to the contrary, nevertheless, incorruptible and coerced by itself, always compares the maxim of the will of an action to the pure will, i.e., to itself, insofar as [reason] regards itself as a priori practical. (*CpV* 5:32; cf. *CpV* 5:91)

In other words, agents who regard themselves as having a will regard the fundamental law of practical reason as the normative principle guiding their choice of maxims (implicitly, or explicitly when they reason correctly). In attributing a will to themselves, agents implicitly acknowledge this principle as the normative standard for the assessment of maxims. This fact, i.e. the consciousness of this fundamental law, on the part of agents who regard themselves as facing choices regarding maxims, is exactly what Kant calls a fact of reason.

Given Kant's equation of the fundamental law of pure practical reason with the *moral* law, this comes down to saying that agents who regard themselves as facing a decision about maxims must (and do, at least implicitly) recognize the moral law as valid. But the core argument can be crafted entirely without using moral terms, and no parochial set of moral intuitions or moral values is presupposed. All that Kant presupposes, beyond the earlier argument in the Analytic, is the perspective of agents who regard themselves as reasoning about courses of action and who construct and assess maxims on the basis of this reasoning. And what Kant claims to prove is that such agents are conscious of the fundamental law of practical reason as the normative standard that should guide their deliberation.

If this reconstruction of Kant's argument is correct, a common view on the relationship between the *Groundwork* and the second *Critique* needs adjustment. It is often thought that Kant's claim that moral consciousness is a 'fact of reason' reflects his recognition of the futility of his earlier attempt, in the *Groundwork*, to offer a 'route into morality from truly non-moral

concerns and self-conceptions'.[23] In some sense, Kant does offer a non-moral route in the *Critique of Practical Reason*: the entire argument can be cast in (presumably 'non-moral') terms of a theory of action and be regarded as the articulation of the self-understanding of agents who take themselves to be reasoning about which maxims to adopt and why. But this argument is not a mysterious 'route into' morality. The fundamental law of rational agency *is*, as Kant puts it, 'what we call the moral law'. Morality is nothing other than the (set of) standard(s) determining how we ought to act. If how we ought to act is determined by a single rational principle, as Kant claims it is (viz., the categorical imperative as the fundamental law of pure practical reason), then this rational principle is suitably called the moral law. There is nothing enigmatic about that.

By saying that the relation between the arguments of the *Groundwork* and the *Critique of Practical Reason* needs to be re-evaluated, I do not mean to imply that his arguments in the two books are identical. The most striking difference is Kant's reversal of the order of the argument concerning freedom. In the *Groundwork* he argues from freedom of the will to the moral law, whereas in the second *Critique*, the argument proceeds from consciousness of the moral law to freedom of the will. Thus, even though the *aim* of Kant's argument in the second *Critique* shows more continuity with the *Groundwork* than is usually thought, Kant's argumentative *strategies* in pursuing this aim are different in important respects.

Now someone might wonder whether this reconstruction of Kant's argument in the *Critique of Practical Reason* does not wrongly present the belief in freedom of the will as an inference from the mere capacity to reason as such, and whether in doing so it does not lose sight of the fact that Kant's defence of freedom appeals specifically to *moral* consciousness. In the false testimony example, Kant relies on the agent's recognition of the authority of morality, and not on a general capacity to reason, to back up the belief in freedom. Hence, one could think, there is still a 'moral remainder' that cannot be captured by an analysis of the argument in terms of practical reason.

In response to this worry, it is important to point out that Kant certainly did not argue, in the second *Critique*, that the mere capacity for means–ends reasoning could suffice to show us that our will is free. The employment of practical reason in the service of an end given by inclination, even if this end is a very general one such as one's happiness, does not suffice to show us our freedom. Thus, whether or not one is able to resist one's lustful inclination when threatened with execution at the gallows should one give in to it, one

[23] Sussman, 'From Deduction to Deed', 65; see also Ameriks, *Interpreting Kant's Critiques*, 256–9.

does not learn of one's freedom here (see *CpV* 5:30). Only when one realizes that one could act against *all* of one's inclinations, as in the false testimony example, does one become aware of one's freedom in the sense required by Kant. In such cases it is not inclination which is conceived as determining the will, but pure practical reason. Only such cases, and not instrumental practical reasoning, disclose the freedom of the will.

One objection has not yet been touched upon, and this is the claim that the *belief* that we are morally obligated does not prove we *are*, that morality might be a figment of the brain, and that freedom of the will (of the Kantian variety) could be a mere illusion.

In assessing the strength of this objection, it is crucial to remind oneself of Kant's distinction between theoretical and practical reasoning. From a theoretical perspective, Kant holds, all our actions (not just some of them) as appearances are entirely determined by natural causal laws. But Kant denies that it is possible for agents qua deliberating agents (that is, from the practical standpoint) consciously to regard their own judgments as determined by natural causes.[24] From the perspective of deliberating agents no inclination sets absolute limits to their own freedom. Kant's argument for the assumption of freedom has a conclusion that is indeed radical. This is what the false testimony example is meant to illustrate: agents judge that they *can* determine their will in accordance with what is right, because they judge that they *ought* to do so.

But what are we to say about the possible illusoriness of this judgment, i.e. this belief in freedom of the will? Believing something does not make it so, and therefore it has been objected that having the consciousness of the fundamental law, and hence having the belief in freedom of the will, does not guarantee that we are 'really' free.[25] Note, however, again looking back to the false testimony example, that Kant's argument establishes that one must *judge* that one can (i.e. that one must judge that one is free), not that one 'really' can. The latter claim would entail metaphysical knowledge of freedom from an objective and theoretical point of view. Kant, however, does not claim to prove freedom in that sense; his project in the *Critique of Pure Reason* led to the conclusion that doing so is impossible. But the first

[24] See his comment in the *Groundwork* that it is impossible to conceive of a rational being that consciously regards its own judgments as determined by natural causes (*G* 4:448), because then such a rational being would not count these judgments as its own.

[25] Barbara Herman presses the point that *believing* we are able to act from respect for the moral law 'does not bootstrap us up to transcendental freedom'. Barbara Herman, 'Justification and Objectivity: Comments on Rawls and Allison', in Eckhart Förster (ed.), *Kant's Transcendental Deductions* (Stanford: Stanford University Press, 1989), 131–41, here 135.

Critique also showed that there is no viable perspective from which to prove that the belief in freedom (in the moral context, by the deliberating agent) is illusory – for this proof similarly requires metaphysical knowledge of a kind that is not available to us. In this way, the assumption of freedom of the will, which is a necessary presupposition from the perspective of the deliberating agent, can be 'defended' against the charge of illusoriness.

Thus, starting from the *consciousness* of the fundamental law of pure practical reason, Kant justifies the assumption of freedom of the will. The assumption of freedom of the will in turn establishes the *validity* of the fundamental law for us. At the point at which the argument starts, Kant has not yet shown that the belief that one is obligated by the law is non-illusory. But after justifying the assumption of freedom (on the basis of this belief), he establishes the validity of the fundamental law, and at that point the consciousness of obligation can no longer be judged to be illusory (e.g. as a mere product of imagination). Given the justification of the validity of the fundamental law (also known as the moral law), the consciousness of this law must be regarded as a product ('fact') of reason. This argument is convincing only from the agent perspective, but in so far as this is a perspective which is inescapable for us humans and which cannot be undermined from a theoretical perspective, it is for us, in Kant's view, fully convincing.

Kant's account of the fact of reason is, therefore, radically agent-based. The argument is developed from the agent's point of view, because it concerns practical reasoning from the perspective of agents, not theoretical reasoning *about* agents from an external perspective. It is an argument not from the concept of morality, nor from the concept of an agent, nor even from the fact that there are agents. Rather, it is an argument developed from and valid for the perspective of those who take themselves to be agents. The argument establishes what such subjects must judge when they engage in practical deliberation about maxims. Kant grants the point that, merely theoretically speaking, freedom of the will might be an illusion, but settling this question lies entirely outside the purview of theoretical reason. Freedom of the will is a necessary assumption for someone judging from a practical point of view, and it can be defended on the basis of the consciousness of the fundamental law of pure practical reason, a consciousness which Kant calls a fact of reason.[26]

[26] I am grateful to the Alexander von Humboldt Foundation for the research fellowship that made work on this article possible. I benefited greatly from comments from other authors in this volume (especially Andrews Reath, Marcus Willaschek and Jens Timmermann) and from Joel Anderson, Claudia Blöser, Antonino Falduto, David Sussman and Micha Werner.

CHAPTER 4

Reversal or retreat? Kant's deductions of freedom and morality

Jens Timmermann

I. KANT'S CHANGE OF MIND

It is one of the unsettling lessons of Kant's critical philosophy that even the most painstaking conceptual analysis does not justify applying the concepts analysed. On a purely analytic level, much can be said about unicorns, bachelors or God, but it is an entirely different question whether there are such fantastic creatures, i.e. whether there is something to which we refer when we avail ourselves of these words in conversation. In philosophy, practical as well as theoretical, this problem is particularly acute when these concepts rest on synthetic judgments a priori. For instance, the analytic sections of the *Groundwork of the Metaphysics of Morals* are intended to convey a precise impression of what morality means. The principle of morality, Kant argues, is stated with due philosophical precision for the first time ever in the history of ethical thought. But to dispel the sceptical concern that morality might not be real he has to embark on quite a different project. He needs to show that morality is more than a nice idea that quick-witted philosophers can analyse to their hearts' content, while in practice it does not, and cannot, determine human action. After all, it is impossible to point to a morally valuable action in experience, even in principle, be it from the third- or the first-person point of view. What is more, experience reveals all human actions to be determined by natural processes, not by an unconditional moral command that expresses a different, normative kind of necessity. If they occur at all, the value of actions determined by this does not surface. For all we know there has never been, nor ever will be, a morally worthy deed.[1]

[1] It is worth noting that it is *this* sceptical challenge, not that of the modern amoralist who refuses to acknowledge the authority of morality altogether, that motivates Kant's justificatory efforts in the moral sphere. Anyone who is otherwise accustomed to using reason would, on reflection, like to be a decent person, even the famous 'most hardened scoundrel' (*G* 4:454). We shall return to this character in due course.

Kant's attempt to vindicate the authority of a categorical moral imperative is obviously a matter of significant philosophical, exegetical and historical interest. However, it is also one of the rare cases in which Kant comes close to admitting that he has changed his mind.[2] In a surprising reversal towards the end of the deduction section of the *Critique of Practical Reason*, Kant tells us that the 'vainly sought deduction of the moral principle' (*CpV* 5:47) is replaced by another deduction, namely the deduction of freedom. Of course, Kant does not explicitly say that *he* is the person who tried in vain to provide a deduction of the moral principle, but as *Groundwork* III contains precisely such a deduction, and proving the possibility of the categorical imperative was hardly fashionable amongst his philosophical colleagues, we can safely infer that Kant had come to reject his earlier justificatory attempt by the time he composed the second *Critique*. It is also worth noting that the sceptical worries that pervade the *Groundwork* are almost completely absent from the second *Critique*.

However, this initial diagnosis raises more questions than it helps to answer. How far-reaching are these changes? What motivates them, and what exactly do they consist in? How much of the substance of the earlier account survives? What are the implications of Kant's change of mind for the architectonic of his system? Finally, should the new approach be regarded as progress, perhaps as an acknowledgement of due philosophical humility, or as a relapse into an inappropriate pre-critical dogmatism? I hope that this chapter will shed some light on these matters.

II. DEDUCING THE PRINCIPLES OF PURE PRACTICAL REASON

Let us turn to the opening sentences of the section entitled 'Of the Deduction of the Principles of Pure Practical Reason', appended to the first chapter of the Analytic:[3]

[2] The only other passage I can think of can be found in the 'First Introduction' to the *Critique of Judgment*, which in the end Kant decided not to commit to print: he declares that this 'is the place to correct a mistake I made in the *Groundwork of the Metaphysics of Morals*', namely that the least kind of imperative should be called 'technical', not 'problematic', a comparatively minor terminological correction (see 20:200fn.).

[3] The heading is curious because Kant proceeds to deny that a deduction of the principle of morality is possible, even if he eventually offers a deduction of freedom. Does he consider freedom a principle of pure practical reason? This would explain the plural 'principle', but it would be odd, first because it is a principle of theoretical reason – if one substantiated by pure practical reason – and second because only one of the two principles mentioned would be capable of being deduced (freedom, but not the moral principle). Should the loose '*Of* [*von*] the Deduction' be understood as less than a promise that a

The Analytic shows that pure reason can be practical, i.e. can by itself, independently of anything empirical, determine the will – and it does so by a fact in which pure reason proves itself in fact [*in der Tat*][4] practical in us, namely autonomy in the principle of morality, by which reason determines the will to the deed [*die Tat*]. – At the same time it shows that this fact is inseparably connected with, and indeed identical with, consciousness of freedom of the will, whereby the will of a rational being that, as belonging to the sensible world, cognizes itself as, like other efficient causes, necessarily subject to laws of causality, yet in the practical sphere is also conscious of itself on another side, namely as a being in itself, conscious of its existence as determinable in an intelligible order of things – not indeed, by a special intuition of itself but following certain dynamic laws that can determine its causality in the sensible world; for it has been sufficiently proved elsewhere [*ist anderwärts hinreichend bewiesen worden*] that freedom, when[5] it is attributed to us, transfers us [*versetzt uns*] into an intelligible order of things. (*CpV* 5:42)

We shall return to the details later. For now, let us note that the expression 'sufficient proof' indicates that one central move of the earlier justificatory project is still intact. There can be little doubt that the 'transferral' mentioned in the last sentence, which makes us take up another, practical point of view in moral deliberation, refers to the third and final section of the *Groundwork*.[6] If so, Kant's rejection of the deduction of the moral law must be limited. A cursory glance at *Groundwork* III from the point of view of the second *Critique* shows that Kant does not jettison the reciprocity thesis that freedom and morality are both identical with autonomy, i.e. that there are

deduction of the supreme principle of morality is to follow, as Karl Ameriks has suggested ('Pure Reason of Itself Alone Suffices to Determine the Will', in Otfried Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft* (Berlin: Akademie Verlag, 2002), 107)? This argument is not compelling, because the corresponding section in the first *Critique*, which is clearly meant to contain a deduction, has a similarly vague heading (see *CrV* A84/B116). This suggests that Kant simply construed the title of the new deduction section along the lines of the old one. However, the plural might refer to the two types of principle he declares the *Critique* has to start with, i.e. subjective and objective practical principles (see *CpV* 5:19.7, 5:16.23, 5:89.25 – references refer to volume, page and line numbers of the Academy Edition). It would admittedly be curious to speak of the deduction of a subjective principle in isolation, but maxims do enter the justificatory project. After all, the fact of reason proves that maxims (subjective principles) can be determined by purely rational laws (objective principles), whether the object intended is realized or not. For a description of the task in those terms see explicitly *CpV* 5:45.

[4] Given the sentence structure, it seems unlikely that the expression is a direct allusion to the fact of reason as a 'deed': pure reason, Kant explicitly says, proves itself practical *in the fact of reason* as presented in the Analytic. There is therefore no room for a second place *in which* pure reason proves itself practical, which rules out the literal meaning 'in the deed' or 'by means of a deed'. No additional reference to the authority of reason is needed. Still, Kant is clearly saying that practical reason is *indeed* or *actually* practical.

[5] *wenn*; most translators have 'if', but 'when' or even 'whenever' is required by the argument because there can be no question *whether* we attribute freedom to ourselves – we do.

[6] The very word used, *versetzen* (to 'transfer' or 'transport' oneself, *CpV* 5:42.19, 5:43.30, 5:34.4 and cf. 5: 92.37 in a less terminological sense), occurs no less than three times in *Groundwork* III: *G* 4:453.12, 4:454.32 and 4:455.2, followed by the cognate *setzen* a couple of pages further down at *G* 4:457.9.

analytic arguments from freedom to morality and back again, or that they are really just two sides of the same transcendental coin (see *CpV* 5:29.24). Nor does he abandon the idea that there are two points of view that we take up in looking at ourselves – one passive, theoretical, explanatory, the other active and practical – and that when we take up the latter we consider ourselves members of a world of pure intelligences. These ideas cover the first three sub-sections of *Groundwork* III, taking us up to *versetzen* at *G* 4:453.12.

If we turn to the next occurrence of *versetzen*, at *G* 4:454.32 in the fourth sub-section, we notice that Kant does not abandon the idea that 'the practical use of common human reason' can be adduced to confirm the results of philosophical ethics. On the contrary, the role of common human reason is strengthened in the second *Critique*, and the paragraph beginning at *G* 4:454.20 is very close in spirit to what the later work dubs the 'fact of reason'. The *Critique of Practical Reason* similarly endorses the substance of the fifth sub-section of *Groundwork* III, the thesis that the interest we take in moral action is inexplicable,[7] and the concluding remarks in the sixth, that we cannot comprehend the nature of moral interest, but at best its incomprehensibility (see *CpV* 5:72.21–4). Accordingly, the two occurrences of *versetzen* at *G* 4:453.12 and 4:454.32, which circumscribe the first and second paragraphs of the fourth sub-section, demarcate the passage we have independent reason to believe contains the actual deduction of the principle of morality as a principle of autonomy.[8] If so, what does this deduction consist in? And why, writing the *Critique of Practical Reason*, did Kant judge it to be a failure?

III. THE 'DEDUCTION' IN THE *GROUNDWORK*, AND WHY IT FAILS

To understand the strategy of *Groundwork* III, it is important to bear in mind that for Kant, at least initially, there is not much of a question as to *whether* the moral law applies to human beings at all. Morality – like free will – is the default position that outside the lecture theatre few would seriously dare to challenge. Sceptical worries arise because we realize that we

[7] However, in 1788 the inscrutability of moral interest no longer represents the 'extreme boundary' of practical philosophy (*G*4:455), which has shifted to include any account of how the moral law is binding upon us as a synthetic a priori principle, i.e. a deduction of the categorical imperative.

[8] Note that the first paragraph of sub-section 4, which argues that the intelligible world is the ground of the sensible world and as such more authoritative, does *not* contain the deduction. It is preparatory in character. The second paragraph does.

live in a world of unremitting causal regularity, in which morality does not appear. We start wondering *how* a morality is possible that revolves around a categorical imperative as a principle of freedom and self-determination. There is plenty of evidence that this is Kant's position – see for example the very beginning of Section II of the *Groundwork*, 4:406.

The *Groundwork* project of a deduction of the moral principle is thus broadly explanatory.[9] A convincing account of how the moral law can guide human action would successfully silence sceptical objections like those mentioned above. The deduction is meant to show *how* a categorical imperative is possible; *if* this can be done, we can rest assured *that* moral commands actually do apply to us, naturalistic worries notwithstanding. As we were favourably disposed towards that assumption from the start, no further argument is needed to persuade us that the categorical imperative is real. According to the *Groundwork* position, scepticism of the empiricist variety gets a foot in the door only if this project fails, i.e. if we are still at a loss as to how Kantian morality can be anything other than an endearing but impracticable idea. At no point in *Groundwork* III is Kant addressing the amoralist.

At the outset of Section III (*G* 4:447) Kant describes the task at hand as follows. Synthetic propositions are possible if two 'cognitions' (*Erkenntnisse*) are conjoined by their connection with some third thing or element, in which both can be encountered. Kant says that this third thing is 'provided' (*schafft*, in the sense of *verschafft*) or 'indicated' (*weiset*) by the concept of freedom. After some preparation that involves our escape from a vicious circle, which comprises the following two sub-sections, this 'third something' turns out to be the idea that we are members not just of the world of sense but also of some supersensible 'world of understanding' or 'intelligible world'. With these philosophical resources in place, the human will can now be shown to be linked to the moral law. If we want to locate the 'deduction' at one specific point in Section III it is this paragraph:

> And thus categorical imperatives are possible because the idea of freedom makes me a member of an intelligible world, by virtue of which, if I were that alone, all my actions *would* always conform with the autonomy of the will, but as at the same time I intuit myself as a member of the world of sense, they *ought* to conform with it; which *categorical* ought represents a synthetic proposition a priori because to my will affected by sensuous desires there is added the idea of the same will, but

[9] *Broadly* explanatory, i.e. not in Kant's narrowly terminological sense of empirical explanation (for which see, e.g., *G* 4:459.3). Kant is trying to provide a philosophically respectable account of how human morality as a matter of categorical obligation is possible.

belonging to the world of the understanding, pure and practical by itself, which contains the supreme condition of the former according to reason; roughly in the way that concepts of the understanding, which by themselves signify nothing but lawful form as such, are added to the intuitions of the world of sense, and thereby make possible synthetic propositions a priori, on which all cognition of a nature rests. (*G* 4:454.6–19)

I shall resist the temptation to dwell on the details of the argument. What is important for our present purpose is that Kant expressly compares the deduction of the categorical imperative with the justification of the theoretical principles of pure reason through which we come to cognize nature, which are brought together by being encountered in intuition. What seems to have happened between 1784 and 1787 is that Kant realized – perhaps in the course of revising the first *Critique*, or as a result of the criticisms levelled against his earlier works – that the deduction at *G* 4:454.6–19 was *in principle* cut off from any kind of confirmation by facts that are independently given, however indirect. Freedom *transfers* us into an intelligible order of things, but that does not suffice as the basis of a deduction because we are in no position to intuit it; and independent evidence of the 'third something' that freedom points to would be needed to make the deduction work.[10]

As a consequence, *Groundwork* III now is deemed insufficient to establish the objective validity of the principle of morality for beings like us, who possess a finite will.[11] That is why in the second *Critique* Kant expressly rejects the idea that such a deduction was ever or will ever be devised. In fact, his denial of the possibility of a deduction refers to this problem. After a rough summary of the first-*Critique* deduction, which by comparison (he argues) was plain sailing, he says:

With the *deduction*, i.e. the justification of its objective and universal validity and insight into the possibility of such a synthetic proposition a priori, one cannot hope to get on so well as was the case with the principles of the pure

[10] Moreover, it is now clear that freedom, which transfers or transports us into an 'intelligible order of things' (*CpV* 5:42), is not given prior to the moral law. It is unclear whether in the *Groundwork* Kant is committed to the view that it is given prior to the moral law – there are hints that we take up the point of view of freedom even when we exercise theoretical spontaneity or act on prudential grounds, which need not, however, prove the strong form of practical freedom required for moral action – but his position in the second *Critique* is unequivocal: the two gallows examples at *CpV* 5:30 are supposed to demonstrate that moral consciousness reveals free will to us whereas prudential action does not. Re-enter the spectre of circularity – now addressed in the preface by means of the distinction between *ratio essendi* and *ratio cognoscendi*, see *CpV* 5:4fn.

[11] For the need to involve intuition in the substantiation of synthetic principles a priori see *CrV* A157/B196. If that cannot be done, there is a danger that we are occupying ourselves with a mere 'phantasm' (*Hirngespinst*), which is exactly the worry Kant intended to put to rest in *Groundwork* III.

theoretical understanding. For these referred to objects of possible experience, namely to appearances, and it could be proved that these appearances could be *cognized* as objects of experience only by being brought under the categories in accordance with these laws and consequently that all possible experience must conform to these laws. But I cannot take such a course in the deduction of the moral law. For the moral law is not concerned with cognition of the constitution of objects that may be given to reason from elsewhere, but rather with a cognition insofar as it can itself become the ground of the existence of objects and insofar as reason, by this cognition, has causality in a rational being, that is, pure reason, which can be regarded as a faculty immediately determining the will. (*CpV* 5:46)

In other words, there is no indication that Kant considered the *Groundwork* account of how the human will is tied to the law of autonomy erroneous – in fact, the elements of this argument reappear in the deduction chapter of the *Critique of Practical Reason*, *CpV* 5:46.5–15, where the concept of one's existence in an intelligible world is said to be the foundation of morality that takes the place occupied by intuition in the first *Critique* – while as indicated already at *CpV* 5:42 an intuition of one's existence in such a world is not and cannot be available. It is simply inadequate philosophically for the task it was intended to do. Kant is no longer prepared to call it a deduction, and denies its very possibility in the subsequent paragraph (the paragraph just quoted). On his considered view, the discarded 'deduction' of *Groundwork* III is a nice story about how our will is necessitated by the laws of reason. It is a story that in 1787 Kant can still tell on the strength of the 'fact of reason'. But it is not something that by itself suffices to dispel the worry that morality is just that: *another* nice story. It is still true that freedom, when attributed by the agent to himself, transfers him into an intelligible order, but this does not amount to intuiting oneself as part of that order, which renders a deduction – an account of how the moral law is binding – impossible.

Consequently, in the *Critique of Practical Reason* the differences between theoretical and practical philosophy come into much sharper focus. In the *Groundwork*, Kant still tried to assimilate the theoretical and the practical justificatory tasks, but now he insists that they are radically different. The former is concerned with the cognition of objects and its a priori foundations; the latter operates in a sphere where objects are not yet given, but must be brought about as a result of our will's being determined. In the essay on *Theory and Practice* this is put in terms of what we would call the 'direction of fit'; and very much in the spirit of the second *Critique* it is argued that doubts about the reality of moral norms, about their 'empty

ideality', drop out of the picture because after all, the factual on this occasion has to conform to theory (*TP* 8:276–7).[12]

IV. COMMON MORAL CONSCIOUSNESS: HOW SCOUNDRELS VINDICATE KANTIAN ETHICS

The paragraph that immediately follows upon the attempted deduction in *Groundwork* III is extremely interesting. Making good his promise to return to his starting point in the synthetic third section of the book (see *G* 4:392), Kant affirms that 'the practical use of common human reason confirms the correctness of this deduction', i.e. the problematic argument contained in the previous paragraph. He uses the famous example of the 'worst villain' or 'most hardened scoundrel' (*der ärgste Bösewicht*) to substantiate this claim:

> There is no one, not even the most hardened scoundrel, if only he is otherwise in the habit of using reason, who – when one presents him with examples of probity of purpose, of steadfastness in following good maxims, of compassion and of general benevolence (even combined with great sacrifices of advantages and comfort) – does not wish that he might also be so disposed. (*G* 4:454)

Anyone who is open to normative practical constraints in deliberation[13] would like to be morally good, even those who act abominably. Of course, the reason for this desire is to be found in the autonomy of the agent's own will. Crucially, it is induced and strengthened by the use of *examples* of decent conduct. Kant continues:

> But only because of his inclinations and impulses he cannot easily [*nicht wohl*] bring this about in himself; while at the same time he wishes to be free from such inclinations he himself finds so burdensome.

This sentence is commonly mistranslated. In Kant's day, the word *wohl* – the English 'well' – was still used as the adverb of *gut* (good) and carried overtones of pleasure and ease (this sense is still present in *das Wohl*, *Wohlergehen*, *Wohlgefühl*, and so forth). In other words, Kant is not telling us that the scoundrel *cannot* become a good man, as his translators would have it; which would be extremely surprising. Kant is merely saying that

[12] As a result, the second *Critique* must start with principles as the first data of practical philosophy (*CpV* 5:91.30, cf. 5:42. 22; 5:16.26). That concerns the order of exposition (which starts with principles in section 1 that are as yet unsubstantiated, and generally begins the Analytic with a chapter on principles), but also the philosophical foundations of the science (where there is a clearer distinction between analysis (sections 1–6) and substantiation, which follows in section 7).

[13] Note the disclaimer, which is not uncommon; see e.g. Kant's illustrations of the law-of-nature variant of the categorical imperative at *G* 4:421–3.

owing to his habits and likings this will be a very painful process. The couple of lines that follow are crucial:

> By this he [the scoundrel] therefore proves that, with a will free from impulses of sensibility he transfers himself [*sich versetze*] in thought into an order of things quite different from that of his desires in the field of sensibility, since from that wish he can expect no gratification of his desires, and hence no condition that would satisfy any of his actual or otherwise conceivable inclinations (for if he expected this, the very idea which elicits that wish from him would lose its pre-eminence), but only a greater inner worth of his person. But this better person he believes himself to be when he transfers himself [*sich versetzt*] to the standpoint of a member of the world of understanding, as the idea of freedom, i.e. independence from *determining* causes of the world of sense, without his choosing necessitates him to do, and from there he is conscious of a good will that, by his own admission, for his evil will, as a member of the world of sense, constitutes the law, the repute of which he recognizes in transgressing it. The moral ought is thus one's own necessary willing as a member of an intelligible world, and is thought by him as an ought only in so far as he considers himself at the same time as a member of the world of sense. (*G* 4:454.20–455.9)

Not only does Kant take up the language of transferral again – the second and third occurrences of *versetzen*; this passage is highly reminiscent of the many occasions in which the fact of reason is discussed in the second *Critique*. In the two gallows cases following section 6 (*CpV* 5:30), it is precisely such an example of honesty of purpose and steadfastness in acting on good maxims that makes us realize that we can, if necessary, act contrary to our strongest natural interest, and are thus autonomous agents: he who refuses the request of his prince to bear false witness against an honourable man, on pain of immediate execution, 'judges, therefore, that he can do something because he is aware that he ought to do so and cognizes freedom within him, which, without the moral law, would have remained unknown to him' (*CpV* 5:30). The consciousness of autonomy cannot be empirical in character, or, as Schiller aptly put it: *Du kannst, denn du sollst*.

Examples of moral behaviour play an important role in the chapter on respect or reverence as the motive of moral action; and they loom large in Kant's theory of moral education as presented in the final chapter of the *Critique of Practical Reason*, the Doctrine of Method, where the case of the princely bully – now identified with Henry VIII – is taken up again. Henry wants someone[14] to betray an innocent Anne Boleyn in court because he wants to be rid of her. The refusal of the witness is intended to inspire in a ten-year-old boy the wish 'that he himself could be such a man (though

[14] Kant is thinking of Sir Henry Norris. He was familiar with this story from Hume's *History of England*.

certainly not in such circumstances)' (*CpV* 5:156). It seems to me that this – note the caveat – is a perfectly reasonable and sane position.

This much the two books on moral theory have in common. However, by the time Kant writes the second *Critique* the role of common moral conviction has changed. In the *Groundwork*, it is used merely to *confirm* a deduction. In the *Critique* common moral judgment, somewhat more clearly stated, strengthened and christened the 'fact of reason', serves to defend and justify the categorical imperative. In short: whereas in 1785 ordinary moral consciousness was used to *support* a deduction of the categorical imperative, in 1788 it is meant to stand on its own to justify – as far as is possible – the principle of morality just by itself.[15]

V. MORALITY AS THE *RATIO COGNOSCENDI* OF FREEDOM

If this account is correct, there is no reason to believe that normative considerations were not intended to be the *ratio cognoscendi* of practical freedom prior to the second *Critique*, if in the somewhat weaker sense of being the ground of awareness (rather than cognition), though Kant was still prepared to go one step further and not make the justification of morality rest entirely on a 'fact of reason'. It is uncontroversial that freedom was used in the deduction of morality in Section III of the *Groundwork* – but that does not necessarily mean that Kant thought we had direct evidence of this freedom that is independent of the recognition of normative laws. That need not be the case if we conceive of the *Groundwork*'s deduction as a philosophically respectable account of what we should like to believe in (human autonomy of

[15] The vindication of objective moral principles now entirely depends on moral consciousness; moral thought presents us with 'a fact absolutely inexplicable from any data of the sensible world and from the whole compass of our theoretical use of reason' (*CpV* 5:43.5–7). The expression is mentioned already in the first *Critique*, where 'facts' (*facta*) of reason are the *deeds* of reason, e.g. reason's attempts to go beyond the limits of experience in its speculative use (*CrV* A760 f./B788 f.). This is obvious from the juridical context (cf. A764/B792, where 'facts' are associated with failed dogmatic attempts; these *facta* must be subjected to the censorship of reason). By contrast, the *factum* at *P*4:274.34 is something one already knows to be reliable on which the *Prolegomena* rest, whereas the first *Critique* does not rest on a *factum*. In the second *Critique*, the 'fact of reason' is either the 'deed of reason' or the 'product' of such a deed, something that is encountered as given in the consciousness of agents because reason has put it there. It is a *datum*, but not one that is given in experience or intuition. That explains Kant's cautious qualification that the moral law is given 'as it were' (*gleichsam*) as the fact of reason, or that freedom can be established 'as it were' by the fact of reason. This fact, as a product of a synthetic process of pure practical reason, cannot be justified theoretically. Rather, there is no other explanation of *this* datum other than that it is produced or given by reason. That is why Kant now calls it the *sole fact of pure reason* (*CpV* 5:31.33), which is said thereby to proclaim itself to be originally lawgiving (*sich als ursprünglich gesetzgebend ankündigt*).

the will) but is entirely unobvious in the face of that which we know (empirical cognition). Indeed, there are hints of normativity as the reason why we are aware of our freedom in Kant's earlier writings. For instance, the following statement can be found in the chapter on the resolution of the antinomy of free causality in the *Critique of Pure Reason*:

> Now that this reason has causality, or that we can at least interpret something of the sort in it, is clear from the imperatives that in everything practical we impose as rules upon our powers of execution. (*CrV* A547/B575)

But in this passage, it is imperatives in general – rather than narrowly the categorical imperative – that make us aware of our (presumed) freedom. Moreover, it is arguable that Kant intends to make the same point in the well-known passage in the Canon where practical freedom is said to be clear from the 'experience' of imperatives (*CrV* A802/B830). However, there is no indication in the writings before the *Groundwork* that the *ratio cognoscendi* of practical freedom must be narrowly moral – rather than broadly normative or imperatival. It is not difficult to see why this should be so. In 1781 Kant was as yet unaware of the fact that moral commands rest on a principle of autonomy, which puts them in a different sphere altogether. On the later picture, moral consciousness reveals to us that we are *absolutely* free to do as reason commands.[16] Already in the *Groundwork*, Kant argues that it is only 'the pure thought of duty and in general of the moral law, mixed with no foreign addition of empirical inducements' that 'first' makes reason aware of the fact that 'it can of itself be practical', i.e. that it is free (*G* 4:410.25–9).

VI. THE REVERSAL: KANT'S DEDUCTION OF FREEDOM

Kant never quite identifies freedom and the fact of reason, which is our awareness of the authority of the moral law. Still, there is a close connection. Freedom, while not identical, is *confirmed* (*bestätigt*, *CpV* 5:6.12) or can be proved (*beweisen könne*, *CpV* 5:104.33) by means of this fact. In the present section Kant returns to the role of morality as the *ratio cognoscendi* of freedom and accords the support freedom receives from these quarters the status of a deduction.[17]

[16] Of course, the Doctrine of Method was not revised for the 1787 edition.

[17] In Kant's text, the connection between the moral law as the *ratio cognoscendi* of freedom and freedom's being established by the fact of reason in a 'deduction' seems so close as to render the expressions synonymous. However, the previous section indicates that there are two potentially independent points at stake because a normative law may well make us realize that we are free without itself, strictly speaking, providing the justification for this belief.

The topic is obliquely introduced at the top of *CpV* 5:47, where he mentions the idea of deducing the existence of basic capacities of faculties (*Gundvermögen*). Even in theoretical philosophy, there is no direct proof that we possess basic cognitive faculties, but *experience* – i.e. empirical cognition – justifies us in our assumption that we do. The idea is that because we are capable of some activity or other we must possess a corresponding capacity. Kant calls this a 'surrogate' of a deduction (*CpV* 5:47.4). This route is not open to us when we seek to establish the supreme principle of practical reason because morality does not appear in experience.[18] But the idea of deducing a faculty reappears in the paragraph where the idea of a 'deduction of an inscrutable faculty' – namely freedom – is said to *take the place* of the deduction of the moral principle that certain philosophers for so long sought in vain:

> Something different and quite paradoxical [*ganz Widersinnisches*][19] takes the place of this vainly sought deduction of the moral principle, namely that conversely it does itself serve as the principle of the deduction of an inscrutable faculty. (*CpV* 5:47)

Then, in a rather surprising move, Kant argues that the very fact that we realize our self-subjection to the moral law *proves* that we are free. An argument that establishes the capacity for free action is needed at this point in the argument of the second *Critique* because freedom, as a keystone, is supposed to complete the dome of the edifice of pure reason – hence the second section on practical and theoretical reason that follows. The moral law thus provides freedom with 'a kind of credential' (*CpV* 5:48.1).[20]

Kant thus uses our recognition of moral commands to establish something different, namely that we possess a will that is free. The problem of

[18] However, in the *Critique of Judgment* Kant calls freedom a matter of fact that can be 'demonstrated in experience' (*CU* 5:468). Of course, freedom's being a *res facti* (*Tatsache*) is compatible with moral consciousness's being the *factum* (*Tat*); there is no change of mind here.

[19] Kant's *widersinnisch* usually means 'paradoxical', but there seems to be more than a hint of the literal meaning here: contrary, reverse, opposed to the usual way, pointing in the opposite direction (*Sinn*), as in his *widersinnig gewundene Schnecken* (*P*4:286), which are not very paradoxical in the standard sense. Freedom played a major role in the deduction of the categorical imperative in *Groundwork* III; now the categorical imperative in turn serves to verify our belief that we are free.

[20] Note that the 'credential of the moral law' (*das Creditiv des moralischen Gesetzes*) is the letter of confirmation the moral law issued for freedom, not a 'credential for the moral law' itself, as Rawls and Allison have argued in the wake of Beck's translation and commentary. (See Rawls, *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, MA: Harvard University Press, 2000), 261, 266–8; and Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), chapter 13.) Kant does not intend to return to the justification of moral norms at this point, or provide a coherentist argument. I argue for this point in detail in 'Das Creditiv des moralischen Gesetzes', *Studi Kantiani* 20 (2007), 111–15.

free will was the great desideratum of the first *Critique*. Now the reality of the concept of freedom is proven by an apodictic law of *practical* reason and can thus be regarded the 'keystone' (5:3.25) of the edifice of pure reason. So even if the idea of normativity as the reason why we know about the capacity of freedom is present in Kant's earlier writings, it is meant to do much more philosophical work once the fact of reason enters the scene; and the focus is now narrowly moral.

VII. THE OPENING PARAGRAPH REVISITED

This reconstruction of the strategy and purpose of the *Critique of Practical Reason* helps us to understand Kant's initial description of its method and purpose. The first paragraph of the preface now presents itself as an outright rejection of his justificatory endeavours in the *Groundwork*. Kant states that, the parallel case of the first *Critique* notwithstanding,[21] the second is entitled a *Critique of Practical* – not *Pure Practical – Reason*. In dispensing with the longer title, Kant renounces the core of the only part of a *Critique of Pure* [sic] *Practical Reason* hitherto published, namely Section III of the *Groundwork*, entitled 'Transition from the Metaphysics of Morals to the Critique of Pure Practical Reason', i.e. the section that contains as much of such a *Critique* as he deemed necessary for the project of founding a metaphysics of morals at the time. By contrast, the *Critique of Practical Reason* no longer pretends to offer an insight into the possibility[22] of the categorical imperative as a synthetic principle a priori. It merely intends to show '*that there is pure practical reason*' (*CpV* 5:3). For that purpose, 'it criticizes its entire *practical faculty*' (*CpV* 5:3), an expression that refers to the quasi-chemical procedure of separating the pure and the empirical adopted in the Analytic. If successful, Kant now argues, the second *Critique* 'has no need to criticize the *pure faculty itself*, in order to see whether by this reason is making a claim in which it merely presumptuously *oversteps* itself', as theoretical reason was shown to in the Dialectic of the first *Critique*. Recall that the perilous status of the claim that pure reason can be practical – that maybe (practical) reason must be restricted to the empirically justifiable – is evident throughout the *Groundwork*. Kant concludes his programmatic

[21] See Kant's notes on the project of a *Critique of Pure Practical Reason* at *G* 4:391. In 1784, Kant starts calling the first *Critique* a *Critique of Pure Speculative Reason* because he thought it no longer sufficient to ground both the metaphysics of nature and the metaphysics of morals; see Jens Timmermann, *Kant's 'Groundwork of the Metaphysics of Morals': A Commentary* (Cambridge: Cambridge University Press, 2007), *ad loc*.

[22] See *CpV* 5:46.21–2 for that formulation.

statement for the new *Critique*: 'For if as pure reason it is really practical, it proves its reality and that of its concepts by a deed [*durch die Tat*], and all subtle reasoning [*alles Vernünfteln*] against the possibility of its being practical is futile' (*CpV* 5:3). The 'deed' – or *Factum* – in question is the act of reason's giving the moral law to the agent.[23] Once again, Kant's rejection of the foundational project of the *Groundwork* is manifest. Subtle arguments against the possibility of morality are mentioned in Section II, and addressed explicitly in Section III, where he seeks to explain how morality as grounded in a synthetic principle a priori is possible. In the new *Critique*, these sceptical worries are brushed aside on the grounds that morality is – and can indirectly be shown to be – real. Its possibility can safely be inferred from its reality. The proof of the moral pudding is in the eating. Or so Kant argues.[24]

VIII. THE DUISBURG *NACHLASS*

In a striking note from the Duisburg papers, Kant explicitly turns to the task of a 'critique of practical [sic!] reason'. The passage – contained in Reflection 7201 (19:274–6), from which all subsequent quotations are taken – appears to be a draft that Kant decided not to include in the final manuscript of the second *Critique*. Both the proposed book title and the general tone of the note suggest that it reflects his thoughts on the method of justifying the moral principle at a time when he had already abandoned the strategy pursued in *Groundwork* III but was not quite prepared to drop the project of a deduction altogether – which explains its omission from the published work. A critique of practical reason, he argues, relies on

> the differentiation of empirically conditioned practical reason from pure and yet practical reason and asks whether there is such a thing as the latter. The critique cannot have insight into this possibility a priori because it concerns the relation of a real ground to a consequence; consequently, something must be given which can arise from it alone; and from reality possibility can be inferred. Moral laws are of this sort, and this must be proven[25] in the same way we proved the representations

[23] See M. Willaschek, *Praktische Vernunft. Handlungstheorie und Moralbegründung bei Kant* (Stuttgart: Verlag J. B. Metzler, 1992), 179, who also emphasizes the inference *ab esse ad posse*, 229.

[24] Thus in the *Groundwork*, the 'that' (of the categorical imperative) was finally established by means of showing 'how' (a categorical imperative is possible). In the *Critique of Practical Reason*, the question 'how' (a categorical imperative is possible) is rejected on the grounds 'that' (the categorical imperative is real). An explanation over and above the fact of reason is neither possible nor required.

[25] I am using Fred Rauscher's translation in Kant, *Notes and Fragments*, trans. Paul Guyer, Curtis Bowman and Frederick Rauscher (New York: Cambridge University Press, 2005). H. Allison mistakenly reads 'and these [laws] must be proven', etc. (*Kant's Theory of Freedom*, 234) and then proceeds to explain that the validity of *particular* moral laws 'is to be established by showing that the principle in which they are grounded, the moral law, is a product of pure practical reason' (235). But in the

of space and time as a priori representations, only with the difference that the latter concern intuitions but the former mere concepts of reason.[26]

By the time he completed the manuscript of the second *Critique*, he clearly still believed that insight into the possibility of pure practical reason was impossible, probably for the reason mentioned: the 'consequence' in question must be moral consciousness, the 'real ground' the synthetic act of self-legislation. Yet in the *Critique* as we have it the project of 'proving' moral laws was further weakened; the possibility of a deduction is explicitly denied. Validating moral laws is no longer separable from the process of distinguishing pure and empirical determining grounds, which in the note still appear to represent distinct stages ('differentiation' of pure and empirical vs. 'existence' of pure practical reason). Note in particular that the 'fact of reason' is not explicitly mentioned.

However, themes from Reflection 7201 reappear in the work as it was eventually published. In the second *Critique*, moral principles still fill the place that space and time occupy in the first.[27] Perhaps more importantly, the note contains the characteristic inference from the reality to the possibility of morality, which is ubiquitous in the *Critique of Practical Reason* and renders the deduction of the categorical imperative superfluous. The thing that is 'given which can arise from [pure practical reason] alone' is our consciousness of the supreme normative force of moral commands, later called the 'fact of reason', i.e. the product of reason, a *Tatsache*. The strategy is reminiscent of the thought that underlies the gallows case at *CpV* 5:30: that our consciousness of an unconditional moral law can be explained not by natural causes but only with reference to our actually being members of an intelligible world and thus subject to the moral law.[28] But the argument that pure practical reason is the only possible explanation of our moral

second *Critique*, Kant is not concerned with the validity of individual laws but rather with the validity of morality as such, and there is no attempt to validate particular moral laws with reference to the purity of the moral law itself.

[26] See Kant, *Notes and Fragments*, 464. Kant adds: 'The only difference is that in theoretical knowledge the concepts have no meaning and the principles no use except with regard to objects of experience, while in the practical, by contrast, they have such wider use, namely they apply to all rational beings in general and are independent of all empirical determining grounds, indeed, even if no object of experience corresponds to them, the mere manner of thinking and the disposition in accordance with principles already suffice'.

[27] See Reinhard Brandt's explanation of the fact that the second *Critique* turns the structure of the first on its head: in the shape of the 'fact of reason', principles take the place of space and time, but both are generated by the self, *selbsterzeugt*. See Brandt, *Die Bestimmung des Menschen bei Kant* (Hamburg: Felix Meiner Verlag, 2007), 364. There is a hint of this at *CpV* 5:46.6, as mentioned above.

[28] This in turn reminds us of Descartes's version of the ontological proof in the *Meditations*, according to which we could not possess the concept of God if he had not himself given it to us. Presumably, Kant would not like to be seen to endorse that kind of argument.

awareness is never quite explicit in the published work.[29] Evidently, Kant no longer thinks it constitutes a *proof*.

IX. CONCLUSION

In conclusion, let us examine the only sentence in the *Critique of Practical Reason* in which Kant explicitly discusses the *Groundwork* and its relation to the new book. The second *Critique*, he argues in the preface, does indeed presuppose

> the *Groundwork of the Metaphysics of Morals*, but only insofar as it provides preliminary acquaintance with the principle of duty and states and justifies [*angibt und rechtfertigt*] a specific formula of it;* otherwise it stands on its own. (*CpV* 5:8)[30]

A little earlier, Kant reminds us that 'matters that have been decided [*einmal abgeurteilte Sachen*] should only be referred to and not raised again' in the new book (*CpV* 5:7).[31] How does this affect the relationship between his two foundational works in moral philosophy? There are (at least) two plausible ways of reading Kant's cautious endorsement of the earlier book. The most straightforward interpretation takes it to be restricted to the analytic part: the expressions 'preliminary acquaintance' and 'specific formula' refer to Sections I and II of the *Groundwork* respectively, whereas Section III is quite explicitly excluded because of Kant's loss of faith in the deduction of the categorical imperative.

There is, however, another possible interpretation. Recall that the *Groundwork* itself defines its remit as the 'identification' (*Aufsuchung*) and the 'establishment' (or 'corroboration', *Festsetzung*) of the supreme principle of morality (*G*4:392), terms commonly taken to refer to Sections I and II on the one hand and Section III on the other.[32] The above statement of the

[29] The only thing that can account for our conviction that we can do what we ought to do, no matter what, is that we are members of a purely intellectual world. On his way to the gallows example at *CpV* 5:30.3–9, Kant addresses the question of how our awareness of the moral law is possible. We can, he argues, become conscious of pure practical laws by paying attention to the *necessity* with which reason prescribes them to us. (For the parallel in the theoretical case, see *CrV* B5–6; for the separation of pure and empirical elements, see the chemical analogy in the Critical Elucidation, *CpV* 5:92–3.) There is a hint of this already in the preface to the *Groundwork*, 4:389.

[30] The footnote defends the categorical imperative as a formula against G. A. Tittel's charge that the *Groundwork* provides 'merely a new formula': Gottlob August Tittel, *Ueber Herrn Kant's Moralreform* (Frankfurt and Leipzig: Pfähler, 1786).

[31] Kant feels the need to justify his return to topics that had been dealt with in the *Critique of Pure Reason*, such as the use of the categories and the topic of freedom and determinism, which is revisited in the Critical Elucidation.

[32] For this point, see Allen Wood's 'Preface and Introduction', in Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft*, 35.

relevance of the *Groundwork* to the second *Critique* – 'stating' and 'justifying' a specific formula of the principle of duty – can be construed to echo that description. After all, in the earlier work morality is corroborated as a principle of autonomy, which is a specific formula of the moral principle.[33] We saw above that the second *Critique* declares the deduction of the *Groundwork* to be unnecessarily and unrealistically ambitious, while large parts of Section III – such as the theory of transferral into a realm of intelligible laws – stay intact; and what was previously used to *confirm* the deduction – common moral consciousness – now takes its place. As one would expect, these arguments are indeed 'referred' to in the second *Critique* but by and large not 'raised again'. If so, even without the actual deduction, the *Groundwork* is still sufficient to justify morality, even though Kant is retreating in the later work.

There is no need to decide the matter at this point. In either case, the second *Critique* was by no means intended to replace the *Groundwork*, even though it rejects the central move of Section III. This explains why it contains three (or so) passing references to the project of the *Metaphysics of Morals*, which is never explicitly discussed or even mentioned.[34] The *Critique* does not need to lay the foundations of this discipline because that has already been done – its task is to complete a readjusted project of the 'critique of reason'. Perhaps surprisingly, the *Groundwork* suffices to ground the Kantian ethical system even if we give up on the project of a moral deduction and surgically remove the offending lines (*G* 4:454.6–19) from its final section, to let common moral consciousness take centre stage. The less ambitious attempt of moral justification may well be the more persuasive.[35]

[33] A description of Section II as stating and justifying a specific formula would actually be rather odd. The formula is already stated towards the end of Section I; Section II provides several *additional* formulations; and the justificatory project is consistently put off until the final section (see e.g. *G* 4:420 and 429fn.). If the *Groundwork* – minus Section III – is meant to contain the 'justification' of a 'formula', it would thus have to be limited to proving that the formula of the moral principle has to take the shape it actually takes, without establishing its validity (unless, again, Kant now expects the two analytic sections implicitly to furnish a proof because the clear division between analysis and synthesis is no longer in place – but that would probably be going too far).

[34] *CpV* 5:8.23, 5:161.25–31 and 5:12.5, which alludes to the bipartite philosophical system that follows the *Critique*. See also 5:67, which promises a new system of morals organized according to the categories of freedom.

[35] Over and above the contributors to this volume, I should like to thank the organizers and participants of the 2007 UK Kant Society Annual Conference as well as Bernd Ludwig (Göttingen) and Michael Wolff (Bielefeld) for many fruitful exchanges.

CHAPTER 5

The Triebfeder *of pure practical reason*

Stephen Engstrom

I.

At the beginning of the *Critique of Practical Reason*, Kant tells us that its aim is merely to show that there is pure practical reason, or that pure reason is practical. It won't, however, be by an analytical argument that this task is discharged: if pure reason is actually practical, he adds, 'it proves its reality and that of its concepts through the act [*durch die Tat*]' (*CpV* 5:3).[1] Though somewhat cryptic, this remark anticipates the doctrine of the *Factum der Vernunft* – the 'fact' of reason – presented in Chapter I of the Analytic of Pure Practical Reason, where Kant explicitly denies that the proposition that pure reason is practical can be demonstrated from antecedently available premises (*CpV* 5:31). Recognizing that pure reason's practicality cannot be established by conceptual argument, Kant endeavours instead to employ such argument to reach a position from which we can gain a philosophical recognition of this practicality – this act – of pure reason, a recognition based in the consciousness of the fundamental law of pure practical reason, a law we have recognized all along, under another name, as the moral law.

But Kant also denies that this *Factum* is an empirical fact. While we cannot establish pure reason's practicality by conceptual argument, neither is it through any 'experience' of this practicality that we first recognize the moral law. If it were by such means that we came to know this principle, it would have to be through a receptive capacity. We would need something like a perception or intuition of it, but not a rational perception or an intellectual intuition, for that would be wholly spontaneous. The perception would rather need to lie in the exercise of a moral sense or feeling. But no sense impression or sentiment could acquaint us with a principle, especially one having strict universal validity, for all rational beings.

[1] Translations are my own, though I have consulted the commonly used English translations.

Such epistemic considerations are enough, Kant thinks, to rule out appeals to a moral sense. The unrestricted validity of basic moral requirements constitutes their cognitive standing, which cannot be secured by a sensible capacity. The moral-sense philosophers, however, make their case largely on motivational grounds. In criticizing rationalist accounts, Hutcheson and Hume argue that moral judgments must be able to move us to act.[2] 'Reason', Hume famously claims, 'is perfectly inert.'[3] To account for the motivating power of moral judgment, we must acknowledge the role of sentiment and human sensibility. Kant is not unmoved by these considerations. He does not, of course, accept Hume's assumption that reason has no practicality, but he recognizes that his aim of showing the practicality of pure reason obliges him to provide an account of how reason, through the law it prescribes, comes to be efficacious in a human subject, where the exercise of choice is subject to sensible impulses. Hence, the full execution of the second *Critique*'s task calls for more than the exposition (and deduction or credential) of the highest principle of practical reason offered in Chapter I of the Analytic of Pure Practical Reason. Thus, after determining the concept of the object of pure practical reason in Chapter II, Kant takes up the motivational issue in Chapter III, which deals with what he terms the *Triebfeder* of pure practical reason.

II.

That Kant is investigating pure reason's power to move us to act is not obvious from the currently standard English translations of the title of this chapter, which render *Triebfeder* somewhat misleadingly as 'incentive'. In ordinary usage, 'incentive' typically signifies what incites to action. It refers to some *object* that attracts or repels rather than to something subjective in the agent. When we say that a certain tax scheme gives taxpayers an incentive to save more of their income, it is not the desire to save that we are

[2] Francis Hutcheson, 'Illustrations on the Moral Sense', in *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*, ed. Aaron Garrett (Indianapolis: Liberty Fund, 2003), Section I; David Hume, *An Enquiry Concerning the Principles of Morals* (Oxford: Clarendon Press, 1975), Section I and Appendix I. It seems reasonable to suppose, however, that doubts about the capacity of reason to move us to act are at least in part supported by doubts about whether reason is capable of yielding substantive moral principles. See Christine M. Korsgaard, 'Skepticism about Practical Reason', in her *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press: 1996). Ultimately, Hume's doubts reflect his across-the-board scepticism about reason, which does not credit reason with the capacity to furnish substantive principles even for theoretical cognition.

[3] *A Treatise of Human Nature*, 2nd edn (Oxford: Clarendon Press, 1978), III.i.1; cf. II.iii.3.

normally thinking of as the incentive, but lower tax payments. Incentives are carrots and sticks, not the desires and aversions they may trigger. But when Kant speaks of a *Triebfeder*, he almost always[4] has in mind something in the subject that generates the action, rather than an object or circumstance that prompts it. It is, he says, 'the subjective determining ground of the will of a being whose reason is not already in virtue of its nature necessarily in accordance with the objective law' (*CpV* 5:72). And the only things he designates as moral *Triebfedern* in Chapter III – the moral law and the feeling of respect for it – are sources of willing and action, not their ends, outcomes, or effects.[5]

In its original literal meaning, *Triebfeder* refers to the mainspring of a clock. But Kant is not thinking of anything mechanical,[6] and the mechanistic meaning itself rests on a more basic biological idea of an inner driving force. So we should think of a *Triebfeder* as an inner spring or source of choice and action. In fact, Hutcheson, Hume, and other eighteenth-century British authors use 'spring' in much the way Kant uses *Triebfeder*.[7] The term is a bit old-fashioned and not a perfect fit for every passage, but there is much to be said for it. It suggests unambiguously an inner source rather than an object.

There is a different though related way of interpreting *Triebfeder*. According to Lewis White Beck, the term refers to 'the dynamic or conative factor in willing', as opposed to the cognitive or representational factor.[8] On this view, the objective determining ground of the will is a certain representation, whereas the *Triebfeder*, or subjective determining ground, is the representation's efficacy, the 'force' it has to determine the will. The contrast is one of aspect. The *Triebfeder* isn't so much the spring itself as it is the 'spring' that's in the spring. It's the pressure, the conatus, the living force, the potential energy – whatever you want to call it – through which the

[4] Not quite always. At *CU* 5:266, for instance, the agreeable, or the object of sensible desire, is said to be the *Triebfeder* of such desire.

[5] The most common alternative translation is 'motive'. This word has in its favour that it can be used in a subjective as well as an objective way: we can say that a politician's motive in embracing a popular cause is to win the election, but we can also place it in ambition or a desire for power. 'Motive' does not solve the problem, however, since we need it to translate *Bewegungsgrund*, which Kant seems to regard as the German equivalent of *motivum* (see 28:587) and which differs in sense from *Triebfeder* (*G* 4:427).

[6] Though he is of course aware of the mechanical metaphor. See the use of *Feder* and *aufziehen* in the Menzer edition of student notes of his lectures on ethics (Gerd Gerhardt (ed.), *Immanuel Kant: Eine Vorlesung über Ethik* (Frankfurt am Main: Fischer Taschenbuch Verlag, 1990), 55).

[7] This usage is registered in T. K. Abbott's nineteenth-century translations of Kant, which sometimes use 'spring' as an alternative to 'motive'.

[8] *A Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960), 216; cf. 90.

mainspring drives the gears in the clock. So understood, *Triebfeder* might be best translated as 'force', or perhaps as 'determining force', 'motivating force', or 'driving force'.[9]

Yet while there is surely something right in this interpretation, it seems unable to accommodate Kant's insistence that the moral law is *itself* to be the *Triebfeder* of the human will, and that 'the objective determining ground must always and all by itself be the subjectively sufficient determining ground of action' (*CpV* 5:72). Even if Kant does occasionally use *Triebfeder* to refer to the conative or causal aspect of a representation, this insistence on an identity of subjective and objective determining grounds suggests a different distinction. It indicates that he is contrasting objective and subjective in the way he does when speaking of practical principles, where he also holds that what's objective can also be subjective. An objective principle, he says, is one on which we *ought* to act, but it is also a subjective principle, or a maxim, in so far as it is a principle on which a given subject *does* act (see *G* 4:420n.). Similarly, I suggest, the moral law, conceived as the will's objective determining ground, is the representation of how the will ought to be determined; that same law, conceived as the subjective determining ground of a given subject's will, is that same representation *actually determining* that subject's will. So far as the moral law is a subjective determining ground of the will, it is a spring. It is a cognitive representation, but one that has force, or efficacy.[10] Even when he speaks of the feeling of respect as a spring, Kant has this identity in view: respect for the law serves 'merely as a spring to make this [sc. the objective moral law] a maxim in oneself'; it is just 'morality itself, regarded subjectively as a spring' (*CpV* 5:76).

III.

In his chapter on the springs of pure practical reason, Kant undertakes to explicate, in the light of the *Factum* of reason, *how* pure reason is practical in the case of the human being, and more generally in that of a finite subject having a share in this *Factum*. What distinguishes this case is a certain lack of perfection in the way such beings are related to the fundamental law of pure practical reason, the law identified as the moral law in section 7 of Chapter I.

[9] 'Motivating force' is one of the terms Greene and Hudson use for *Triebfeder* in their translation of Kant's *Religion*. 'Driving force' is suggested by the presence of *Trieb* in *Triebfeder* and by Kant's Latin gloss, *elater animi* (I am grateful to Alp Aker for helpful discussion of the meaning of this expression).

[10] A qualification is needed here, which I will touch on presently, namely the presupposition that there are obstacles to the moral law within the subject.

As Kant puts it, here 'reason is not already by virtue of its nature necessarily in conformity with the objective law' (*CpV* 5:72). This lack of perfection is ascribed to reason itself, but since it pertains to reason's *practical use*, we can place it more specifically in practical reason. And since Kant identifies practical reason with the will, we can also attribute it to the will, as Kant himself does when, in the *Groundwork of the Metaphysics of Morals*, he marks a distinction between perfectly and imperfectly good wills (*G* 4:412–14).

Because the lack of perfection under consideration resides in the relation practical reason, or the will, bears to its own objective law rather than in its relation to the action that results, it is clear that Kant is not just thinking of the deficiency Aristotle called *akrasia*, a mere defect in respect of execution, which can be present even where the agent's end, deliberation, and choice are all in conformity with that law. And because the relation at issue is one of conformity, it is likewise clear that he is not contemplating a mere deficiency in the practically deployed power of judgment, owing to which a person might, despite recognizing the law, occasionally fail to see how it is to be carried out and so be left suspended, like Buridan's ass, incapable of deciding what to do. Rather, he has in mind a will whose nature does not rule out the possibility that it might fail to agree with that law, by willing what it forbids or not willing what it requires.

To clarify further the sort of imperfection Kant has in view, two points are worth noting. First, in speaking of a rational being 'whose reason is not already by virtue of its nature necessarily in conformity with the objective law', Kant is not saying that the nature of this reason could by itself alone give rise to a lack of such conformity. He well recognizes that there's no way even to conceive of a rational being in whom the exercise of reason might not agree with its own objective law except by supposing that in such a being reason is exposed to something outside it that can affect it – namely the subjective or sensible component of this being's nature (see *R* 6:24, 36). Thus, later in the chapter he says, 'the recognition of the moral law is the consciousness of an activity of practical reason from objective grounds, which fails to express its effect in actions *only because* subjective causes (pathological) hinder it' (*CpV* 5:79; my emphasis). And much earlier in the Analytic he indicated that it is where the will of a rational being is 'pathologically affected' that a conflict can arise between its maxims and the objective practical laws it itself cognizes (*CpV* 5:19); 'if reason wholly determined the will, the action would occur without fail' according to reason's objective rule (*CpV* 5:20). In describing the affection as pathological, Kant signals that he does not mean the sort of affection on which theoretical knowledge depends, where objects outside the mind, by acting

on the outer senses, produce in us the sensations requisite for our experience of them. He means the kind of affection that constitutes sensible desire, an affection of which we are immediately conscious through feeling – that is, through our capacity to be pleased or displeased. Such affection first arises with that very consciousness: to the extent that the experience of some object is pleasing, there arises a sensible desire for the object (see *CpV* 5:21). Thus, the human will is pathologically affected in that the component of human sensible nature that can affect it is determined through pleasure and displeasure. In sum, the chapter on the springs of pure practical reason is concerned with expounding how pure reason is practical in a rational being whose will has a nature that, on account of its exposing that will to pathological affection, does not already, by virtue of itself alone, entail the will's necessary conformity with its own objective law.

Second, although the imperfection at issue depends on the will's being pathologically *affected*, this *im*perfection, positively described, is a kind of perfect*ability*, in the form of a capacity for self-determination, and the will's possession of this form of perfectability entails that it is not *determined* through being pathologically affected. Since the imperfection belongs to the will's relation to its own objective law, it can also be said to be located in the will's freedom. For freedom, so far as it belongs to the will's nature (as the *can* implied by the *ought* of the objective law), is the will's *capacity* to determine itself to be in necessary conformity with that law. On account of the will's exposure to pathological affection, this necessary conformity is not already given by virtue of the will's nature alone, but realized only through self-determination. This freedom is therefore imperfect in the sense that it is not already the actuality of necessary conformity, but only the capacity for such conformity. Yet though a mere capacity, it is still a capacity, and moreover a capacity for self-determination. So given that the will is free, pathological affection does not entail determination; were the will determined through such affection, the possibility of its being determined by its own law – the possibility of self-determination in which its freedom consists – would be excluded. Indeed, Kant states explicitly that though the will of a finite rational being is pathologically affected, it is not thereby determined (*CpV* 5:32). Thus, any lack of necessity in the will's conformity with its objective law could no more be a natural and necessary consequence of the will's being pathologically affected than it could be something to which the will's nature by itself alone gives rise.

From the points just noted, we can see that where a will is by nature both free and pathologically affected, necessary conformity with its objective law is neither precluded nor already given by virtue of its nature alone.

Necessary conformity cannot be naturally impossible in a will that is free, nor can it be a matter of natural necessity in a will that is pathologically affected.

We can also see that since it belongs to the *nature* of the will of a finite rational being to be both free and pathologically affected, it is not possible to express in terms of bare freedom and affection alone the difference between a particular will that is in necessary conformity with its objective law and one that is not. To capture the difference, we need to introduce a further notion Kant sometimes employs, that of *influence*.[11] We can then say that whether a will fails to conform to the objective law is a matter of whether the sensible impulses to which it is exposed in being pathologically affected also influence it. Influence requires affection but where freedom is involved does not reduce to it. For if a will is free, its being influenced always depends in addition on its allowing itself to be influenced, whereas it is affected willy-nilly, simply on account of its nature. The notion of influence accordingly lies in a position intermediate between those of affection and determination, standing to them respectively as the concept of actuality stands to those of possibility and necessity: pathological affection is the condition without which influence is not even possible, and were determination through such affection not excluded by the will's freedom, it would be the condition under which such influence is necessary.

Given that a human will is pathologically affected, so that the mere recognition of the objective law does not by itself already yield this will's necessary conformity with it, some further account is needed to articulate how that law determines such a will, how the moral law, which Kant has already identified as the *objective* determining ground of the will, comes to be also its *subjective* determining ground. In a being whose will does not already by its nature necessarily conform to the moral law, such conformity will be an *achievement*, a matter of acquired disposition, a matter of *virtue*. 'The greatest moral perfection of the human being', Kant later says, is 'that the law be not merely the rule but also the spring of actions' (*MdS* 6:392). What is needed, then, is an articulation of how this perfection of necessary conformity is achieved in such a being. Or, as Kant puts it, the task is 'carefully to determine in what way the moral law becomes a spring and, inasmuch as it is, what happens to the human faculty of desire as an effect of that determining ground upon it' (*CpV* 5:72). Kant makes clear, however, that he will not be undertaking an empirical investigation: what is needed, he says, is to indicate *a priori* what the moral law effects, or must effect, in

[11] *Einfluß*; see, e.g., *G* 4:458, *CpV* 5:118.

the mind (*CpV* 5:72). Since the act of pure reason is not an occurrence encountered in experience, but a law, we can have no ground for tracing an actual mental determination or affection to this act as its cause unless we can determine in advance, a priori, that pure reason, so far as it is practical, must produce such an effect.

Kant emphasizes from the start that the nature of the problem entails that the conformity of will to law cannot be secured indirectly. If the conformity is to have the required necessity – the necessity figuring in virtuous conduct – the law's determination of the will must be immediate. No means–end or consequentialist account could satisfy that requirement. If my choices conform to the moral law only through a judgment that such conformity will promote my happiness or further some other antecedently determined object of my will, the conformity will lack the requisite necessity, and my actions will be legal, but not moral.

Kant is equally emphatic in rejecting another approach, which also runs afoul of this immediacy requirement, but in a different way. The very character of duty and of obligation, he holds, prevents us from basing our account on the idea of a moral sentiment that has the moral law as its object. One apparent reason why Kant would regard such an account as unacceptable is that it would leave reason's practicality dependent on a special kind of feeling, a special property of human nature (see *G* 4:425). But Kant thinks this view founders on a more fundamental problem: 'there is no *antecedent* feeling in the subject that is attuned to morality. For that is impossible, since all feeling is sensible, whereas the spring of the moral disposition must be free of all sensible conditions' (*CpV* 5:75). As 'the mere form of law', the moral law 'can be represented only by reason' (*CpV* 5:29); to suppose that a law can be felt would be 'to make into an object of sensation [or feeling] what can only be thought through reason', and this, Kant says, borders on a 'plain contradiction' (*CpV* 5:38–9). If we turn Kant's point around, we can see that, to this extent, he is in agreement with Hume. So far as we are asking just about reason's capacity to produce an immediate effect on feeling, the answer must be that reason is 'perfectly inert'. So although Kant insists that morality requires that the moral law immediately determine the *will*, he also denies that this law can immediately determine *feeling*.

In rejecting such an appeal to moral feeling, Kant does not, of course, mean to reject the notion of moral feeling altogether or to deny that the moral law can influence feeling indirectly. Indeed, the centre-piece of his account of the moral law as a spring is a description of how this law can 'exercise an effect on feeling' (*CpV* 5:74). The challenge is to show, in this description, that 'no special kind of feeling need be assumed, under the

name of a practical or moral feeling, as preceding the moral law and lying at its basis' (*CpV* 5:74–5).

Kant does not say why indicating a priori what the moral law effects in the mind requires an explanation of the law's effect on *feeling* – that is, on the capacity to feel pleasure and displeasure. But if we bear in mind his repeated claim that it is through the pleasure and displeasure involved in our experience of objects that we are determined to have inclinations, or sensible desires, regarding them (*CpV* 5:21–2), an answer readily suggests itself. In so far as the human will fails to conform to the moral law, the failure is due to the influence that inclinations have on it, so if the moral law is to be a subjective determining ground of the will, it must somehow limit or modify those inclinations, and to do this it must exercise some effect on the feeling on which they are based.

But we must also bear in mind the other point just noted, that the moral law has no direct effect on feeling. As the law of the will, its immediate efficacy extends to the will alone; any further effect must arise through the willing it determines. Also to be kept in mind is that the will, according to Kant, is just practical reason. Since reason is the cognitive power, willing is a form of cognition, what Kant calls *practical cognition*, which lies in practical judgments as to what one is to do or ought to do, the most basic of which are the practical principles discussed in Chapter I. So it is in such judgments that the moral law determines the will, and through such judgments alone will this law have any further effect in the mind.

IV.

How then does the moral law become a spring? What effect does it produce in the mind? Kant begins his a priori description in much the way we might expect. He again draws attention to a point made earlier in Chapter I, that precisely because it is a law, the moral law requires the will's unconditional conformity with it, regardless of whether one's inclinations favour the required exercise of the will, whether they are indifferent, or whether they are all opposed. This necessity belongs to the formal character of the moral law's determination of the will and is the basis, as Kant claimed in Chapter I, of our consciousness of our freedom, of our capacity to conform our will to the law, regardless of the tendency of our inclinations.

Now from this point about the formal character of obligation, Kant notes, we can see a priori that a negative effect on feeling must follow – namely a feeling of *pain*. This is a familiar observation; Kant makes it in many places. But let's ponder it for a moment. Consider how things will

have to look from the point of view of one's inclinations. Inclinations arise from pleasing experiences. When I find the experience of some object pleasing or enjoyable, the pleasure, according to Kant, is just my awareness of the effect this experience is having on my desire; it's my awareness, as he puts it, of the determination of the (lower, pathologically determinable) faculty of desire, through which I acquire a sensible desire for that object (*CpV* 5:21). If I find something pleasing, I want more of it. So this pleasing experience determines me to have this sensible desire. This determining ground, obviously, is fundamentally distinct from the moral law. Through my acquiring such a desire, my capacity to feel pleasure and displeasure is modified accordingly. Since I can now anticipate that further experience of the object will be pleasing, and deprivation of it displeasing, any judgments I might make that have a bearing on my prospects for attaining this object – grounds for hope that I will acquire it, or for fear that opportunities for its enjoyment are being closed off – will also produce effects on my feeling, in the form of joys of anticipation and pains of frustration and disappointment. Since on account of its formal character the moral law can be recognized to be a different and independent determining ground of the will, it is clear that so far as my will is determined by this law, my choices may not augur well for the attainment of the agreeable objects of my inclinations. So even if, in a given instance, the moral law's determination of my will would not block the attainment of those objects, and even if, adopting a more general vantage point, I regard such blockage as unlikely or wholly hypothetical, the *form* of my will's determination by this law is enough to upset my inclinations.

These considerations apply to every inclination. Indeed, they rely on no special assumptions about the constitution of the capacity to feel in this or that finite rational being, but rest only on a general understanding of the way sensible desire is related to this capacity in such a being. So we can see in advance, a priori, that even though the inclinations may compete with one another, they are *all* pained at the prospect of the will's being determined by the moral law. Thus, as Kant says, the moral law, 'by thwarting all our inclinations, must produce a feeling, which can be called pain' (*CpV* 5:73). And the feeling, as we can also see, arises *indirectly*, through the awareness of possible conflicts within the faculty of *desire* between sensible desires, which belong to the lower faculty of desire, and the moral law, the law of the will; it is the prospect of the possible frustration of the inclinations that engenders the pain.

But Kant does not content himself with this observation about the moral law's *negative* effect on feeling. Far from being the whole of his account of

the moral law as a subjective determining ground, it is but the beginning. Why is it not enough? Kant does not explain, but the following considerations seem pertinent.

First of all, the pain is not adequate to the moral law in the sense that it involves no distinctive reflection of it. Such a feeling would be produced just as well by any other obstacle, such as a rain shower that spoils one's plans for a picnic. Indeed, the inclinations themselves can be such obstacles to one another. The feeling reflects the nature of the inclinations rather than the moral law; nothing has yet emerged that qualifies as moral feeling in any interesting sense.

We can carry these considerations further, by noting that Kant's description does not bring into view any judgment regarding the goodness or worth of these sensible desires or their objects. As so far characterized, these desires involve no more than judgments that their objects are agreeable, or pleasing to experience. But as Kant made clear in Chapter II, the judgments governed by the moral law concern not the agreeable but the good. It is as if the inclinations and the moral law are entirely unrelated, operating not as two components of a single faculty of desire, but in complete independence from one another, except of course that they must compete for control of the will. To the extent that the moral law determines the will, the inclinations are simply thwarted, as if by some external power. The point is not that Kant's characterization is inaccurate. It correctly registers the fundamental difference between the determining grounds of sensible desire (pleasing experiences) and of the will (the moral law). It is rather that the description seems incomplete.

The point could also be put by saying that if there were no further effect on feeling, then the moral law functioning as a subjective determining ground could never bring us to the condition of virtue. The result would be Aristotelian continence rather than virtue proper. So far, Kant's account has merely described how the moral law can infringe the inclinations. It has not offered a prospect of a deeper influence of reason upon the mind, one that would so constitute sensibility that the inclinations themselves, as they arise, might be amenable to being ordered, trained, and disciplined by reason.[12] Kant has yet to indicate how the moral law can have a positive global effect on sensibility – on the very capacity to feel, not just on its particular operations on particular occasions.[13]

[12] Kant makes much the same point in describing the 'negative effect' of the moral law at *CpV* 5:78.

[13] In this respect, Kant's account of the *Triebfeder* of pure practical reason parallels his account of space and time in the first *Critique*. He is concerned not just with the effect the intellect has on this or that occasion of sensible affection, but more fundamentally with the effect it has on sensibility itself.

In sum, Kant has so far identified a pain we can anticipate just by considering our sensible nature. In the next stage, he will point out a feeling we can anticipate once we consider our practical capacity to judge and to cognize in connection with our sensible nature.

v.

To extend his account, then, Kant needs to consider the inclinations in respect of the common relation of dependence they all bear to our sensibility, or to feeling, and to take account of the regard we have for them so far as they stand in this relation. Thus, he begins by directing attention to all of one's inclinations taken together, remarking that they can be brought into a tolerable system, where their satisfaction would constitute one's own happiness. Taken together, he says, they 'make up' *selfishness* (*Selbstsucht*, *solipsismus*), or *self-regard*, as it is usually called in English translations.[14] The thought here is plainly not that selfishness reduces to a mere bundle of inclinations. The name itself implies a certain concern for oneself, as the pathologically affected subject to whom all those inclinations belong; it also suggests that this concern constitutes an inclination to seek one's own happiness.

Kant then locates under this generic heading of selfishness two important concepts, *self-love* (*Selbstliebe*, *Eigenliebe*, *philautia*) and *self-conceit* (*Eigendünkel*, *arrogantia*). Self-love is initially characterized as 'a predominant *benevolence* toward oneself', self-conceit as '*satisfaction* with oneself'. Each is said to be a source of claims. Self-conceit is described as a source of 'claims of self-esteem' and of a 'pretension' to the 'worth' of one's own person (*CpV* 5:73). And self-love, along with self-conceit, is characterized as a striving or a propensity of 'our pathologically determinable self' to secure a certain validity for its claims, which are also referred to as maxims (*CpV* 5:74). Self-love and self-conceit are thus not to be placed on a footing with the sensible desires we attribute to animals generally, including those lacking reason – appetites for food, drink, and sex, aversions to extremes of temperature, and so forth. Self-love and self-conceit are closely bound up

[14] It is doubtful that 'self-regard' captures all that Kant intends by *Selbstsucht*. In the *Anthropology*, he observes that (cases of the sexual appetite aside) the term *Sucht* is used to designate passion (*Leidenschaft*) (*Anth* 7:266). Pluhar uses 'selfishness'; he notes that 'self-seekingness' would be the literal rendering and cites Kant's remark that a self-seeker (*solipsista*) is one who is indifferent to how things go for others, provided only that things go well for himself (*MdS* 6:450). The criticism implicit in the term can be seen in Kant's own usage (e.g. at *MdS* 6:206). Each of these renderings can be misleading by itself, but together they go some way towards conveying the concept's core idea of self-centred or self-directed concern.

with the inclinations, but they are distinctive in that they involve, in their maxims, certain claims, which contain an essential presumption of and concern for their own validity. They are propensities to assert, or maintain as valid, certain claims, or judgments, regarding one's worth and the goodness of one's happiness, and their efficacy with regard to action operates *through* these judgments. Or, as Andrews Reath puts the idea, 'self-love [is] a tendency to treat one's inclinations as objectively good reasons for one's actions, which are sufficient to justify them to others'.[15]

The nature of these claims will be examined more closely below. For the present we need only remark that, since the capacity to make claims of goodness (unlike the capacity to be pleased or displeased) resides not in our sensible nature but in reason (*CpV* 5:58), Kant's characterization of self-love and self-conceit as strivings or propensities of 'our pathologically determinable self' to secure validity for its claims is another way of saying that they belong to the pathological affection to which the human will – human practical reason – is naturally subject. The striving of the pathologically determinable self is thus just a liability of the will in a pathologically determinable subject to assert, to maintain as valid, certain claims reflecting self-directed concern, a liability that consists in the will's being naturally pathologically affected by the pathologically determinable self, or by the subject's lower faculty of desire, to which all its inclinations belong.

In describing our propensity to maintain such claims, Kant offers what appears to be an empirical observation about our sensible nature:

> Now we find our nature as sensible beings to be so constituted that the material of the faculty of desire (objects of inclination, whether of hope or fear) first presses upon us, and our pathologically determinable self, even though it is entirely unsuited through its maxims for universal legislation, nevertheless strives to make its claims valid antecedently and as the first and original ones, just as if it made up our entire self. (*CpV* 5:74)

Kant seems here to be reporting a finding about our nature as sensible beings. This may strike us as surprising when we recall that his aim in this chapter is to indicate *a priori* what the moral law must effect in the mind (*CpV* 5:72).

[15] 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination', in his *Agency and Autonomy in Kant's Moral Theory* (Oxford: Oxford University Press, 2006), 15. Kant's interpreters have differed, however, on whether he employs a notion of objective goodness in connection with personal ends of self-love. For an argument in favour of reading Kant as employing a subjective conception, see Thomas E. Hill, Jr, 'Personal Values and Setting Oneself Ends', in his *Human Welfare and Moral Worth* (Oxford: Clarendon Press, 2002), 244–74.

We need not suppose, however, that when he described his task as one of indicating a priori what the moral law must effect in the mind, Kant meant to imply that it could not be carried out with regard to an empirically determined conception of the mind. His description of his task leaves open the possibility that it will involve two steps: first, to introduce a determinate conception of our sensible nature by supplementing the abstract, a priori notion of the sensible nature of a finite rational being (a notion implicit in willing's self-consciousness) with an empirically acquired conception of 'our pathologically determinable self' as striving to make its claims valid, independently of consideration of their suitability for universal legislation. Second, with this empirically specified conception in place, to indicate a priori what the moral law must effect in the mind of a being with such a sensible nature. We can confirm that Kant is proceeding along such lines if we can find grounds he himself recognizes to be sufficient to rule out the possibility that we could know a priori that the propensity for such striving belongs to our sensible nature. Such grounds will emerge once we appreciate that self-conceit as well as self-love is integral to this propensity, and once we take note of self-conceit's distinctive features.

That Kant finds both forms of selfishness, or self-regard, to be implicated in our nature can be seen from the way he proceeds to describe them and to relate them to one another:

> This propensity to make oneself, in accordance with the subjective determining grounds of one's power of choice, into an objective determining ground of the will in general one can name *self-love*, which, if it makes itself legislative and the unconditional practical principle, can be called *self-conceit*. (*CpV* 5:74)

Both self-love and self-conceit are here identified with a propensity Kant has just assigned to our nature as sensible beings. This relation to our nature is reasserted a few lines later when he explicitly attributes to 'every human' this 'sensible propensity of his nature'.[16]

Yet while both self-love and self-conceit are thus linked to our sensible nature, they are not related to it in the same way. As the passage just quoted suggests, Kant thinks of self-love as the primitive form or grade of

[16] Another indication that Kant is concerned with our nature – and in particular with our nature so far as it is knowable only empirically – is that he speaks here, not of inclination (*Neigung*), which presupposes experience of its object, but of propensity (*Hang*). Propensity stands to inclination as potentiality stands to actuality: 'By *propensity* (*propensio*) I understand the subjective ground of the possibility of an inclination (habitual desire, *concupiscentia*) so far as it [sc. the inclination] is contingent for humanity in general' (*R* 6:28). And since inclination is contingent and so knowable only through experience, the propensity in our nature that grounds it is likewise knowable only empirically.

selfishness, out of which self-conceit somehow arises. Self-conceit is thus not so directly tied to our nature as is self-love. Kant says self-love is 'natural and astir in us even prior to the moral law' (*CpV* 5:73), but he never says this of self-conceit. Indeed, the description of self-conceit in terms of legislation and the unconditional practical principle seems to imply that it presupposes consciousness of the moral law. For part of what Kant attempted to show in Chapter I is that mere reflection on the idea of an objective practical principle enables us to know what the unconditional practical principle must be – just the moral law itself. Or as he put it in the *Groundwork*, 'If I think of a categorical imperative, I know at once what it contains' (*G* 4:420). Of course, not all of Kant's readers have been confident that they know at once what such an imperative contains. But the important point for present purposes is that Kant supposes we know this at least implicitly, so in characterizing self-conceit as what self-love becomes if it 'makes itself legislative and the unconditional practical principle', he must be thinking that self-conceit depends on consciousness of the categorical imperative, or the moral law. So while self-love is 'natural and astir in us even prior to the moral law', self-conceit must be in a suitably contrasting sense artificial, not natural, and astir in us only posterior to the moral law.[17] I will return to this later. Further reflection on self-conceit's relation to the moral law will establish a priori that only through experience can self-conceit be known to belong to our nature.

VI.

So far we've noted that self-love and self-conceit are inclinations – propensities at root – to maintain as valid certain sorts of efficacious claim, and that the claims of self-conceit presuppose the moral law in a way those of self-love do not. How are the claims different? Let's recall Kant's initial characterizations of self-love and self-conceit as self-benevolence and self-satisfaction. As his terminology suggests, Kant takes self-love to be a form of love, and self-conceit a form of esteem. Love and esteem are of course attitudes that can have others as their objects. Indeed, Kant identifies love

[17] Rousseau makes essentially the same point in drawing his similar contrast between *amour de soi-même* and *amour propre* ('Discourse on the Origin and Foundations of Inequality among Men', in Rousseau, *The Discourses and Other Early Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997), note XV). Two comments: (i) If self-love is a propensity to maintain certain claims as valid, there must be a sense in which it too is not prior to the moral law. For such a propensity depends on the capacity to comprehend the formal standard of validity for such claims, and the moral law is precisely that standard. (ii) As an 'artificial' propensity, self-conceit is self-incurred, even if there is another sense in which it is natural (see section VII below).

and respect – a form of esteem – as the two feelings involved in the virtuous carrying out of duties to others (*MdS* 6:448–9). Self-love and self-conceit are thus forms of love and esteem, but directed towards oneself. How do they arise?

Later in Chapter III Kant associates love closely with the inclinations and suggests that objects of inclination can be objects of love if they are animals, such as horses and dogs (*CpV* 5:76). In ordinary speech, however, our use of the term is not always subject to that restriction. Any object of inclination – anything we enjoy, or find agreeable – may be counted as something we love, if the inclination is sufficiently strong. We can say of Kant, for instance, that he loved carrots, at least the ones his friend Kiesewetter shipped him from Teltow. And elsewhere Kant acknowledges at least in effect that 'love' (*Liebe*) has this broader application, as when he notes that we say of some people that they passionately love such things as drink, gambling, and hunting (*Anth* 7:269). Later in the second *Critique*, he observes that we eventually come to love anything the contemplation of which allows us to feel a more extended use of our cognitive powers. Leibniz, for example, 'gently returned to its leaf an insect that he had carefully examined with a microscope because he had found himself instructed by his view of it and had, as it were, enjoyed a benefit from it' (*CpV* 5:160). Kant relies on this general point about the origin of love when he notes, in the *Metaphysics of Morals*, that the injunction to love your neighbour is to be understood as requiring, not that one make oneself feel love for others by a sheer act of will, which is impossible, but that one do well by one's neighbour, the result being that in time one will come to love those whom one assists (*MdS* 6:402). The beneficiaries, being the conditions of the exercise of one's practical and theoretical cognitive powers in the assistance one provides, come to be objects of one's affection.

A very general version of this point applies across the board to all objects of inclination, to the agreeable in general. Since any object of our enjoyment serves as a condition of the exercise of our powers in the experience that constitutes our enjoyment or use of it, any such object can be an object of love.[18] Leibniz loves his insects, benefactors love their beneficiaries, Kant loves carrots. But an agreeable experience is conditioned on two

[18] Enjoyments are not all on a par, of course. Some pleasures, such as those of learning and friendship, involve the cultivation of our powers. Others (which generally involve only the receptive, or passive, cognitive powers of sensation) are self-depleting and as such support only a qualified form of love; where they are self-destructive, as in the case of addictions and the like, the love can be modified by self-disgust and hatred of the object (see *CpV* 5:24; *Anth* 7:236–7).

sides. It depends as much on the subject as on the presence of the object. So just as the object, as an objective condition of my enjoyable experiences, counts as an object of my love, so does the subject – myself – who finds the object agreeable. And since I am necessarily a condition of all the agreeable activities in which my inclinations find their satisfaction, I am a natural and most intimate object of my love. Self-love is accordingly, as Kant says, 'natural and astir in us even prior to the moral law'. Thus, we naturally come to have, in addition to our other inclinations, a concern for ourselves, a love that inclines us to concern ourselves with how we are faring and to think it good that our inclinations be satisfied, just as Leibniz thought it good to return the insect to its leaf. This concern, I take it, is what Kant has in mind in his wry but insightful description of self-love as 'a predominant *benevolence* toward oneself'. Self-benevolence arises in the same general way as does benevolence toward others.

In self-love, then, one implicitly views oneself as *the pathologically determinable subjective condition of the agreeable activities in which one's happiness consists*. And self-love lies in a benevolence toward oneself conceived as such a condition. But the self-love in question is that of a rational being, a person, and so is an inclination or propensity to maintain as valid a certain claim or judgment. As we have seen, Kant says self-love is the 'propensity to make oneself, in accordance with the subjective determining grounds of one's power of choice, into an objective determining ground of the will in general' (*CpV* 5:74). His mention here of 'the will in general', coupled with his talk of self-love's striving to make its claims valid, suggests that he has in mind a certain judgment that purports and strives to be objective, valid for any person's will. Thus, my benevolent concern for myself is not merely an inclination to look after myself and to pursue the objects of my inclinations; it is originally a tendency to hold as valid that I, as a person who is a pathologically determinable subjective condition of my own happiness, am *a suitable object for the benevolent concern of any person*.

That Kant conceives of self-love along these lines is confirmed by the way he describes it when explaining the duty of beneficence in the *Metaphysics of Morals*:

> That beneficence is a duty follows from the fact that because our self-love cannot be divorced from our need of being loved by others (of receiving help from them when we are in need) so that we make ourselves an end for others, and [because] this maxim cannot be obligatory [for others] except through its qualification to be a universal law and so through a will to make others our ends too, the happiness of others is an end which is also a duty. (*MdS* 6:393)

In addition to illustrating what Kant means when he says the moral law 'restricts' self-love, rendering it 'reasonable' (*CpV* 5:73), this passage throws light on his characterization of self-love as a propensity to make oneself 'into an objective determining ground of the will in general'. In a rational being, self-love is a benevolent concern for oneself that has pretensions to objectivity, making a claim on other persons in that it regards its object – oneself – as worthy of others' concern as well.

VII.

We turn now to self-conceit, starting again from the idea that, just as self-love is a form of love, so self-conceit is a type of esteem. How does this attitude arise, and what sort of claim does it involve?

As we noted, Kant sees self-conceit as arising from self-love. If self-love 'makes itself legislative and the unconditional practical principle', it can be called self-conceit (*CpV* 5:74). How this happens Kant does not say, but the account just outlined of how self-love arises suggests a generically parallel account of how it becomes conceited self-esteem. Benevolent concern for oneself naturally arises through the awareness of oneself as the pathologically determinable subjective condition of the agreeable activities that constitute one's happiness. But as the concern of a person, a rational being, it takes the form of an inclination – originally a propensity – to maintain as valid a judgment deeming oneself worthy of the benevolent concern of any person. Hence self-love belongs to the natural pathological affection of our practical rational capacity, our cognitive capacity to determine what is good, what counts as an end, an object worthy of concern or pursuit. Now just as enjoyable experiences can spawn a concern for oneself as the subjective condition on which they depend, so in parallel fashion the cognitive activity of determining what is good, what counts as an end to be pursued or preserved, can lead to an esteem for oneself as the subjective condition of this activity. Just as enjoyable experiences involve an implicit awareness of oneself as the subject affected (pleased) by them, so one's judgments deeming the objects of those experiences to be good involve an implicit awareness of oneself as the subject making those judgments and maintaining them as valid. Bound up with these judgments is self-love's own judgment that, as the affected subject to whom that enjoyment belongs and on which it depends, one is oneself good, or an object worthy of persons' benevolent concern; so implicit in that judgment too is the awareness of oneself as also the subject and source of that very judgment and its validity. Thus, just as self-love arises from enjoyable experiences through a

first stage of reflection, namely reflection on oneself as the subjective condition of those experiences, so self-esteem arises from self-love through a second stage, reflection on oneself as the subjective condition of the (pathologically affected) judging activity in which self-love itself consists. As this stagewise advance reveals, self-esteem emerges as a distinct grade of self-love in that, though its object is the same, it does not represent that object in the same way. In bare self-love one conceives of oneself as a pathologically determinable subject, that is, a *passive*, *experiencing* subject, liable to feel pleasure and pain; in self-esteem one conceives of oneself as (also) an *active*, *cognizing* subject, capable of knowing and determining what counts as good and as an end to be preserved or pursued. In this way, then, the activity of claiming oneself to have standing as an end worthy of the benevolent concern of any person gives rise to a conception of oneself as worthy of esteem on account of one's cognitive power thus to determine an end and thereby to make claims about what should be done that are valid for persons generally. Since this power is just the practical rational capacity constitutive of personhood, esteem directed at it must be closely related to respect; for respect too, as Kant says, 'is always directed only at persons, never at things' (*CpV* 5:76).

Despite the difference just noted in the way self-love and self-esteem represent their objects, the similarity in the way they arise might encourage the thought that self-esteem is also like self-love in being such as can be made reasonable by being restricted by the moral law. It might seem that just as love of self merely needs to be broadened so as to include a benevolent concern for others, so esteem for self merely needs to be extended so that others too are included as objects of one's respect. But to draw such a conclusion would be to overlook a second difference between self-love and self-esteem, rooted in a general difference in form between the feelings of love and esteem.

Broadly speaking, love relates to its object through its representation of the latter's quality, whereas esteem attends to its object's quantity, its magnitude. This means that the claim or judgment of love is absolute, whereas that of esteem is relative. Loving something, or having a benevolent concern for it, requires no comparison of it with any other thing of its kind. Yet holding a person in esteem always involves an implicit comparison with some standard or measure. This reference to a measure simply reflects the judgment's character as a judgment of quantity, in which magnitude (size or power, for instance) is assigned to a thing. All such judgments rely on some measure, even those that merely deem an object to be great (large, powerful), without determining how great. Thus, the earth is judged small

compared to the sun, but large compared to the moon; and a child may be considered weak or strong according as it is measured against an adult or an infant. There is a further respect, however, in which relativity figures in esteem. When one holds a person in esteem, the estimation is not merely conceptual, but aesthetic in Kant's technical sense (see *CU* 5:251). That is to say, it lies in feeling, in one's consciousness of the effect one's own comparison of the object with oneself has upon oneself. Whether a particular person is held in esteem is thus relative to the judging subject, just as the city of Königsberg might be deemed large and impressive by the residents of a nearby hamlet but disregarded as small and insignificant by the citizens of Berlin. Judgments of esteem are thus relative in two ways, in one respect on account of being quantitative, in another on account of being aesthetic.

The points just noted may occasion some puzzlement regarding the esteem figuring in self-conceit, where the object of judgment is oneself. Since the judgment is aesthetic, the estimation is relative to the subject, which again is oneself. But the relative character of estimates of quantity entails that estimation is impossible where the object serves as its own measure. No estimate of the length of a year can be gained by noting that a year is as long as a year. Self-conceit's esteem for self can therefore only be indirect, through depreciatory judgments concerning others, in which, with a feeling of disregard or contempt, one fancies them small in comparison with oneself. Thus the self-esteem constituting self-conceit rests essentially on comparative aesthetic judgments asserting one's superiority over others. In no other way can self-conceit, as an aesthetic appreciation of one's worth, arise.

A difference in form between the attitudes of self-love and self-conceit is now readily discernible. It is essential to self-conceit that its representation of oneself includes a comparison with others and in particular a placing of oneself above them in point of practical-cognitive power, on account of which one's own judgment, so far as cognitive validity is concerned, prevails over theirs in cases of conflict. But no such implicit ranking or ordering is contained in the representation of the object of self-love.

Here it is important to appreciate that, as a propensity in human nature, self-conceit does not develop from self-love through empirical comparison of one's own actual practical judgments with those of others (even though, once such experience is available, self-conceit tends, so far as it can, to represent itself as supported by well-grounded judgments about how one compares with others in point of virtue and vice, wisdom and folly). It is rather a *propensity* to make a *practical* judgment, manifested in the readiness, prior to any scrutiny of the actual record of one's own judgments or those of

others, to presume, where one's own judgment conflicts with that of another, that it is the latter that needs to be corrected. Others are thus to defer to oneself in such judging rather than the reverse. As if an a priori judgment in its own right, it is the original act of placing oneself, in respect of this capacity, in an imagined position of superiority over everyone else.[19]

Since self-conceit is nevertheless an outgrowth of self-love, its claims still rest on 'the subjective conditions' of the pathologically determinable self that one's self-love inclines one benevolently to concern oneself with. But in contrast to bare self-love, self-conceit 'prescribes the subjective conditions of self-love as laws' (*CpV* 5:74). In self-conceit, one treats these conditions – one's own pleasures and pains – as all-sufficient grounds for practical cognition and regards oneself, uniquely positioned as one is to discern them, as uniquely empowered to determine ends and to make valid claims about what should be done. Self-conceit thus lies in the disposition to suppose oneself to be the one person whose judgment ultimately counts, the one subject in whom practical reason, the full capacity to determine what's good and what's not, is to be found.

In regarding oneself as the seat of such a capacity, one views oneself in the way Kant describes the good will in the famous opening paragraphs of the *Groundwork*. There he proclaims that the good will is the only thing that can be regarded as good without qualification. And his argument in support of this proposition turns on the idea that the will – as practical reason, the capacity for practical knowledge – is the capacity whose exercise constitutes the use of all the other things we might call good, and that the good will in particular is that very capacity, exercised in agreement with its own principle, determining the right and proper use of all those other things. Since the good will lies precisely in the capacity for knowledge regarding the good, it is the one thing that determines the goodness of all other things and the one thing that can never have any bad effects. It is accordingly the only unconditioned, unqualified good. Kant also argues in the *Groundwork* that the moral law is the principle of such a will, the distinctive form of its practical cognitive activity. Self-conceit is a regarding of oneself in just this way, but with the crucial difference that when self-love 'makes itself legislative and

[19] The a priori character of this presumption is well reflected in the first sentence of Descartes's *Discourse on the Method*: 'Good sense is the best distributed thing in the world: for everyone thinks himself so well endowed with it that even those who are the hardest to please in all other things do not usually desire more of it than they already have' (*Discourse on the Method*, AT 6:2). Hobbes makes essentially the same point about the 'vain conceipt of ones owne wisdome': 'such is the nature of men, that howsoever they may acknowledge many others to be more witty, or more eloquent, or more learned; Yet they will hardly believe there be many so wise as themselves' (Thomas Hobbes, *Leviathan*, ed. R. Tuck (Cambridge: Cambridge University Press, 1997), chapter 13). Both authors take this universal conceit to indicate a natural equality among human beings.

the unconditional practical principle', its judgment is that one is oneself the sole bearer of this capacity for practical cognition, to determine what is good and what the proper use of things is. So the judgment of self-conceit is, in effect, the opening proposition of the *Groundwork*, but modified to read: 'Nothing at all in the world can be regarded without qualification as good, except *my* will.'

Kant does not explain how it is that self-love can take this turn, rather than developing directly into a virtuous attitude toward one's capacity for practical cognition, a recognition of it as shared with all other persons. As we noted, he does not suggest that self-conceit is 'natural and astir in us even prior to the moral law', but rather takes himself to have established that mere reflection on the bare idea of an unconditional practical principle suffices to reveal that such a principle must be the moral law (not oneself). So not only does self-conceit arise out of self-love; it can do so only through some antecedent, if only implicit, consciousness of the moral law. Self-conceit must therefore lie in a propensity to turn away from – and against, in an act of psychic violence – this antecedent understanding of what the unconditional practical principle must be. As Allen Wood has observed, self-conceit is what Kant describes in the *Religion within the Bounds of Mere Reason* as the propensity of radical evil, to reverse the proper order of the springs of choice, subordinating the moral law to the condition of self-love in one's fundamental maxim (*R* 6:36).[20] Although Kant attributes this propensity to human nature (*CpV* 5:74; *R* 6:32–3), for the reasons already noted it cannot be natural in the way that self-love is. Kant's recognition of this difference is reflected in his assertion in the *Religion* that we know of this propensity only through experience, only through the fact – not, of course, the 'fact' of reason, but the fact of actual human choices and actions: deeds, or rather misdeeds, in which that reason is misused (*R* 6:32). This may account for why Kant does not attempt to explain how self-conceit arises from self-love. This outgrowth is something into which no insight is possible. Unlike self-love, self-conceit cannot be recognized or understood entirely a priori; it can be assigned to our nature only a posteriori and hence can belong to it only in a secondary way. There is something radically unintelligible about it, an element of the surd.[21] Confirmation that it must

[20] *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999), 290.

[21] It can still be incumbent on the practical philosopher to seek a way of accounting for the presence of self-conceit in human nature. Thus, Kant offers a practical-anthropological view of what he calls our 'unsociable sociability': nature is to be regarded as having so framed us that we are drawn into competitive relations that drive us to cultivate our talents and so to perfect ourselves in the individual and in the species. For discussion, see Wood, *Kant's Ethical Thought*, chapter 6 (esp. 213–15).

be regarded in this way can be found in Kant's account of how the moral law strikes down self-conceit, to which we next turn.

VIII.

We have now to consider the effect the moral law has on feeling through its bearing on the propensities of self-love and self-conceit. Earlier we examined Kant's description of how the moral law must, by thwarting the inclinations, produce a feeling of pain. We noted that this effect reveals nothing distinctive about the moral law. But the propensities of self-love and self-conceit yield inclinations to maintain as valid claims that involve a conception of oneself as not only a passive subject of pleasing and displeasing experiences but also an active, practically cognizing being, a person, and hence as worthy of esteem as well as benevolent concern. Since this self-conception *presupposes* consciousness of the moral law, we can anticipate that this law, as it bears on these propensities, will affect feeling more intimately than it does in thwarting the inclinations.

Let's start with self-love. Kant says, 'pure practical reason merely *infringes* the love of self, in that it merely restricts it, as natural and astir in us even prior to the moral law, to the condition of agreement with this law' (*CpV* 5:73). At one level of description, this infringement is like the thwarting of the inclinations, merely a source of pain. But there is also a striking difference. We noted earlier (section VI, above) that in one of his explanations of the duty of beneficence Kant illustrates how self-love is limited (*MdS* 6:393). The limitation is represented as a restriction, not of an inclination, but of a *maxim* of self-love, a judgment or claim in which we 'make ourselves an end for others'. The imposed limitation – that we 'will to make others our ends too' – is a condition the claim must satisfy to qualify as a universal law, or to be valid and to count as practical knowledge. A similar portrayal of the restriction is offered earlier in the second *Critique*:

> Let the matter be, for example, my own happiness. This, if I attribute it to each (as, in the case of finite beings, I may in fact do), can become an *objective* practical law only if I include in it the happiness of others. Thus the law to promote the happiness of others arises … merely from this: that the form of universality, which reason requires as the condition of giving to a maxim of self-love the objective validity of a law, becomes the determining ground of the will. (*CpV* 5:34)

The important point for our purposes here is that this restricting condition gains purchase only because self-love strives to maintain as valid – in its maxims, or 'volitional opinions' (*CpV* 5:66) – certain claims that *purport* to

have the standing of practical cognition and so are liable to limitation by the standard of 'objective validity' *constitutive* of such cognition, namely the form of universal law, or universal legislation. The moral law thus infringes self-love from within, not as a mere external constraint.

Also noteworthy, and striking, is that this 'infringement' or 'restriction' can be the source of *positive* duties – in the example at hand, the duty to widen one's benevolent concern beyond oneself to include others as well. This is not to say, of course, that the moral law's limitation of self-love can directly produce a feeling of love for others. The immediate effect on feeling is merely negative. But as we noted, Kant holds that fulfilling the obligation to make others our ends by assisting them when they are in need eventually gives rise to a feeling of love for them, in a way that broadly parallels the way in which the enjoyable experience of the objects of one's own inclinations naturally activates the propensity of self-love. So while the direct effect of the moral law's infringing self-love is wholly negative, this infringement can, once the moral law becomes a spring, gradually lead to the love of humanity that would animate the benevolent conduct of virtue, constituting a kind of self-enlivenment of practical reason's own exercise. Nevertheless, the restriction of self-love does not yet make known 'the force of pure practical reason as spring, but only the resistance to springs of sensibility' (*CpV* 5:78–9).

IX.

In the case of self-conceit the impact on feeling is dramatically different. Here again we should start with the act and then turn to the effect on feeling. As what self-love becomes when it makes itself 'legislative and the unconditional practical principle', self-conceit attempts to identify the standard of the objective validity of practical knowledge and practical law with oneself, with one's own capacity to estimate the goodness and worth of things. It usurps the position originally occupied by the moral law, which, as the form of practical knowledge and practical law, is the standard by which the cognitive validity of all practical claims and judgments is to be assessed. Kant says pure practical reason '*strikes down* self-conceit, in that all claims of self-esteem that precede agreement with the moral law are null and without any warrant, in that precisely the certainty of a disposition that agrees with this law is the first condition of all worth of the person … and all pretension prior to this is false and contrary to law' (*CpV* 5:73). Whereas the moral law merely restricts the claim of self-love, it utterly rejects the claim of self-conceit. As we saw, self-love's claim involves making oneself an end for others. This claim is valid as far as it goes, *provided* that it is completed

through the extension of one's judgment to cover others as well, so that the claim has the form of law. The claim of self-conceit, on the other hand, is not similarly amenable to the form of law. Because this claim is concerned with the quantity rather than the quality of its object, and because its estimation depends on comparison with other persons, it *rules out* the extension of its esteem to others. While the concern originally figuring in self-love merely *fails* to extend itself in a way that would give it the form of law, the esteem that constitutes self-conceit essentially *precludes* such extension. In ignoring others' standing as ends, self-love exhibits no more than the sort of negative deficiency characteristic of ignorance; self-conceit, on the other hand, positively errs, or rather lies in the very erring itself. Containing nothing sound that might be retained through mere limitation or restriction, it must be rejected root and branch.

Kant says that when the moral law strikes down self-conceit, the direct effect on feeling is the negative, or displeasing, sentiment of humiliation. But he takes this description to capture only one side of the experience, arguing that the humiliation introduces an indirect effect, in that it awakens the positive feeling of respect for the law. His thought seems to be that, just as the striking down itself is the recognition in practical consciousness of the *relation* between the moral law and self-conceit, so its effect in feeling is a two-sided sentiment, with one aspect corresponding to the downcast presumption and the other to the activity of practical cognition that casts it down.

Let's approach this feeling from the negative side, beginning with a question that might be raised about Kant's description of it. One might grant that the initial effect is negative, but wonder whether it lies simply in humiliation. It is clear that self-conceit and the moral law are, in a sense, in competition. Self-conceit is a propensity to put oneself, one's *own* practical rational capacity, on the throne, in the position of the moral law. Something has to give. Now individuals who stand in an immediate relation of competition characteristically regard one another with animosity. They call one another enemies; they hate one another. These hostile feelings differ essentially from those of humiliation and self-contempt. And though the latter sentiments may be salient when one adversary suffers a defeat at the hands of the other, the former are also implicated, commonly spawning a desire for vengeance, a desire to reverse the outcome. When the moral law lands a blow against self-conceit, then, mightn't the attitude that ensues be better described as a vengeful hatred of this law, or as misology, even if it also involves humiliation?

Kant allows that in certain conditions something that might be described as hatred of reason and the moral law can arise. As we've noted, self-conceit depends on consciousness of the law. So it always involves an at least

obscure awareness of itself as in necessary conflict with the very thing it pretends to be. This recognition, being frustrating as well as humiliating, may prompt it to strike back in anger, in a fanciful attempt to strike down the moral law itself. In this retaliation, self-conceit depicts the law as an object worthy of scorn and disdain by staging a reversal in representation, not only arrogating to itself the mantle of legitimacy but also projecting its own reactive nature onto morality. Thus, it may 'deride all morality as the mere phantom of a human imagination which oversteps itself through self-conceit' (*G* 4:407; cf. *G* 4:405). There are various ways in which this degrading portrayal of morality might be elaborated, but one of them is very familiar, having been urged by Thrasymachus in Plato's *Republic* and by many others since. According to this depiction, morality has no true footing in nature but comes to be through human convention when many inferior, envious individuals, resenting on account of their self-conceit their subjection by the superior few, join together in order to strike back. Incited by clever men, they band together under the fabricated banners of law, justice, and equality (merely imaginary ideals which they secretly hate, just as they are secretly opposed to one another), in order that through their combined forces they might have the strength as a body to usurp the position occupied by their greatest enemy, those few who owing to their natural superiority truly deserve esteem.

Kant's exposition of the moral law, however, according to which it is nothing but the form of goodness itself, entails that a reaction on the part of self-conceit that depicts morality as bad, as an object to be hated and opposed, cannot belong to self-conceit's original relation to this law. To that extent, this hatred is also less stably present than the more fundamental feeling of humiliation, which properly reflects the actual underlying order of representations in the mind. The hatred can be sustained only through self-conceit's active, or rather reactive, occlusion of the original recognition of the moral law as the form of goodness, legislated by one's own practical reason (see *R* 6:23–4n.; *MdS* 6:485).

But how exactly does the feeling of humiliation arise? Here it will help to recall a point touched on earlier, that esteem and contempt represent their objects differently from the way love and hatred regard theirs. Broadly speaking, the objects of love and hatred differ in quality, as positive and negative: love's object is seen as good, hatred's as bad. The objects of esteem and contempt, on the other hand, differ in quantity, or magnitude: we esteem what is great, but think little of what is small. In a word, love looks to the good and moves us to assist it; esteem contemplates the great and prompts us to defer to it.

Humiliation is self-contempt, so in so far as it arises from the moral law's striking down self-conceit, it depends on self-conceit's representing itself as small in relation to the moral law. But small in what respect? Clearly not in respect of its causal power as a spring. For Kant appeals to humiliation as a feeling that plays a role in the moral law's *becoming* a spring. If the humiliation arose from self-conceit's consciousness of being causally overpowered by the moral law, it would come on the scene too late to play such a role. Nor would it then be through its standing as the will's objective determining ground that the moral law comes to be its subjective determining ground; self-conceit would simply be thwarted. So it must be in another respect that the moral law strikes down self-conceit: the difference in magnitude must pertain, not to causality, but to cognition.

Now as we've seen, cognitive capacity lies at the core of self-conceit's self-conception. In addition to presupposing a consciousness of the moral law as the form of practical cognition and law, self-conceit essentially involves a conception of oneself as the ultimate bearer of the practical-cognitive capacity to make law. It is precisely the propensity to regard oneself as the one who really has the insight or understanding or good sense to know what should be done and how things ought to be arranged – in short, to lay down the law. But built into the notions of cognition and of law is the idea of universality, as a formal feature of cognition and standard of its validity.

Kant sees cognitive activity as having both a qualitative and a quantitative dimension. The judgments of discursive cognition are positive (affirmative) or negative in quality, and in quantity they are universal, particular, or singular. The distinguishing mark of *rational* cognition is universality, that is, strict universality, to which no exceptions are possible. In respect of quantity, then, the cognition of reason is infinite, unlimited. The moral law, as the form of all practical cognition, holds for all rational beings. The scope of its validity, the magnitude of the cognition it constitutes, is unrestricted. It is itself the archetype of this very universality – this infinite magnitude – that it holds up, in the idea of law, as the standard of practical rational cognition. Self-conceit, we have noted, is a propensity to regard one's own judgment as occupying this position, even though it is based on the 'subjective determining grounds' of self-love, which are at bottom the particular affections on which one's inclinations are based. Thus self-conceit is a propensity to put what is in fact a radically particular standard of judgment – the feeling of a single individual – in the place of a standard of objective rational cognition, which, to be a standard of all cognition, must itself have unlimited or absolutely universal validity. In point of quantity, then, self-conceit stands to the moral law as an extensionless

point stands to infinite space. Thus Kant describes the moral law as 'infinitely' infringing self-conceit and as deeming its claim to be 'null', 'nothing' (*CpV* 5:74, 73, 78). It seems impossible to exaggerate how mortifying it must be for self-conceit when the grounds of its grand pretensions to universal validity are exposed as microscopically, indeed invisibly, small.[22] In the court of practical reason, these claims are not just defeated but utterly annihilated.

The moral law, in contrast, is the object of 'boundless high esteem' (*CpV* 5:79). While the mortification of self-conceit is the first moment of the moral law's effect on feeling, the other side of this effect is a feeling of respect for the law, which Kant describes as positive. Of course, when one feels humiliated in comparison with another person, there is ordinarily nothing particularly positive in the esteem or respect one feels for the other, a point Kant discusses in some detail (*CpV* 5:77). But as we have seen, self-conceit is a propensity that already presupposes consciousness of the moral law, and because of this it is possible, as Kant notes, to appreciate, even in the experience of humiliation, that the consciousness of the moral law and its validity is the work of one's own practical reason, except that now this capacity is understood to be shared with other persons (*CpV* 5:80–1). As our attention is attracted to the moral law in the feeling of respect, we are able to recognize our practical reason *as* ours and indeed to identify with it, by recognizing it to be more fundamental to our self-conception than is self-conceit. For self-conceit presupposes awareness of the moral law; awareness of the moral law does not presuppose self-conceit. Through the *felt diminution* of self-conceit in one's estimate of the magnitude of its cognitive validity and the corresponding *felt magnification* of the moral law in one's recovered estimate of its boundless validity, and the accompanying elevation that comes with the recognition that we share in the cognitively sublime power of reason, a power more integral to us than the propensity of self-conceit, the moral law has an indirect positive effect on feeling through which it establishes itself as a spring, a *Triebfeder*. This efficacy of pure, unlimited *cognition* on our capacity to *feel* counters the propensities of self-love and particularly self-conceit, weakening their influence and instilling in its place a readiness to defer to the consciousness of the law in the

[22] Like an emperor with no clothes, it is originally the most naked of all lies, which any child can see. Hence only by the imaginary cloaks and distracting regalia of the most violent self-delusion – or in other words by the illicit appropriation of what in truth are mere gifts of nature and of fortune rather than anything properly one's own (see *G* 4:393) – can this root of all vice possibly sustain itself. For this reason 'the first command of all duties to oneself' must be Socrates' Delphic dictum, 'know yourself' (*MdS* 6:441).

exercise of the will, thereby enabling us to take an interest in the moral law itself and to make it the maxim of our conduct, a basis for progress towards virtue and all that it will entail, including the growing concern for others that will arise through the practice of beneficence.

As we noted, Kant has taken care to describe how the moral law can be a spring without assuming any antecedent feeling attuned to morality. The positive feeling of respect that reveals the 'force' of pure practical reason as a spring is indirect, arising only through the humiliation of self-conceit. It reveals a positive strengthening of the moral law's efficacy, but only indirectly, through the removal of an obstacle to it (*CpV* 5:79). By thus eschewing any assumption that our capacity of feeling is specially attuned to the moral law, Kant has been able to describe how – in accordance with the *can* required by pure practical reason's *ought* – the moral law is itself a spring. Of course, we have also seen that this account of the positive force, or motivating power, of the moral law *has* relied on the presence of the propensity of self-conceit, something radically surd, into which no a priori insight is possible. But this does not imply a deficiency or limitation in Kant's view of moral motivation, since it is only because of our empirical recognition of the actuality of the opposing force of self-conceit that it is necessary in the first place to provide a separate account of the moral law as a subjective determining ground of the will.[23]

[23] This is a revised version of a paper presented in September 2007 at a conference on Kant's *Critique of Practical Reason* at the University of St Andrews. I thank the audience and especially the other contributors to this volume for helpful discussion. I am also grateful to Jochen Bojanowski and Steffi Schadow for useful comments.

CHAPTER 6

Two conceptions of compatibilism in the Critical Elucidation

Pierre Keller

The 'Critical Elucidation of the Analytic of Pure Practical Reason' develops two distinctive conceptions of how freedom of choice and causal determinism may be reconciled. These two conceptions of freedom of choice correspond to a distinction between what Kant calls 'psychological' and what he calls 'transcendental freedom'. A person's choices are 'psychologically free' when those choices are determined by the psychological states of the agent rather than by external causes. Psychological freedom in this sense is compatible with causal determinism in the sense traditionally associated with the metaphysical doctrine of compatibilism. While such psychological freedom is involved in our choices, Kant argues that it is not strong enough to support the conception of choice required by the demands of moral responsibility.

Kant endeavours to show that theoretical attempts to explain how the kind of free agency presupposed by moral responsibility can be accommodated in psychological or broadly natural terms are illusory and must fail. His reasons, however, go deeper still. Kant thinks that no comprehensive explanation is forthcoming at all of how such free agency is possible. This leads him to a two-standpoint conception of human agency. When we regard our actions from the vantage point of moral responsibility, we regard them from a different standpoint from that which we occupy when we are explaining them in terms of antecedent causes. This may be regarded as a form of compatibilism, for it aims to show that free will and the thesis of causal determinism are mutually compatible. However, in the form that Kant defends this view it is akin to libertarian forms of incompatibilism in that it insists that actions for which one is morally responsible are not, as such, to be regarded as caused by antecedent events, but caused by the agent independently of antecedent events in the agent's psychology or outside of it. The agent must regard himself as transcendentally free rather than merely psychologically free in order to account for his responsibility for his actions and his character. I would note though that the metaphysical commitment

to independence from antecedent causes involved in this notion of transcendental freedom is a commitment that holds only from the standpoint of moral responsibility. Kant rejects any attempt to provide a comprehensive theoretical account of how causal determinism and the kind of freedom of choice involved in moral responsibility are to be reconciled. Thus I take Kant to reject the shared metaphysical assumption of metaphysical compatibilists and incompatibilists. This assumption is that the theoretical point of view can provide the comprehensive perspective from which the question of compatibility or incompatibility may ultimately be decided. It is in virtue of the standpoint relativity of any determinate sense that can be made of claims of the compatibility and/or incompatibility of free will and determinism to which Kant is committed that leads him to argue for a form of what Allen Wood has aptly referred to as the 'compatibility of compatibilism and incompatibilism'.[1] Kant regards the viability of his two standpoint accounts of individual choice as a function of the complementarity of the standpoints of theoretical and pure practical reason. Theoretical reason demands causal antecedents for all choice and thus allows only causally and empirically conditioned choices from its purview. The standpoint of moral or pure practical reason demands choices that are unconditioned by antecedent causes. These standpoints are mutually exclusive if comprehensive, but complementary if they are not.

1. THE ANALYTIC OF PURE PRACTICAL REASON AS PRACTICAL SYLLOGISM

The Critical Elucidation presents the argument of the Analytic of Pure Practical Reason as a kind of practical syllogism that highlights the distinctive contribution of pure practical reason and the complementary relation of the standpoints belonging to theoretical and pure practical reason (*CpV* 5:90).[2] Kant cites this syllogistic structure of the Analytic as evidence for

[1] Allen Wood, 'Kant's Compatibilism', in A. Wood (ed.), *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), 73–101. It should be noted that Wood takes Kant's thesis that compatibilism and incompatibilism are in some sense compatible to depend on a dis-unifying conception of ourselves as belonging to two different metaphysical worlds. I take the two worlds in question as expressions of the commitments of the theoretical and practical points of view. The metaphysical dis-unity in question would only arise from the vantage point of the kind of comprehensive theoretical point of view that I take Kant to reject.

[2] Kant compares the Analytic of Pure Practical Reason in the second *Critique* to what he here refers to in the Elucidation as the 'Analytic of Pure Theoretical Reason' in the first *Critique*. He argues that the Analytic in the second *Critique* inverts the order in which the Critique of Pure (Theoretical or Speculative) Reason proceeds (*CpV* 5:90). The first *Critique* moves from sensible objects to concepts

the unity of theoretical and practical reason and for the prospect of at some point satisfying the need to derive the principles of theoretical and practical reason from a single principle.

In this syllogism practical reasoning proceeds from fundamental principles of pure practical reason to the recognition of specific reasons for specific actions to be pursued or avoided and then to specific categorical obligations to act. The fundamental principle of pure practical reason (the categorical imperative) is the universal major premise in this syllogism or reasoning process. The minor premise involves the subsuming of possible actions under the fundamental principle of pure practical reason. The principle of pure practical reason is the basis upon which we form our very concepts of an intrinsic good to be pursued and evil to be avoided (Chapter II, 'On the Concept of the Object of Pure Practical Reason'). These concepts of good and evil, and the principles of pure practical reason that define them, need to be applied to specific spatio-temporal actions. Since the identity conditions for events are tied to their causal conditions, the nature of an action as a spatio-temporal event precludes it from expressing pure practical (moral) concepts. The pure practical concepts of morality involve principles of action that are independent of causal antecedents because they are based on the general principle of autonomy. However, Kant argues that moral principles can be indirectly connected to action tokens through the universal natural laws that apply to all events as tokens of a certain type. These natural laws model the unrestricted generality or universality of pure practical principles ('Of the Typic of Pure Practical Judgment'). Finally, in the conclusion of this practical syllogism, the authority of pure practical reason displays itself to us in our moral self-consciousness as a sense of obligation that motivates us to act. This moral self-consciousness is our capacity to reason practically independently of considerations that depend on the vagaries of our experience. Our moral self-consciousness puts impartial principles that are independent of the particular contexts of our experience before us. Yet that moral self-consciousness is nevertheless able to express itself in a particular context as a sensible incentive to act. This is

and then to principles. By contrast the second *Critique* begins with principles and proceeds from there to concepts formed on the basis of those principles and then applies those concepts to sensibility. 'The Analytic of Theoretical Pure Reason' is supposed to be divided into the Transcendental Aesthetic and Transcendental Logic (*CpV* 5:90). However, in the first *Critique*, the Transcendental Aesthetic is a separate section from the Transcendental Logic that is divided into a Transcendental Analytic and Transcendental Dialectic. In consequence the Analytic of Pure Theoretical Reason does not correspond precisely to anything in the structure of the first *Critique*. Kant's general point that the order of exposition in the first sections of the two *Critiques* proceeds in inverse order in respect to sensibility, concepts and principles is not affected by this mischaracterization.

possible by means of the feeling of respect of which our moral self-consciousness is the intellectual cause (Chapter III, 'On the Incentives of Pure Practical Reason').

II. THE DISTINCTIVE AUTHORITY OF MORAL PRINCIPLES

The first chapter of the Analytic of Pure Practical Reason develops the basic principles of practical reason axiomatically from the definition of a practical principle. The syllogistic reconstruction of the argument of the Analytic as a whole takes this a step further. This presentation of practical principles *more geometrico* along the lines of Spinoza's *Ethics* is now put in proper perspective. The authority of the principles belonging to practical and pure practical reason is different in kind from the kind of authority that the principles of geometry have. Kant notes a striking difference between theoretical and practical reason with respect to basic a priori principles of reason and the ultimate significance of syllogistic reasoning. This difference puts the syllogistic interpretation of the structure of the Analytic in a completely different light. In the case of theoretical reason, we have evidence that there are bona fide a priori principles of pure reason in the existence of the disciplines and theorems of mathematics and mathematical physics.[3] The basic principles of pure practical reason unlike those in mathematics and theoretical physics are not constructed and justified within science. The principles of pure practical reason must, 'as the first data, be put at the basis of all science and cannot first arise from it' (*CpV* 5:91).

The theorems of practical reason culminating in the principle of autonomy and the fundamental principle of pure practical reason are to be understood as expressions of the commitments of common-sense reasoning about action. As such practical principles they are derived and gain their justification from the commitments of common-sense reasoning. The task of the moral philosopher is to show that the fundamental principle of pure practical reason (the categorical imperative) is something that 'every natural human reason knows completely a priori and not dependent on

[3] The existence of mathematical science does not mean that a priori principles need no defence and critical evaluation from philosophy. In the first section of the Transcendental Deduction in the first *Critique*, *CrV* A87–8/B119–21, Kant notes that geometry does not need certification from philosophy for its cognition a priori, but he also insists that once pure concepts of the understanding are in play, the significance of space (and geometric structure) itself becomes ambiguous and needs critical reflection. But a priori principles of reason in mathematical science are inherently more secure, both in their substantive claims and in their non-empirical status.

sensible data, as the supreme law of its will' (*CpV* 5:91). The 'purity of the origins of this principle' in other words, that it is not an empirical principle, is established 'in the judgment of this common reason' before science can take it as a 'fact' (*CpV* 5:91).[4] The a priori principles of practical reason are to be established from common-sense reasoning by 'an experiment with every human practical reason' that is reproducible at any time (*CpV* 5:92). Principles of practical reason based on the principle of happiness are empirical and context-dependent. One must distinguish moral principles of practical reason from such context-dependent empirical principles just as the geometer distinguishes his geometrical principles from empirical principles. Kant suggests that one can use examples of conduct in thought experiments to elicit and articulate common-sense reasoning about how to act. Presented with particular cases of moral reasoning in thought experiments, common sense distinguishes what is moral from what is conducive to one's happiness and is in one's self-interest.

Thought experiments can establish a priori principles, since they probe what anyone ought to do in certain factual or counterfactual situations. We have this capacity to probe our commitments in thought experiments because we have the capacity in the purely reflecting 'I' of thought and reason to put ourselves in the place of any other individual in different counterfactual circumstances. We can do this because we can abstract from anything distinctive about our own particular selves in self-conscious reflection.[5] For instance, we take a case of someone who would like to lie in order to acquire something. Then we present that individual with the moral law. Regardless of whether that person is honest, he 'just this once puts himself only in thought in the place of an honest man' (*CpV* 5:92). In the process, he sees that his respect for himself requires truthfulness of him despite the fact that his advantage lies elsewhere. The primary purpose of the thought experiment is to distinguish moral principles from principles of reason based on happiness. Kant compares this thought experiment with an experiment in chemical analysis. Adding an alkali to a solution of calcium (calcareous earth) dissolved in hydrochloric acid will precipitate out the lime. In the same way, the moral law will separate personal advantage from respect for oneself as a person. Then, once the lost advantage has been

[4] See Pauline Kleingeld's chapter in this volume on the 'fact of reason' and *CpV* 5:6, 31–2, 42, 47, 55, 104–5.

[5] 'Discursive consciousness (pure apperception of one's mental activity) is simple. The "I" of reflection contains no manifold in itself and is always one and the same in every judgment, because it is the merely formal element of consciousness' (*Anth* 7:141).

separated and washed of all contributions by reason, the lost advantage involved may be weighed and considered by everyone.

Kant takes the commitments of common-sense morality to reflect this role of pure practical reason in establishing action that is not causally conditioned. Common-sense reasoning is committed to categorical obligations, that is, to obligations that apply unconditionally to us. Kant illustrates this strong commitment to being able to act independently of what we might happen to want to do by the use of a thought experiment, an experiment of pure reason, applied to common-sense moral reasoning. Individuals are able to overcome even the strongest of immediate desires in the face of imminent threats to their own survival and long-term happiness. A person might overcome a desire that she regards as irresistible if faced, upon satisfaction of that desire, with execution by hanging (*CpV* 5:30). Thus self-interest and the desire for self-preservation can overcome the motivational pull of immediate desires. More importantly, human beings have a much more robust form of freedom from both immediate desire and from the various forms of long-term self-love (for instance the kind of self-love displayed in the desire for self-preservation). This freedom is displayed in our recognition of our capacity to be obligated to act in a certain way even though such action may go against our own survival. We think that we could resist even the threat of death if that were the only alternative to doing something that we regard as quite wrong. We take ourselves to have the ability to act in this way, regardless of whether we actually do act this way (*CpV* 5:30). Taking ourselves to be under the obligation to act this way is tantamount for Kant to recognizing that we can act this way, since he is committed to the principle that ought implies can.

III. FREEDOM AND THE LIMITS OF EXPLANATION

The experimental method of reason distinguishes sensible from intellectual reasons and establishes the basic principles of practical reason. This gives the method an immediate relevance to the interpretation of action. It also raises a deeper question about the relation between theoretical and practical commitments with respect to the understanding of action and in this way raises the problem of the relation of theoretical to practical reason and the very nature of the fundamental principles of practical and theoretical reason.[6]

[6] The philosopher establishes the principles of theoretical understanding and practical reason in the analytic of theoretical understanding and the analytic of practical reason by a procedure of experiment and analysis that Kant also compares with that of a chemical reduction or synthetic analysis in the preface to the second edition of the *Critique of Pure Reason* (1787), *CrV* Bxxifn. It is the task of the

The significance of this method is particularly apparent in situations where we are under the pull of powerful seemingly conflicting considerations. Kant regards the antinomy as a kind of consistency test for the distinction between the rational (noumenal) and the sensible (phenomenal). The antinomy in reason is generated by reason's requirement of the unconditioned for explanation. In searching for explanations we cannot regard them as complete if the explanans in the explanation is itself in need of further explanation. On the other hand, it does not seem to be possible to come up with an explanation that is not in need of further explanation. An unconditioned explanans seems to be both necessary for the purposes of explanation and impossible. The antinomy resolves itself once one sees that the only way in which the unconditioned may be thought without contradiction is by thinking of the unconditioned as belonging to how things are in themselves completely independently of their relation to us and not as belonging to how things appear to us and are in relation to us. The distinction between intrinsic intellectual and relational sensible conditions on cognition of objects and actions becomes warranted through the avoidance of inconsistency in one's account of objects and actions that drawing the distinction between the intellectual standpoint on objects and the sensible standpoint on objects makes possible. In a wider sense, this avoidance of inconsistency is also the basis for a fundamental distinction between the theoretical and the practical. If one can come up with a coherent account of action only by distinguishing between the fundamental requirements of the theoretical and practical accounts of action, then this is a basis for regarding such a distinction as fundamental.

One must first distinguish the intellectual or rational from the sensible conditions for action in order to make sense of action. One must then be able to bring these two different standpoints on action together in relation to an individual action. One can only do this, Kant argues, by taking the basis of agency in an uncaused rational cause to be the most fundamental characterization of the action. In this way, one brings together rational and sensible conditions in that which is unconditioned in action. This unification of different perspectives on action is not just a unification of the intellectual and sensible conditions of action. It is also at the same time a unification of the practical and theoretical stances on action.

analytic to break a priori cognition in its theoretical and practical dimensions into their sensible and their rational elements through an experiment of pure reason. The metaphysician breaks a priori cognition into two heterogeneous elements, things as they must appear to us as we sensibly represent them and things as we conceive of them in themselves in analysis (in the Analytic), and then brings these two things together in the concept of the unconditioned as grasped by reason.

The experiment of reason that Kant offers as a warrant for pure practical principles provides a kind of proof of the validity of the moral law. But even if the experimental use of reason in evaluating common-sense moral judgments can in some sense show that the moral law is valid a priori, it cannot offer a proof of this that one can regard as theoretically compelling from outside of the moral point of view. And such arguments based on thought experiments cannot explain how it is possible for the moral law to be valid a priori (*CpV* 5:93). The kind of justification such an experiment of pure practical reason provides thus falls short of what a bona fide transcendental deduction of the kind offered in the first *Critique* for space and time and for the categories of the understanding is supposed to accomplish. Transcendental deductions are supposed to explain how a kind of a priori knowledge is possible on the basis of a certain kind of a priori representation.

The failure of the deduction of the supreme principle of pure practical reason turns on the connection between freedom as an efficient cause and the supreme principle of pure practical reason. The moral law is necessary as the supreme practical law for all rational agents to whom one ascribes freedom of the will; for all such rational agents have the capacity to be moved by the mere fitness of a principle to serve as a universal law: '[T]he two concepts [of freedom as cause and the supreme principle of pure practical reason] are so inseparably connected that one could even define practical freedom through independence of the will from any other [law] than the moral law alone' (*CpV* 5:94). To understand how an individual could be moved by reasons that motivate purely in virtue of their inherent universality, one would have to understand how a person could be moved to act independently of any antecedent states including the antecedent states of that person. One cannot understand how it is possible for an efficient cause to be free to act (independently of temporally antecedent reasons or causes) and so one cannot understand how the will could act independently of any causal law but that of morality. To assume otherwise would be to treat an unconditioned cause as something that is knowable by theoretical reason.

Kant consistently argues that it is not possible theoretically to prove the '*possibility* of freedom' (*CrV* A558/B586; *CpV* 5:94). One can show that the notion of freedom is not logically inconsistent with that of natural causation, but not explain how it is possible to act freely: 'nature at least *does not conflict with* causality through freedom – that was the sole thing we could accomplish, and it alone was our sole concern' (*CrV* A558/B586). The impossibility of explaining how freedom is possible on a priori principles leaves the conception of an intelligible world expressed by such freedom

theoretically incomprehensible. However, the commitment to freedom that is the mark of pure practical reason and the moral point of view commits us to regarding ourselves as agents whose actions can be motivated by nothing other than their intrinsic reasonableness and thus independently of whatever other considerations might cause us to act otherwise. This gives a practical reality to an intelligible world, since such an intelligible world is in effect implied by our notion of overriding moral obligations. We cannot, to be sure, know in the theoretical sense that we belong to such a world, since we have no basis that is independent of our morally obligatory commitments for regarding ourselves as members of such a world.

The logical possibility of an uncaused cause is, according to 'On the Deduction of the Principles of Pure Practical Reason', 'an analytic principle of pure speculative reason' (*CpV* 5:48). The inferential structure of pure practical reason is able to show the existence of an unconditioned causal condition that theoretical reason presupposes in all explanation to be a real rather than the merely logical possibility that it can only be for theoretical reason. The fundamental principle of morality and of pure practical reason becomes the warrant for a conception of uncaused causation that theoretical reason paradoxically needs to be truly possible for its idea of a complete causal explanation of an event:[7] 'For if as pure reason it is really practical, it proves its reality through what it does … With this faculty [of pure practical reason] transcendental *freedom* is also established, taken indeed in that absolute sense in which speculative reason needed it, in its use of the concept of causality, in order to rescue itself from the antinomy in which it unavoidably falls when it wants to think the *unconditioned* in the series of causal connection' (*CpV* 5:3).[8] Practical reason shows that the notion of an uncaused cause or 'absolute spontaneity' is something that we can make

[7] On the coherentist authentication of practical reason, see John Rawls, 'Themes in Kant's Moral Philosophy', in Eckhart Förster (ed.), *Kant's Transcendental Deductions* (Stanford: Stanford University Press, 1989), 108.

[8] Unlike Henry Allison, I wish to argue that it is Kant's theory of moral autonomy that is responsible for his insistence on an irreducibly intelligible dimension to free agency, and not his conception of theoretical reason: 'In order to understand Kant's seemingly gratuitous insistence on a merely intelligible moment of spontaneity in the conception of rational agency, we must look not to his moral theory or motivational psychology but to his views on the spontaneity of the understanding and reason in their epistemic functions' (*Kant's Theory of Freedom* (Cambridge and New York: Cambridge University Press, 1990), 36). Although Kant may have initially been attracted to the idea that self-consciousness provides one with consciousness of absolute spontaneity, at least by the time he wrote the second *Critique* Kant had come to the conclusion that only the kind of acting on reasons that is grounded in one's moral self-consciousness provides compelling evidence of one's ability to reason from absolutely spontaneous and hence unconditional starting points, since those reasons present themselves to one's moral self-consciousness as absolute and unconditional obligations. One is under those absolute obligations, and thus free to choose, even if one acts against those obligations.

sense of through the unconditional character of moral obligations. By showing the 'real possibility' of the kind of unconditioned cause that theoretical reason required to be at least logically possible in order for the notion of a complete explanation to make sense, pure practical reason contributes indirectly to the explanatory purposes of theoretical reason and the overall hypothesis of the unity of reason gains a warrant of its own.

The theoretical standpoint has as its interest the providing of a causal explanation for the occurrence of any given event. The goal of exhaustive causal explanation leads, however, to an antinomy. One can only make sense of a cause as something that precedes a given event. But to specify which of two events is earlier and which is later, one must presuppose a complete set of causal conditions that allow one to discriminate the spatio-temporal locations of these two events. This is why causal explanation seeks the cause of the occurrence of a given event in the complete set of causal conditions that are sufficient for the occurrence of that event. But since every set of causal conditions itself has causal antecedents, no complete set of causal conditions is in principle available to it. The only way one can make sense of a complete set of causal conditions is by thinking of the ultimate causes of a given occurrence as themselves uncaused. But such causes are inherently unknowable and inconsistent with the conditions for causal explanation, since they cannot be assigned a determinate spatio-temporal location. Spatial and temporal position can only be determined with respect to antecedent causes, relative to which it is a determinate fact of the matter whether a given event has occurred or not.

An uncaused cause would be something that cannot be assigned as such to a determinate spatio-temporal location. An event can only be assigned to a specific location in virtue of the events that are its causes and its effects. But the effects of that uncaused cause would be indistinguishable from the effects of a cause that is itself the effect of some antecedent cause. Events are only empirically knowable to the extent that they have causal antecedents that are themselves natural events whose distinctive set of causes and effects distinguish them from other events. The fact that uncaused causes are not theoretically distinguishable in the causal order from causes that have some antecedent cause makes uncaused causes something that the theoretical point of view cannot take to be anything more than a logical possibility, a possibility that is incomprehensible to the purely theoretical point of view. The theoretical incomprehensibility of transcendental freedom becomes key to its defence. For the commitment of the moral law to such freedom becomes the basis for a wider claim of coherence and intelligibility.

IV. WHAT IS WRONG WITH PSYCHOLOGICAL OR COMPARATIVE FREEDOM?

In the Critical Elucidation, Kant argues against those who claim that we can provide an explanation of how freedom is possible. The resources for such an explanation at the non-empirical level are just not there, since one cannot explain something except in relation to something else. To explain freedom in terms of something else would be to treat it as something conditioned rather than as unconditioned. This seems to leave only the possibility of an empirical explanation of how freedom is possible. But Kant sees the same problem here. He rejects the attempt to 'explain this freedom in accordance with empirical principles, like any other natural ability, and regard it as a *psychological* property, the explanation of which requires a more exact investigation of *the nature of the soul* and of the incentives of the will' (*CpV* 5:94). The moral law commits one to thinking of action for which one is morally responsible as something that is both ultimately independent of antecedent causes and that has effects that are spatio-temporal events. So it seems that one needs an account of how the moral law could motivate action, but Kant thinks that any theoretical explanation of such motivation will actually undermine the distinctive practical autonomy of the moral law. Any attempt to provide an empirical or psychological account of free agency will 'deprive us of the moral law itself, which admits absolutely no empirical determining ground' (*CpV* 5:94).[9]

Something needs to be said to defend the transcendental conception of freedom from those who think that freedom can be given an adequate psychological or natural explanation. To do so, Kant sketches the extent and limits of an empirical or psychological account of free agency. Like those who regard freedom and deterministic causal explanation as mutually compatible, Kant thinks that free agency should be compatible with the possibility of an exhaustive causal explanation of a certain action. Absence of causal determination is randomness, and randomness is no more conducive to responsibility for action than is exhaustive causal determination. There must be a sense in which one's action is part of a causal explanation that can

[9] 'But no insight can be had into the possibility of the freedom of an efficient cause, especially in the sensible world: we are fortunate if only we can be sufficiently assured that there is no proof of its impossibility, and are now forced to assume and are thereby justified in doing so by the moral law, which postulates it. For, there are many who believe that they can nevertheless explain this freedom in accordance with empirical principles, like any other natural ability, and regard it as a *psychological* property, the explanation of which simply requires a more exact investigation of the *nature of the soul* and of the incentives of the will, and not as a *transcendental* predicate of the causality of a being that belongs to the sensible world (although this is all that is really at issue here)' (*CpV* 5:94).

legitimately purport to capture it as an event, otherwise one is 'handing it over to blind chance' (*CpV* 5:95). If one's actions are random, then they do not have the requisite relation to one's choices that we require for personal responsibility for action.

Kant is willing to take the extent to which a deterministic account of a person's action is in principle possible to the limit. He maintains that if we could know everything about a person's life and the circumstances and incentives that motivate his or her behaviour, that behaviour would be as predictable, indeed as 'calculable', as that of an eclipse (*CpV* 5:99). This does not mean that he thinks that such predictability is a real possibility. One would have to assume at the very least that one could base one's predictions of a person's behaviour on observations of a person's incentives and the causal conditions under which they have a causal role in a person's behaviour without altering the psychological states involved. However, inner experience has the special difficulty that its object is changed through observation.[10] Even if precise prediction of psychological states were possible, then one would need to be able exhaustively to inventory all of the causal conditions relevant to a person's behaviour before one could offer such a prediction. Strict psychological and hence psycho-physical laws like the physical laws to be found in physics are not possible, since mathematics cannot be applied with the same kind of substantive success to inner experience as it can to events in outer experience.[11]

Kant's rejection of strict psycho-physical laws has suggested to some recent interpreters that his position might be a form of anomalous monism and that his account of freedom could be given an interpretation under which freedom of choice is fundamentally compatible with causal determinism. On that reading, he would take the relation between the mental and the physical to involve token–token identity between mental and physical events without psycho-physical laws governing the relation between mental and physical event types. Mental events themselves would not be subject to laws at all under their identity as mental events; the mental would be inherently anomalous.[12]

[10] Kant, *Metaphysical Foundations of Natural Science*, 4:471, in *Theoretical Philosophy after 1781*, ed. Henry Allison and Peter Heath (Cambridge and New York: Cambridge University Press, 2002), 186.

[11] *Metaphysical Foundations of Natural Science*, 4: 471, in *Theoretical Philosophy after 1781*, 186.

[12] Ralf Meerbote, 'Kant on the Nondeterminate Character of Human Actions', in W. L. Harper and R. Meerbote (eds.), *Kant on Causality, Freedom, and Objectivity* (Minneapolis: University of Minnesota Press, 1984), 138–63, and also Hud Hudson, *Kant's Compatibilism* (Ithaca: Cornell University, 1994).

Kant thinks that the principle that 'every event is determined by a cause according to constant laws' belongs to a universal science of nature. Such a science of nature brings together 'the object of the external senses or that of the internal sense (the object of physics as well as psychology), under universal laws'.[13] This might still be consistent with anomalous monism if the universal laws governing objects of psychology govern those objects only in so far as they are physical objects, that is, as purely physical laws. While Kant does not actually deny the existence of natural laws governing psychological states, he does maintain that there are no substantive natural laws for psychology that can be known a priori, since there are insufficient resources for the construction of objects in inner experience a priori. The existence of natural laws governing psychological events is not entirely ruled out; these natural laws would, however, be merely empirical psychological laws.[14] It is true that in his Remark II to Theorem II, Kant not only demands of any law of nature that it be knowable a priori, but also expressly rejects anything as a practical law that would have the natural necessity in question. However, even there Kant does not want to rule out the possibility of a way of looking at things in which 'the action is as unavoidably forced upon us by our inclination as is yawning when we see others yawn' (*CpV* 5:26). What he does argue is that one cannot capture the principles in terms of which agents make their choices from the vantage point of such causal laws.

The anomalous monist regards choices as causal only under a physical or at least natural description of choices. Mental events are token–token identical with such physical events and function as singular causes that are not covered by psychological or psycho-physical laws. There is little evidence to support the attribution to Kant of a singular causal view of event causation.[15] But it does not help one either. One can treat event causation as a relation between individual events, and still if the identity of an individual event is tied to its causal antecedents as their effect, then that event couldn't as such be otherwise, unless its individual cause were different. This is a

[13] Kant, *Prolegomena* section 15, 4:295, in *Theoretical Philosophy after 1781*.

[14] Evidence for Kant's commitment to the existence of psychological laws is advanced by Gary Hatfield, 'Empirical, Rational, and Transcendental Psychology: Psychology as Science and as Philosophy', in P. Guyer (ed.), *The Cambridge Companion to Kant* (Cambridge and New York: Cambridge University Press, 1992), esp. 218–23.

[15] It is possible to read Kant's position in the Second Analogy of the first *Critique* (where he develops the general causal principle) to commit him only to individual causes for every event. A singular causal account is defended by Lewis White Beck, 'A Prussian Hume and a Scottish Kant', in *Essays on Kant and Hume* (New Haven: Yale University Press, 1978), 11–129, and Henry Allison, *Kant's Transcendental Idealism* (New Haven: Yale University Press, 1983) and *Kant's Theory of Freedom*, 33–4. I have argued against such a singular causal view in my *Kant and the Demands of Self-Consciousness* (Cambridge and New York: Cambridge University Press, 1997), 141 ff.

matter not just of the causal covering laws that support counterfactual claims about the event, but about the causal relation itself. It does not matter whether one thinks of causal conditions as individual conditions or as general (lawlike) conditions. To the extent to which one takes oneself to be motivated to act by a reason to act that can be understood without remainder in terms of a particular temporal position in one's life history, one must take oneself to be exhaustively subject to the antecedent causal conditions that determine the location of events in space and time. Any reason is a cause only in so far as it is identical with a physical cause that is in turn the effect of antecedent physical causes.

Such a view allows for a kind of freedom of choice. However the kind of choice that it allows is a causally conditioned choice. Regardless of whether mental events are type–type or only token–token identical with physical events, or even distinct but correlated token–token or even type–type by means of psycho-physical laws, the conception of choice entailed by propositional mental life is only a causally conditioned choice.[16] Whether mental events are caused or causes only under a physical description or not is of some interest to Kant, but it is largely beside the point when it comes to the problem of freedom of choice for him. The important point is that they are caused. Reasons are causes, but such causes can in turn be caused reasons for choice as well as cause of action. What matters is the basis upon which a choice is made.

In a certain sense when we act on reasons, we are acting independently of our causal environment. We are not being caused to do what we are doing by anything outside of us. This is why acting on reasons involves a certain kind of freedom of choice. However, even if we think of the agent as being motivated 'through instinct or determining grounds thought by reason' those motivations give us only a very thin 'psychological freedom' that is tied to a 'natural necessity' if they have 'the ground of their existence in time and indeed in the *antecedent state*, and this in turn in a preceding state' (*CpV* 5:95). Kant rejects the view that reasons-based action is already as such free action in no uncertain terms: 'One assumes … for instance Wolff and Baumgarten, that the human being who acts is independent of all natural

[16] I do not think that proponents of the anomalous monism interpretation have made the case that the anomalousness of the mental, even if it obtains, renders the mental causally unconditioned. If anything, it shows that the mental allows for a kind of psychological freedom that falls short of unconditioned freedom. See Ralf Meerbote, 'Which Freedom?', in P. Cicovacki (ed.), *Kant's Legacy* (Rochester: University of Rochester Press, 2001), 197–225: 'Appropriation and apperception are needed for freedom in their own right. This is so because anomalist monism treats all propositional mental life as anomalous and hence as unconditioned, but not all mental life of this sort is free' (225n.).

necessity, in so far as his actions are governed by motives, that is, determined through understanding and reason; but this is false. A human being is not freed from the mechanism of nature through undertaking an actus of reason in his action. Every act of thought, of deliberating, is itself an event of nature in which the understanding searches for a connection of the things' causes with their effects, and chooses the means for acting accordingly: only this actus is an inner occurrence, since it takes place in the human being himself.'[17] If reasons for action are understood as psychological facts in a person's history, then our judgment merely rationalizes a motivation to act that can be understood purely in terms of its position in an agent's causal history. They have 'psychological instead of mechanical causality' (*CpV* 5:95). Yet they are nevertheless determined by antecedent causes and thus involve a form of psychological mechanism. Kant makes this point elsewhere, as well: 'Supposing, now, that every action a man undertakes by the use of his reason were also to be grounded in the time preceding, then it would surely take place only with respective, not absolute spontaneity; for though it lay with reason in the first place, the latter was determined in the time preceding, and thus unconditioned self-activity would not be present in it.'[18]

A certain kind of freedom of choice is possible within the framework of predictable mental and physical events. Such choice is free in the sense that the choice is made by the agent and not directly caused by something or someone outside of the agent. This freedom of choice belongs to a causal series and is conditioned by antecedent events and is in this sense only what Kant calls a 'comparative notion of freedom' (*CpV* 5:96). Such comparative freedom is the kind of freedom to which soft determinists subscribe. A comparative notion of freedom would involve a form of spontaneity or self-directed activity. That is, if an agent has comparative freedom, the agent's actions are free in the sense that the actions of the agent are not constrained by something outside of the agent. Kant invokes for this purpose the example of a projectile in flight. Things have comparative freedom or relative spontaneity that operate according to an internal mechanism. Things that operate according to an internal mechanism are governed in their activity by an internal principle, and thus free in comparison with something that has no internal mechanism at all. However, such things are

[17] See 'Notes on the Lectures of Mr. Kant on the Metaphysics of Morals', Vigilantius (1793–4), 27:503, in Kant, *Lectures on Ethics*, trans. Peter Heath and ed. Peter Heath and J. B. Schneewind (Cambridge and New York: Cambridge University Press, 1997), 269–70.

[18] 'Lectures on the Metaphysics of Morals', 27:505, in *Lectures on Ethics*, 270–1.

comparable to projectiles in flight in that the internal principle according to which such things operate is itself outside of the control of the things that are operated by those internal principles.

Kant sets up a problem for the compatibilist who advocates a 'comparative notion of freedom'. He argues that a conception of free agency according to which an individual's action is in fact exhaustively determined by antecedent causes would prevent the kind of control over one's actions that are required for free agency and moral responsibility. According to the comparative notion of freedom it is logically possible for the individual to make a different choice from the one that he does make. However, this does not mean that it is 'really possible' for the individual to make a different choice, since that choice would require a different causal history for the choice. Since the agent's past actions are one and all determined by reasons for action that are themselves causally conditioned ones with a specific, determinate temporal position, such events are outside of the agent's control.

Not only are one's actions outside of one's control. One's choice is outside of one's control. One's very character is in some sense outside of the person's control, since one's character is the expression of innate characteristics or those acquired from one's past actions, and thus outside of one's present control. Even complete independence from outside causal influences would not give one the requisite control as long as one's past psychological states have a determinate connection to one's present choice. For in this case one's present state of choice would always be determined by a past state over which one now no longer has control (*CpV* 5:97).

If we were limited to this comparative notion of freedom, we would have the same kind of freedom to act that a turnspit (rotisserie) or a clock has: once the thing is set in motion it operates according to its own internal mechanism.[19] It may encounter both internal and external impediments to its action, but these impediments are nothing but the limitations imposed

[19] Kant's pre-critical *Lectures on Metaphysics* edited by Pölitz develop this distinction between comparative and absolute freedom in terms of the distinction between relative and absolute spontaneity: 'Spontaneity is either *absolute or simply of a certain kind* (*secundum quid*). *Spontaneity* is *of a certain kind* (*secundum quid*) when something acts spontaneously *under a condition*. So, for example, a body that is propelled moves itself spontaneously, but in a certain way (*secundum quid*). This spontaneity is also called *automatic spontaneity* when namely a machine moves from itself according to an inner principle, e.g., a watch or a turnspit. The spontaneity is however not *simply such* because the inner principle is determined through an external principle. The inner principle of the watch is the spring, of the turnspit, the weight; but (the external) principle is the artisan who determines the inner principle. The spontaneity that is *simply such* is an absolute spontaneity.' See *Metaphysik L*$_1$, ed. Pölitz, 28: 267–8, in Kant, *Lectures on Metaphysics*, ed. Karl Ameriks and Steve Naragon (Cambridge and New York: Cambridge University Press, 1997), 80.

on the unconstrained operation of its internal mechanism. We could think of ourselves as material automata, on the model of a clock that winds itself up, a position famously defended by La Mettrie.[20] Alternatively we might think of ourselves, with the Leibnizians, as self-conscious, spiritual automata wound up by God, with all the experiences, motivations and actions that we will ever have (*CpV* 5:97).[21] In neither case would we be truly free.

Kant accepts the idea that God creates us, and even that he creates our natures, our substance. However, he argues that our actions would not be free, we would be a form of self-conscious automaton, if all of our actions were ultimately caused by God.[22] Kant proposes to avoid this kind of fatalism by insisting that although God may have created us as individuals or substances, God is not the creator of our psychological and physical states.[23] The latter are mere ways in which our nature as individuals must appear to us in experience. The claim that God creates us as things or substances – rather than creating the states that we must ascribe to ourselves (given how we must appear in experience) – has to be understood to involve a more limited notion of creation than that to which the Leibnizian is committed if it is to help.

For Kant, God creates a set of capacities in us through which we are able to choose who we are to be. God is responsible for creating the (non-phenomenal) character that gives us the capacity to make moral choices. God is not, however, responsible for creating the whole character that each of us has. God creates only a specific capacity to act morally and rationally in a creature with needs. He is not responsible for creating the specific needs

[20] Julien Offray de La Mettrie, *Machine Man and Other Writings* (Cambridge and New York: Cambridge University Press, 1996): 'The human body is a machine that winds up, a living picture of perpetual motion … I am not deceived the human being is a clock', 6.

[21] G. W. Leibniz, *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil* (London: Routledge and Kegan Paul, 1951): 'All is therefore certain and determined beforehand in man, as everywhere else, and the soul is a kind of spiritual automaton', Section 52, 154. Kant argues that the Leibnizian compatibilist, the defender of comparative freedom, is forced into a kind of fatalism that is akin to the position of Spinoza (*CpV* 5:100–1). If you regard all of our states and actions as causally determined by our inherent natures and you regard our natures as created by God, then God becomes the cause of everything we do.

[22] In an early work, the *New Elucidation of the First Principles of Metaphysical Cognition* of 1755, Kant argued that all actions might have an ultimate cause in God, but nevertheless be free. He took actions to be free so long as they derived from an internal principle of the agent. He was not yet committed to the strong conception of alternative possibilities for action in that early work that he would later require; freedom did not require the capacity to act otherwise. See *New Elucidation*, 1:400 ff., in Kant, *Theoretical Philosophy, 1755–1770*, trans. and ed. David Walford and Ralf Meerbote (Cambridge and New York: Cambridge University Press, 1992), 22 ff.

[23] Kant is probably responding to an objection made by his anonymous critic Pistorius in a review of the *Groundwork*, as has been pointed out by Eberhard Schulz, *Rehbergs Opposition gegen Kants Ethik* (Berlin: de Gruyter, 1975), 146.

and desires or even the neediness of the creature in question. Neediness together with specific needs and desires arise from the limitations of being created. Thus God is responsible for the positive capacity to act according to the moral law, while our tendency to act contrary to the moral law is connected with being the sensible needy creatures that we are. However, even our neediness does not explain our tendency to give preference to our own self-interest over what we recognize to be reasons. Here is where Kant thinks that our explanations must come to an end.

V. THE THIRD ANTINOMY AND TWO STANDPOINTS ON ACTION

We can look at the person's causal history to try to understand why the person did what he did. Thinking of what the person does in terms of the person's causal context compels us to think of the person's behaviour as if it were in principle completely predictable given a sufficient amount of information. There is, however, a different standpoint that we take on a person's behaviour in so far as we hold the individual responsible for what he or she does. We look at the rightness or wrongness of the action and not the causal antecedents to answer the question why the person did what she did. From this standpoint, the individual 'also views his existence *in so far as it does not stand under conditions of time* and himself as determinable only through laws that he gives himself by reason and in this existence of his nothing is, for him, antecedent to the determination of his will' (*CpV* 5:97). From the standpoint of moral autonomy the individual regards himself as being able to use his pure practical reason to legislate principles to himself that are not determined by antecedent causes.

Kant's fusion of incompatibilism and compatibilism needs to be able to explain how one and the same action can be both caused by antecedent events and nevertheless be an action for which one can be held responsible. The causal explanation of action seems to entail rejection of the principle of alternative possibilities and that a person could have done otherwise. On the other hand, Kant's commitment to moral autonomy and the principle that 'ought implies can' requires real alternative possibilities: '[If a] deed is, in accordance with the natural law of causality, a necessary result of determining grounds in preceding time, then it was impossible that it could have been left undone; how, then, can appraisal in accordance with the moral law make any change in it and suppose that it could have been omitted because the law says that it ought to have been omitted?' (*CpV* 5:95). This leads to a conundrum. How can we hold individuals morally responsible if their

actions are already causally necessitated by antecedent events? If actions of persons must occur as a consequence of antecedent causes their actions would seem to be outside of their 'control'. Kant has a version of the same problem that he poses for the proponents of the comparative or psychological notion of freedom. How can we make sense of the fact that an action can be regarded as both causally necessitated and yet also as independent of antecedent causes?

Kant makes his case for responsibility particularly difficult for himself. He argues that every unlawful action that one does while one is in command of one's faculties is one that one could have avoided doing. He offers two different kinds of examples of this moral culpability. One might be tempted to excuse oneself from an unlawful action as an 'unintentional fault', a bit of misconduct due to bad habits or inattentiveness. One is tempted to offer a causal account of why one has mitigated responsibility and blame in such cases. Still one's conscience is not disposed to let one off so easily, but finds one guilty nevertheless (*CpV* 5:98). Even the person who seems to have a bad character from the outset is to be held responsible for this character. An individual, 'taken to be a born villain', to whom a 'hopeless natural constitution of mind' is ascribed, is nevertheless censured for his crimes (*CpV* 5:100). If you think that blame should be mitigated by causal circumstances, then Kant argues that you should think that the person who is born with a set of bad dispositions should not be held accountable. Similarly, one ought then to hold a person who grew up under extreme circumstances to be less accountable for her actions than a person who did not. While Kant does not think that one's causal history can undermine one's control over what one does in such a way as to mitigate moral responsibility, he does think that it can make it harder to do the right thing and thus make an action more meritorious.

Kant appeals to the first *Critique* and the Third Antinomy in order to 'resolve the apparent contradiction between the mechanism of nature and freedom in one and the same action' (*CpV* 5:97).[24] In the Antinomy, Kant argues that a metaphysical realist must either assume that every event has a set of causes that exhaustively determine its occurrence or that there are some events that do not have an exhaustively determining set of antecedent causes. Either of these positions threatens the kind of freedom to which Kant takes the moral law to commit us. The thesis shows that the notion of an exhaustive set of causal conditions makes sense only if you can assume

[24] We hold individuals responsible and yet, as he puts it in the first *Critique*, 'before they ever have happened, their actions are one and all predetermined in the empirical character' (*CrV* A553/B581).

that there is a way of articulating causal conditions that is not itself contextual. In order even to make sense of causal determination, you have to treat causal explanation as if it were possible in principle exhaustively to inventory the causal conditions that are relevant to the occurrence of any event. Assuming the standpoint of metaphysical realism, one must then assume that these causal conditions have a reality that is completely independent of the context of inquiry. But this is incompatible with the assumption that these causal conditions are in principle fully specifiable, since this specifiability presupposes the determination of causes in terms of temporal antecedents that in turn presuppose antecedent causes.

The conclusion of the thesis that some events are not exhaustively determined to occur by antecedent events seems initially promising for the notion of free agency, as Kant notes in his comment on the thesis of the Antinomy (*CrV* A450/B478). However, in the antithesis of the Antinomy, he shows that if one assumes such indeterminacy, then one is incapable of making sense of the relation of an action to antecedent events, either to those antecedent to the performance of the action or even of the action as outcome to the event that was the bringing about of that outcome, since one has no adequate basis for relating these events to each other in time (*CrV* A445/B473). The result is that one cannot make sense of agency as a causal temporal notion. Basically Kant argues that the metaphysical realist is impaled on Hume's Fork. We cannot make sense of action that is exhaustively determined by antecedent causes, since we do not have a determinate notion of cause that satisfies the constraint in question. But we cannot seem to make sense of action that is not determined by antecedent causes, since that action then seems to be to that extent merely random and have no determinate temporal relation to what comes before it.

In his Resolution of the Antinomy, Kant argues that we can avoid the problem for explanation and at the same time show how we can make sense of action. The explanation of an action seems to commit us to the existence of a complete set of causal conditions. This complete set of causal conditions makes sense only as a distributive rather than as a collective unity. Kant argues that we ought not and indeed cannot give up on the principle that every event has a cause from which that event necessarily follows. But this cause is itself only sufficient for the occurrence of that event if it is understood as a complete set of the causal conditions relevant to the occurrence of that event. For any event it is true that that event would not have caused another event, were it not the case that certain conditions were true. Since these causal conditions can in turn be specified only in relation to further causal conditions which would allow us to fill out the

ceteris paribus clauses implicit in the articulation of the relevant conditions, we seem to require a complete set of causes in order to make sense of the cause of any given event. Kant thinks that we can treat such a complete set of causal conditions as if it were something that could be specified, so long as we think of this merely as a regulative ideal and imaginary focal point for inquiry, rather than as something that has a reality that is completely independent of the needs and demands of inquiry. To think otherwise is to fall victim to the transcendental illusion implicit in the aim of causal inquiry and in the explanation of causal events and actions. Once we see the notion of a complete set of causes as the imaginary focal point of inquiry, we are also free to take an alternative standpoint on the explanation of action. This is one that explains action in terms of normative reasons. These normative reasons are not the psychological states that explain why one has done something, since these psychological states are every bit as much a part of the causal sequence of events as are physical causes. Rather they are the considerations that are independent of one's particular spatio-temporal situation that justify the course of action in question and require the action in question of us.

VI. TIMELESS CHOICE AS AUTONOMY

There are two different ways of regarding one and the same action. From the temporal, and hence phenomenal point of view, one looks at an action as a part of a temporal sequence. The action is situated at a temporal position relative to one's own capacity for temporal (and spatial) self-location. This involves the temporal position of the action in a causal series in which the event is uniquely determined by means of causal laws. The action understood in this way is an expression of one's empirical capacity for choice. The empirical character of practical reason and the power of choice is one's capacity for rational decision-making in so far as the exercise of that capacity is embedded in the overall context of the causal order of nature in which its choices all have antecedent causes (*CrV* A549/B577). In respect to this empirical character of reason every choice that an individual makes is completely predictable (*CrV* A550/B578). The noumenal character is the intellectual power of choice that is the 'sensibly unconditioned condition of appearances' (*CrV* A557/B585). The distinction between the phenomenal and noumenal character of rational choice corresponds to the distinction between psychologically or comparatively free choice and transcendentally free choice. Choices based on one's phenomenal character are rational, but are based on antecedent causes that determine one's choices. Choices based

on the noumenal character are based on the intrinsic reasonableness of the action independently of how it bears on one's particular spatial and temporal location. This point of view on an action in which we look at the reasonableness of the action in unconditional terms is independent of time and antecedent causal conditions.

From the standpoint of the phenomenal character and its psychological freedom, the agent's choices are determined by natural necessity and outside of the person's control. From the standpoint of our noumenal character, each of us views our actions as stemming from reasons that are unconditioned by causal antecedents.[25] From this standpoint, you are the ultimate source of your own choices and all of the events that affect you. What that means for Kant is that the choices that you make are regarded from the vantage point of their intrinsic rationality, rather than as choices that are determined for the individual by that person's causal history. To the extent that you are guided by or even take the moral law to apply to you, you are determining yourself with respect to a space of reasons that allows you to determine yourself according to universal principles that are independent of your causal situation. It is in this sense that one must understand the timelessness of the noumenal character.

Kant might seem to envisage an individual who timelessly chooses a set of basic principles for action that underlie his whole conduct throughout his life. The conception of freedom as a timeless choice of character is strongly suggested by the idea that one provides oneself with a character that includes all of one's causal history in it and that makes one responsible for all of the choices that one makes.[26] Such a timeless choice might be thought to be 'simultaneous with each act as it occurs in the temporal order' in so far as the timeless choice manifests itself in an intelligible cause that is the source of its efficacy.[27] One obvious problem with the notion of a timeless choice is that it seems to be unintelligible. Choice seems to be an event, and as such something temporal, and the simultaneity of a timeless choice with temporal choices also seems to be a difficult notion to make sense of in non-temporal terms.

[25] Andrews Reath, 'Kant's Critical Account of Freedom', in Graham Bird (ed.), *A Companion to Kant* (Oxford: Blackwell, 2006), 275–90, has a helpful discussion especially of the intelligible character and the way that it grounds the empirical character of choice. Reath notes but does not emphasize that the empirical character 'includes a person's reasons (the rational grounds of action) and subjective principles', 284.

[26] Allen Wood reads Kant this way in 'Kant's Compatibilism', 90 ff.

[27] Wood, 'Kant's Compatibilism', 90 ff.

Given its limited intelligibility, one will only want to ascribe the notion of timeless choice to Kant if there is no other alternative. Kant is committed to the possibility of reasons for choice that are reasons for choice independently of particular circumstances. But there is no reason that he need be committed to acts of choosing to do something that are outside of time. In any case in which I choose, my choosing to do what I do may be traced to vicissitudes of my causal history. But then the reason upon which I am acting cannot be regarded as a truly general one or universal reason for doing what I do. When I regard myself by contrast as a rational agent who is part of the space of reasons, then I regard my choices as subject to a generality constraint that renders them in principle independent of my causal and my temporal circumstances. This is the principle involved in moral autonomy. It is the way that I must regard myself in so far as I regard myself as an agent endowed with the capacity for moral autonomy and morality.

Kant thinks of the commitment to a self-determining causal law of practical reason to be the hallmark of the intelligible character of choice. This intelligible character of choice and self-determination is warranted in so far as it is implicit in the general principle of choice that is built into common-sense morality: 'Hence nothing remained but that there might be found an incontestable and indeed objective principle of causality that excludes all sensible conditions from its determination, that is, a principle in which reason does not call upon something *else* as the determining ground with respect to its causality but already contains this determining ground by that principle, and in which it is therefore as *pure reason* practical. Now, this principle does not need to be searched for or devised; it has long been present in the reason of all human beings and incorporated in their being, and is the principle of *morality*' (*CpV* 5:105). Moral autonomy provides us with an unconditioned causal law of reason that allows us rationally to determine ourselves independently of antecedent states. This does not mean that I can or thereby do (theoretically) know myself to be acting on reasons that are independent of my causal circumstances. Instead when I look at myself from the vantage point of my moral responsibility, I take myself to be capable of acting on unconditioned reasons, reasons that are completely independent of my specific causal situation.

Any causal account of how I act will have to be interpreted in such a way that it can do justice to this basic commitment of practical reason. One can accommodate such an account in the causal explanation of my behaviour to the extent that one acknowledges that the causal account of my behaviour in terms of my phenomenal character and other causal conditions governing my choice is never complete. The timelessness of the noumenal character

allows one to take it to determine a different outcome for the empirical character in a situation of choice, even though the empirical character is determined in what it does by antecedent causes in its own history and context (*CpV* 5:98).[28] It is only because the empirical character has 'at its basis a free causality' that is grounded in the noumenal character that one can act otherwise (*CpV* 5:100).[29]

On the reading of the noumenal character of rational choice that I am offering, this noumenal character is nothing other than the principle of moral autonomy. I take autonomy for Kant to be fundamentally a capacity. We exercise this capacity to some degree whenever we make reasoned choices. Such choices involve a commitment to the intrinsic reasonableness of what is chosen, although we may be mistaken, even fundamentally mistaken, in what we take to be reasonable. In engaging in practical reasoning, we are guided by generality constraints that have an implicitly universal character to them. We cannot reason about what to do without doing so in general terms that involve thinking of what we are doing as what is appropriate in some sense for anyone to do. This is in some sense true even when we are guided by purely selfish considerations. For even in putting such considerations to the fore, we do so because we think that

[28] The capacity to regard oneself and one's reasons for choice in a timeless, or temporally neutral, way allows one to make sense of the kind of agent regret that one feels in having done wrong, an agent regret that hard determinists like Priestley are compelled to treat as irrational. See Joseph Priestley and Richard Price, *Free Discussion of the Doctrines of Materialism and Philosophical Necessity* (London: Johnson and Cadell, 1778), 302. Priestley holds that no one will feel shame, remorse or repentance, but that one cannot avoid attributing actions to oneself rather than to their primary causes. A view along this line was also defended by Kant's German contemporary, Johann Heinrich Schulz, *Versuch einer Anleitung zur Sittenlehre für all Menschen, ohne Unterschiede der Religionen* (Berlin, 1783). Kant argued in his review of Schulz's work that even the fatalist must treat his actions 'as if he were free' (8:13 in Kant, *Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1996)).

[29] When we hold ourselves and other agents responsible for what we do, we think of the reasons upon which we act as 'principles freely chosen' (*CpV* 5:100). What this means can be illuminated by looking at the principle of incorporation that Kant uses in his later work to define freedom of choice. According to the principle of incorporation, an incentive can determine choice only to the extent to which the individual has taken it up in a maxim or general rule according to which he intends to behave. See *Religion within the Limits of Pure Reason Alone*, 6:52, and *Conflict of the Faculties*, 7:44, in Kant, *Religion and Rational Theology*, ed. Allen W. Wood and George di Giovanni (Cambridge and New York: Cambridge University Press, 1996). For further discussion of this principle see Andrews Reath, 'Kant's Theory of Moral Sensibility', in *Agency and Autonomy in Kant's Moral Theory* (New York: Oxford University Press, 2006), 12 ff. The principle of incorporation means that a desire or inclination does not itself constitute a reason to act, but it does so only as incorporated in a general principle of action, as noted by Henry Allison, *Kant's Theory of Freedom* and especially 'Kant on Freedom: A Reply to My Critics', in *Idealism and Freedom: Essays On Kant's Theoretical and Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1996), 118–19. However, the act of incorporation is itself only uncaused choice when it is understood not from the vantage point of the agent's causal history, but as an expression of the fundamental maxim of choice that makes a person the person who she is.

the desirability of the ends that they proffer is enough to confer rationality on the best means of their achievement.

One need not act in a manner that is genuinely morally autonomous in order to hold oneself responsible to the principle of moral autonomy. But when one does act in a manner that is genuinely morally autonomous, one's choices are ones that are independent of one's causal history and thus express something more than a relative freedom of choice. Transcendental freedom, or what Kant also calls 'absolute spontaneity', is something that has an essential relation to our capacity for moral autonomy. By this Kant understands the capacity to act according to universal principles that are internal to our capacity for choice. We know ourselves to be free in the sense of having absolute spontaneity in so far as we are conscious of moral obligations, and the nature of the principle of moral obligations is something that emerges from the structure of absolute spontaneity as determining oneself to act independently of antecedent causes.[30]

Thinking of spontaneity as self-determination links the notion of spontaneity to that of the seemingly very different conception of freedom as autonomy and simultaneously illuminates the relation between negative and positive freedom that Kant insists upon in Theorem IV of the Principles in the Analytic of Pure Practical Reason (*CpV* 5:33). The negative notion of freedom is freedom from alien or outside influences. Now this freedom from alien influences cannot be defined or articulated without reference to what is not alien, what belongs to someone, something or some state itself. In other words, an implicit form of self-reference of some kind is built into any conception of negative freedom. This leads Kant to argue that negative freedom presupposes a positive notion of freedom as some form of self-determination. Kant argues that true self-determination is only possible in the form of self-legislation or autonomy. This means that whenever one acts

[30] Absolutely spontaneous causality, causality from freedom in the cosmological sense, is causality 'whereby a series of appearances, which proceeds in accordance with laws of nature, *begins by itself*' (*CrV* A446/B474). Absolute spontaneity is the 'capacity to begin a state by itself'. The temptation is to interpret spontaneity along the lines of randomness, as in the phrase 'spontaneous combustion'. Beginning a state spontaneously or 'by itself' is not a pellucid turn of phrase, so it is tempting to understand the term 'spontaneity' merely as the capacity to act without being acted on. But one needs to take the reflexive characterization of spontaneity seriously. Spontaneity is really the capacity for determining oneself to act, as opposed to being determined to act by something outside of oneself or one's state. This is why Kant refers to spontaneity as an act of 'self-activity' or of self-directed activity: 'Freedom signifies absolute spontaneity, and is the self-activity [*Selbsttätigkeit*] out of an *inner principle* according to the free choice [*Willkühr*].' See *Metaphysik L*$_1$, ed. Pölitz, 28.: 267–8, in *Lectures on Metaphysics*. The use of self-activity to describe spontaneous activity is also prominent at the beginning of the second edition of the Transcendental Deduction in the first *Critique*, at section 15, *CrV* B130.

in a manner that is self-determining one stands under the normative principle of autonomy.

Christine Korsgaard has given the argument from spontaneity to autonomy a characterization that is helpful in linking spontaneity to self-legislation, but that is also cryptic on precisely why spontaneity requires the strong notion of self-legislation that Kant thinks that it does. In developing what she calls the argument from spontaneity to freedom, Korsgaard argues that one is led inexorably from spontaneity to self-legislation and autonomy. If one thinks of an agent as purely spontaneous in his or her choices, Korsgaard maintains that there is nothing left to determine the will apart from the law. She can appeal to Kant's own talk of freedom as something that requires self-determination. More importantly, she can appeal to Kant's thesis that all self-determination requires self-legislation. Following Kant, Korsgaard argues that this is so because nothing at all can operate without operating according to a law. If something is to operate in a manner that is self-determining it must give itself its own law. It must be autonomous. However, this still leaves it rather obscure as to why self-determination requires self-legislation.[31]

Self-determination, self-directed activity, is largely a relative notion, hence the notion of relative spontaneity. However, implicit in all self-determination is an ideal of determination by self rather than by facts that are independent of the self. This pushes the notion of self-determination towards a standard that is purely self-imposed and that applies to one independently of whatever matters of fact may obtain. This is what Kant calls 'absolute spontaneity'. Such absolute spontaneity can be adequately expressed only in terms of the notion of universal self-legislation, that is, determining oneself to act from a principle merely because that principle is appropriate to be adopted by absolutely everyone no matter what the circumstances. It is to choose a principle to guide one because this principle would be an appropriate principle for anyone in any circumstances to adopt for themselves. In this way, one abstracts from anything about one's own past history that might be determining one's choice behind one's back. Thinking of oneself as having the capacity for such self-legislation, a self-legislation built into Kant's conception of moral autonomy, allows one to see oneself as having (without theoretically knowing that one has) the kind of control over oneself that we require in order to make sense of our capacity for moral responsibility.[32]

[31] See especially Christine Korsgaard, 'Morality as Freedom', in *Creating the Kingdom of Ends* (Cambridge and New York: Cambridge University Press, 1996), 159–87.

[32] I want to thank the other contributors and especially Andrews Reath for many interesting and helpful suggestions and critical comments.

CHAPTER 7

The Antinomy of Practical Reason: reason, the unconditioned and the highest good

Eric Watkins

The Antinomy of Practical Reason has received far less attention than most other sections in the *Critique of Practical Reason*, despite the fact that few are seriously tempted to reject its fundamental systematic and philosophical importance within Kant's practical philosophy.[1] This comparative neglect is both unfortunate and remediable. It is unfortunate because focusing on it makes clear the conception of practical reason that Kant uses and needs to use within his practical philosophy as a whole. It is remediable because its basic features can be determined without excessive difficulty if one pays attention to the significant parallels it displays with the *Critique of Pure Reason*'s Antinomy of Pure Reason.

In a first section, I briefly present the basic structure of the Antinomy of Pure Reason, paying special attention to the role that reason plays in generating this antinomy. In the second section, I reconstruct the Antinomy of Practical Reason with the help of several basic structural similarities that it bears to the theoretical antinomy. In a third section, I turn to several basic questions that arise with respect to the Antinomy of Practical Reason's central concept of the highest good, and use insights gained from the comparison of the theoretical and practical antinomies to develop detailed answers.

[1] The Antinomy of Practical Reason has been discussed most fully by Lewis White Beck in the course of his commentary on the *CpV* (*A Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960)); by Allen Wood in his systematic treatment of Kant's moral religion (*Kant's Moral Religion* (Ithaca: Cornell University Press, 1970)); by Michael Albrecht (*Kants Antinomie der praktischen Vernunft* (Hildesheim: Georg Olms Verlag, 1978)); and by Victoria Wike (*Kant's Antinomies of Reason: Their Origin and Resolution* (Washington, D. C., University Press of America, 1982)). Beck's remarks are themselves remarkable: '[W]hen we examine the antinomy we shall find that it is really quite a poor thing, wholly unable to carry this great historical and systematic burden. We shall also find, regrettably, that Kant's usual high-quality workmanship is not much in evidence in the discussion of the antinomy' (246).

1. THE ANTINOMY OF PURE (THEORETICAL) REASON

In the Antinomy of Pure Reason of the *Critique of Pure Reason*, Kant claims that if one accepts what he calls Transcendental Realism – the view that appearances and things in themselves are identical – then reason can come into conflict with itself, because it can prove contradictory propositions with respect to certain features of the world. Kant counsels us to respond to such antinomial conflict by rejecting Transcendental Realism in favour of Transcendental Idealism, which, as a first step, distinguishes between things in themselves and appearances, before then asserting, among other things, that the former can be neither spatio-temporal nor known, while the latter are 'ideal' in that they depend for their existence on purely subjective forms of intuition. In so far as Transcendental Idealism is Kant's most distinctive and most important doctrine, the Antinomy of Pure Reason plays a central role in his entire philosophical system and it is worth attaining clarity about the commitments he makes that give rise to it.[2]

Specifically, the contradictions expressed in the Thesis and Antithesis in each of the four Antinomies are generated not simply by the assumption that things in themselves and appearances are identical, but rather by reason's attempt, given this assumption, to make assertions about *determinate* features of the *world as a totality*. Thus, the Thesis and Antithesis of the First Antinomy argue that the world as a totality has and does not have a first moment in time and an outer edge in space, i.e. that the world as a totality has a determinate magnitude that is, impossibly, both finite and infinite. Similarly, the Second, Third and Fourth Antinomies each focuses on different kinds of determinate features that the world must have as a totality, specifically, on its ultimate quality or composition (as composed of simples or something infinitely divisible), on its causal origin (as containing or precluding first or uncaused causes, such as freedom) and on its modal status (as containing or not containing necessary beings), arguing in each case that the world must have both finite and infinite instances of these features.

Kant's resolution of the Antinomy of Pure Reason can be described at a general level in these very same terms. For if one distinguishes between things in themselves and appearances, then it is clear that the world of

[2] For discussion of Kant's intentions in the Dialectic as a whole, see Karl Ameriks, 'The Critique of Metaphysics: The Structure and Fate of Kant's Dialectic', in P. Guyer (ed.), *The Cambridge Companion to Kant and Modern Philosophy* (Cambridge and New York: Cambridge University Press, 2006), 269–302.

appearances, which essentially depends on us for its determination, does *not* have the *determinate* features that are at issue in considering what the world *as a totality* must be like. As a result, the concept of the world of appearances and the concept of the world as a totality that is presupposed by the attributions asserted and argued for in the Thesis and Antithesis are different concepts, which undermines the claims of the Thesis and Antithesis as applied to appearances. It is equally clear that the world of things in themselves is fully determinate and is what the concept of the world as a totality is supposed to refer to, but given that we have no intuitions of things in themselves, we lack sufficient knowledge to successfully mount the Thesis and Antithesis arguments. As a result of disambiguating things in themselves and appearances and the concepts of the world that they can enter into, the arguments for the Thesis and Antithesis fail and the contradictions they purport to generate no longer arise, even if there is still an unavoidable *illusion* of conflict.

With the First Antinomy, for example, both the Thesis and the Antithesis end up being false if one takes the world in question to be the world of appearances. For given that it is always possible to have additional intuitions of earlier moments in time and farther reaches of space (though with no guarantee that objects will thereby be given therein), the world of appearances is necessarily *indeterminate* with respect to its magnitude, i.e. is neither finite nor infinite. More specifically, the magnitude of the world is indeterminate because appearances are essentially dependent on the subjective forms of intuition through which the world is given to us. If the world is taken, by contrast, to consist of things in themselves, which have their properties independently of us and our intuitions, then the world must be fully *determinate* and thus also have, when taken in its totality, a magnitude – whether finite or infinite we may not be able to know, given the constraints that Kant attributes to our theoretical knowledge of things in themselves.[3]

What is not yet clear at this level of description, however, is (a) how the world as a totality comes to be such an important concept and (b) what this line of argument has to do with pure reason such that it would be appropriate to call it an antinomy *of pure reason*. The answers here arise from Kant's view that (a) reason generates the idea of the world as a totality and (b) it is for that reason that the antinomies that arise for the idea of the world

[3] However, in the transcripts from his metaphysics lectures (28:850), Kant maintains that we can know (as an analytic truth) that the world of things in themselves consists of simple substances.

as a totality are an antinomy of pure reason. But how does reason generate the idea of the world as a totality?

Kant's essential idea is that reason should be characterized as a faculty that searches for conditions for whatever is conditioned that is given to us. Thus, in formal logic reason searches for the conditions for a conditioned judgment, since if successful, it can then formulate a syllogism, with the premises serving as the conditions for the conclusion, which is 'conditioned' by its premises. However, this characterization of reason is not restricted to judgment and the realm of formal logic, as it also applies to objects and scientific inquiry.[4] For any conditioned object that is given to us, reason necessarily searches for the conditions that would explain it. This view is justified by an analytic connection between something being conditioned and its conditions; without its conditions an object would not in fact be conditioned. And reason is the faculty that is interested in, or pursues, this connection.

However, reason goes beyond the analytic connection between the conditioned and its conditions by also being interested in the *un*conditioned. For the unconditioned offers it a resting place, a point at which it can stop its inquiry without being dissatisfied because of conditioned objects whose conditions have not been ascertained; after reason has attained the unconditioned, nothing remains for it to pursue.[5] Kant is quite clear, however, that reason's interest in the unconditioned is distinct from its interest in the conditions for what is conditioned. While there is an analytic relation between the conditioned and its conditions, to posit a condition that is itself *unconditioned* is a synthetic claim for which a further justification must be sought. Indeed, Kant thinks that the fate of metaphysics rests on synthetic a priori claims about certain unconditioned objects, such as God, the immortality of our soul and freedom, and it is because no theoretical vindication of such claims is ultimately possible that metaphysics cannot be a science in the way in which geometry or the pure part of natural

[4] One of the great difficulties that one encounters in trying to make sense of Kant's account of reason is that he uses the notion of a 'condition' in a variety of senses. In logic, he obviously intends a logical conditioning relation. However, in discourse about (e.g. physical) objects, he employs a range of meanings. The most obvious is a causal meaning (in the Second Analogy of Experience), but he also has spatio-temporal conditioning in mind in the First Antinomy and part–whole constitution relations in view in the Second Antinomy. What makes it particularly difficult is that Kant provides no clear guidance regarding these different notions and their relations to each other.

[5] This is not quite true without further qualification. Even having found the unconditioned totality of conditions for a set of conditioned objects, reason could pursue what further conditioned objects *follow* from the original set of conditioned objects. However, Kant interestingly and plausibly notes an asymmetry here, namely that reason is interested only in regressive, not progressive series of conditions.

science is, though Kant does not dismiss metaphysics altogether, but rather attributes a non-constitutive, regulative status to it.

At the same time, Kant maintains a strong connection between reason's search for conditions and its positing something unconditioned. As he explains in 'On the Transcendental Ideas' in the Transcendental Dialectic: 'the transcendental concept of reason is none other than that of the *totality of conditions* for a given conditioned. Now since the *unconditioned* alone makes possible the totality of conditions, and conversely the totality of conditions is always itself unconditioned, a pure concept of reason in general can be explained through the concept of the unconditioned, insofar as it contains a ground of synthesis for what is conditioned' (*CrV* A322/B379). There are a number of claims being made, or at least implicitly assumed, here. The two that are most important for our purposes are as follows. First, reason searches not only for conditions, but also for the *totality* or completeness of conditions. This is not at all a trivial claim. One could imagine a less ambitious faculty that has the need to find only one condition (A) or even one series of conditions (A_1, A_2, A_3) for something conditioned, rather than all of them (A, B and C, or A_1, A_2, A_3, B_1, B_2, B_3, and C_1, C_2, C_3), i.e. without always beginning the search for conditions anew. Second, the totality of conditions must itself be unconditioned. The idea here seems to be that the totality of conditions cannot be a totality if it does not contain *all* of the conditions, but if it contains *all* of the conditions, then it must be unconditioned, because if it were conditioned, it would not have included the condition that conditions it and would thus not in fact have contained the totality of conditions. As a result, if one can find all of the conditions for a conditioned, one will have found the unconditioned as well.

Given this characterization of reason, one can understand how it is that the idea of the world as a totality is generated by reason (rather than simply abstracted from experience) or is, as Kant puts it, 'a transcendental concept of reason'. For the object of cosmology, which is the subject matter of the Antinomy of Pure Reason, is 'the sum total of all appearances (the world)' (*CrV* A334/B391). And one can see how this plays out in the particular Antinomies, since what the Thesis and Antithesis claims are about in each case is not this or that particular empirical object that can be found in the world, but rather some ultimate, fundamental or unconditioned aspect of the world as a totality. Thus, in the case of the First Antinomy it is about the magnitude of the entire world, which obviously concerns a spatio-temporal totality. Now the Second Antinomy, which affirms and denies that the world is ultimately made up of simples, is obviously not about the universe on a large scale, but it does still concern a totality, only on a small scale, since

in resolving an object into its ever smaller constituent parts, it in effect considers what the constitution of the object will be after *all* of its composition has been removed, that is, after it has progressed through the totality of intermediate conditioned constituent parts to the unconditioned (which would be either simple or nothingness, see *CrV* A413/B440).

However, not only does Kant's account of the nature of reason explain the formation of the concept of the world as a totality, but he also uses it in a crucial way in his general resolution of the Antinomies. For reason's search for the conditions for any conditioned object applies to things in themselves and appearances in different ways. As he notes: 'if the conditioned as well as its condition are things in themselves, then when the first is given, not only is the regress to the second *given as a problem*, but the latter is thereby really already *given* along with it; and, because this holds for all members of the series, then the complete series of conditions, and hence the unconditioned, is thereby simultaneously given' (*CrV* A498/B526). That is, for things in themselves it is the case that the totality of conditions and thus the unconditioned as well must exist if the conditioned exists. This contrasts with the situation for appearances. 'On the contrary, if I am dealing with appearances, which as mere representations are not given at all if I do not achieve acquaintance with them … then I cannot say with the same meaning that if the conditioned is given, then all the conditions (as appearances) for it are also given; and hence I can by no means infer the absolute totality of the series of these conditions' (*CrV* A499/B527), nor the unconditioned, given the inference Kant endorses from the totality of conditions to the unconditioned. Even independently of this inference, however, Kant's idea is simply that the mere fact that a conditioned object happens to be given to us does not entail that all of its conditions are given to us as well. For those objects to be given to us, we would have to be affected by them, and there is nothing about the objects that have been given to us that requires that distinct objects be given to us as well.

As a result, the contradictions that are generated in the Antinomies arise because our idea of the world as a totality represents the unconditioned totality of conditions, but what is given to us in intuition (and is thus knowable) is, he claims, always conditioned so that there is a necessary mismatch between appearances and the idea of reason, which only things in themselves are adequate to. If one does not distinguish between things in themselves and appearances, one thus has a necessary mismatch and the Antinomy's contradictions. However, as soon as one draws the distinction, reason can require that its demands be satisfied for things in themselves, though it has no way of knowing how they are, but it cannot require that

this very same demand be satisfied for appearances as a result of their essential lack of complete determinacy; instead, reason can demand only that one continue to search for ever further conditions, even if one knows that the totality of conditions (and thus a determinate answer in regard to the world as a totality) can never be given in experience. The Antinomy of Pure Reason is thus resolved not only by distinguishing between things in themselves and appearances, but also by noting that our idea of the *world as a totality* can apply only to the fully *determinate* world of things in themselves and that the sensible world is unlike it in so far as it is *indeterminate* and does not contain anything unconditioned.[6]

While there are myriad details of the Antinomy of Pure Reason not addressed by this description, it does offer a brief account of the very basic structure of Kant's understanding of a theoretical antinomy. Reason, in its search for an unconditioned totality that underlies all conditioned objects we encounter in experience, forms the idea of the world as a totality that it would find satisfying, and attempts to determine its basic features. In so far, however, as contradictory propositions can be proved regarding these features, reason has come into conflict with itself. It can resolve this conflict only by distinguishing between things in themselves and appearances, recognizing that the world of appearances is essentially indeterminate and therefore does not fit the concept of the world as a totality, and that the world of things in themselves, which must be a totality, lies beyond what we can know theoretically.

II. THE ANTINOMY OF PRACTICAL REASON RECONSTRUCTED

Given this understanding of the Antinomy of Pure Reason, we can turn to the Antinomy of Practical Reason, with the expectation of considerable similarities. However, what is immediately striking about the text of the Antinomy of Practical Reason is that it contains significant and surprising obstacles to any attempt at understanding even the main idea behind its argument. For one, Kant does not label any sentences in this text 'Thesis' or 'Antithesis', nor does he explicitly formulate propositions that obviously constitute a contradiction. For another, Kant's statement of the entire antinomy is unusually brief, occupying a single paragraph, which increases the difficulty of identifying formal proofs or arguments of any sort. Finally,

[6] For further discussion, see my 'Kant's Antinomies: Sections 3–8', in G. Mohr and M. Willaschek (eds.), *Kooperativer Kommentar zu Kants Kritik der reinen Vernunft* (Berlin: Akademie Verlag, 1998), 445–62.

in addition to the regular store of puzzling sentences, there are numerous sentences that seem completely out of place.[7] This situation has led the most respected of commentators brave enough to tackle this challenge to depart from Kant's claim that there is a single antinomy of practical reason. Thus Lewis White Beck claims that there is no antinomy of practical reason 'in any strict sense', while Allen Wood has argued that there are actually two.[8] The consensus is that Kant's text is sloppy and that one must react accordingly. However, I suggest that if one keeps in mind the structure of the theoretical antinomy from the first *Critique* one can identify without undue strain at least the basic structure of the antinomy of practical reason in the appropriate passages.[9]

The basic idea of the Antinomy of Practical Reason is that given the assumption that the world of appearances is identical to the world of things in themselves, practical reason comes into conflict with itself when it attempts to establish the possibility of the highest good in the face of our experience of the world. This argument can be reconstructed in the following steps:

(1) Assume Transcendental Realism.
(2) If morality is not false, then the highest good must be possible as an object of pure practical reason.
(3) The highest good can be possible as an object of pure practical reason only if there is a (synthetic) necessary connection between virtue and happiness.
(4) There can be a (synthetic) necessary connection between virtue and happiness only if either happiness necessarily causes virtue or virtue is necessarily the cause of happiness.
(5) If (1), happiness does not necessarily cause virtue.
(6) If (1), virtue does not necessarily cause happiness.
(7) If (1), then (a) the highest good is not possible as an object of pure practical reason and (b) morality is thus false.
(8) Morality is not false.
(9) Therefore one must reject (1).

[7] For example, when Kant talks about the 'desire for happiness' being 'the motive' for virtue (*CpV* 5:113), it seems as if this is the wrong subject. For the question at hand concerns, not desires and motives, but rather the object of practical reason and either happiness or virtue.

[8] Beck, *Commentary*, 247, and Wood, *Kant's Moral Religion*, 104–5. Beck nevertheless goes on to formulate three different possible reconstructions of the antinomy.

[9] For exhaustive discussions of earlier interpretations of various aspects of the antinomy of practical reason, see Albrecht, *Kants Antinomie*. After noting the importance of reason to the concept of the highest good and providing some historical context to Kant's position, Albrecht brackets further discussion of how reason might generate the specific content of the concept of the highest good (55).

This reconstruction of the antinomy of practical reason is formally valid (even if (7) is technically redundant). Perhaps surprisingly, however, the textual evidence in its favour is, I contend, relatively clear. Except for (1) and (2), which are background assumptions made explicit earlier in the second *Critique*, each of the premises of the argument can be seen as an expression of the eight sentences that make up this section of the text, and, what's more, in exactly that order. (3) is stated in the first sentence of the Antinomy of Practical Reason: 'In the highest good which is practical for us, that is, to be made real through our will, virtue and happiness are thought as necessarily combined' (*CpV* 5:113). (4) is supported by the line of argument expressed in the next three sentences, where Kant asserts (in the second sentence) that if the combination of virtue and happiness is necessary, then it must be either analytic or synthetic (in the third sentence), that since they cannot be related analytically, they must be related synthetically, specifically, as cause and effect, and (in the fourth sentence) that the relevant cause–effect relations are that virtue causes happiness or that happiness causes virtue. (5) is supported by the fifth sentence ('The first is absolutely impossible because …'), while (6) is supported by the first part of the sixth ('But the second is also impossible because …'). (7a) is clearly stated in the second part of that same sentence ('consequently, no necessary connection of happiness with virtue in the world, adequate to the highest good, can be expected …'), while (7b) is explicitly inferred in the seventh sentence ('since the promotion of the highest good … is an a priori necessary object of our will and inseparably bound up with the moral law, the impossibility of the first must also prove the falsity of the second' (*CpV* 5:114)) and then immediately repeated in the last sentence. (8) is not explicitly stated here, but is an obvious assumption in the larger context.

If one reconstructs the Antinomy of Practical Reason in this way, the basic line of its resolution turns out to be surprisingly straightforward as well. If one rejects Transcendental Realism in (1) and distinguishes between things in themselves and appearances, then one can claim that while it is 'absolutely false' (*CpV* 5:114) that happiness will necessarily cause virtue, it is possible that virtue might cause happiness. For even if virtue does not directly cause happiness in the sensible world, it is possible that virtue could cause happiness indirectly, by way of some connection in the world of things in themselves. In the Postulates of Practical Reason Kant goes on to argue that God's existence and the immortality of the soul must be presupposed for the highest good to be possible in this way, but the first minimal step necessary for the resolution occurs in so far as one notes that the relation between virtue and happiness may be different from what

the argument assumes on account of the possible influence of things in themselves.

Moreover, this reconstruction is confirmed by several basic parallels with the Antinomy of Pure Reason, even if there are, and in fact must be, significant differences, given that the one antinomy is theoretical and concerns the world as a totality, whereas the other is practical and concerns the highest good. First, the argumentative structure of the practical antinomy (including its set-up and resolution) is analogous to that of the theoretical antinomy. For the antinomy proceeds by assuming Transcendental Realism (in (1)), which leads to antinomial conflict (by the highest good being both possible and not possible), which is then resolved by rejecting Transcendental Realism (in (9)) and explaining how the distinction between appearances and things in themselves allows one to avoid the argument that delivers the problematic conclusions.[10]

Second, the practical antinomy requires not just the initial assumption of Transcendental Realism, but also an idea in the form of the idea of the highest good, just as the idea of the world as a totality was required for the theoretical antinomy. For it is in the attempt at understanding specifically the possibility of the highest good that the antinomy arises. That is, simply assuming Transcendental Realism in the moral realm is not obviously immediately problematic on its own. It is the highest good that forces the conflict. What's more, the idea of the highest good is generated by reason as an unconditioned totality, again, just as the idea of the world as a totality is. Kant is perfectly explicit about this point in Chapter I of the Dialectic of Pure Practical Reason: reason 'seeks the unconditioned for the practically conditioned … not indeed as the determining ground of the will, but even when this is given (in the moral law), it seeks the unconditioned totality of the object of pure practical reason, under the name of the *highest good*' (*CpV* 5:108).

Against this background, we can now appreciate Kant's explanations of several features of the concept of the highest good at the start of Chapter II of the Dialectic.[11] First, he begins by noting an ambiguity in the concept of the highest in general. It can mean 'either the supreme [*das Oberste*] (*supremum*) or the complete [*das Vollendete*] (*consummatum*). The former is

[10] This interpretation allows one to avoid the problem that (5) and (6), which one might naturally take to be the Thesis and Antithesis, do not in fact contradict each other. I thank Andy Reath for helping me to think more clearly about how to reconcile my reconstruction with the text on this point.

[11] For discussion of Kant's notion of the highest good as treated outside the *CpV*, see Pauline Kleingeld, 'What Do the Virtuous Hope For?: Re-reading Kant's Doctrine of the Highest Good', in H. Robinson (ed.), *Proceedings of the Eighth International Kant Congress*, vol. I.i (Milwaukee: Marquette University Press, 1995), 91–112.

that condition that is itself unconditioned, i.e., is subordinate to no other (*originarium*); the second is that whole that is not a part of a still greater whole of the same kind (*perfectissimum*)' (*CpV* 5:110). He then applies it to the highest good and suggests that the Analytic has established virtue as the supreme condition of everything that could appear desirable to us, while the addition of happiness to virtue constitutes the complete good for rational finite beings.

Second, Kant spends several paragraphs explaining the relation between virtue and happiness within the highest good. As a general point, Kant notes that two 'determinations' (or concepts) necessarily combined under a single concept must be connected as ground and consequent and such that their unity is either analytic (based on a logical connection) and according to the law of identity (i.e. principle of contradiction), or synthetic (based on a real combination) and according to the law of causality (*CpV* 5:111). He then argues that with respect to the highest good, the concepts of happiness and virtue cannot be identical, as Epicureans and Stoics had maintained in attempting to reduce the one to the other (albeit in different directions). Rather, the concepts of happiness and virtue are extremely heterogeneous and thus too different from each other for any such reduction to be possible. Instead, one must synthesize them and the only way to do so, he thinks, is by way of a causal relation. Moreover, because the idea of the highest good must be a priori, this synthesis of the concepts of happiness and virtue within it must be a priori as well. For the conditions of the possibility of the highest good cannot depend on empirical principles; rather 'the deduction of this concept [of the highest good] must be transcendental' (*CpV* 5:113).

While we have not explored in any detail the consequences that follow from the Antinomy of Practical Reason, we are now in possession of a formally valid reconstruction of the practical antinomy itself, a reconstruction that is squarely based on the text and on a structural similarity with the theoretical antinomy. This should give us confidence, despite the difficulties inherent in the actual text of the Antinomy of Practical Reason, that the basic structure of the practical antinomy has been captured accurately. We have also clarified the content and basic structure of the notion of the highest good, which plays a central role in the practical antinomy.

III. THE HIGHEST GOOD AND THE NATURE OF REASON

If the structure of the Antinomy of Practical Reason has been reconstructed and if both what leads to it and the fundamental terms it employs have been

clarified, we can now turn to the task of subjecting it to more detailed critical scrutiny. While much attention has been devoted to the question of whether the Postulates of Practical Reason are in fact necessary for the possibility and promotion of the highest good, it is possible to focus on earlier steps in the argument, with respect to which Kant's argument can appear vulnerable from the perspective of common sense.[12] Why must the highest good be an object of pure practical reason at all (in the consequent of (2))? Couldn't one simply follow the moral law and not have any greater ambitions about the extent of the good that must be thought as attainable thereby? Moreover, why must the highest good include virtue *and* happiness (in the consequent of (3))? And even if one concedes that virtue and happiness are not reducible to each other, why must they be related as ground and consequent as opposed to some other way, for example as competing goods?

These queries can be stated systematically in the form of the following questions. (A) Must there be an object of pure practical reason at all? (B) Must there be a *single* object of pure practical reason? (C) Must the object of pure practical reason be the *highest good*? (D) Could the highest good not be simply the supreme good rather than the complete good? (E) Why must virtue and happiness be related by means of a one-way causal relation in the complete good? These questions will be discussed in turn.

(**A**) The first question is either surprisingly easy or surprisingly difficult to answer. It is quite easy if one simply accepts as a basic fact that action is essentially teleological.[13] For in that case, every action must have an object that it attempts to bring about in the world, and the will is a (special) kind of causal power, whose exercise produces this object as its intended effect.[14] Kant acknowledges this feature of action by distinguishing between the motivation (*Bestimmungsgrund*) for an action and its object or end.[15]

[12] For discussion of the promotion (as opposed to the possibility) of the highest good, see Andrews Reath, 'Two Conceptions of the Highest Good in Kant', *Journal of the History of Philosophy* 26 (1988), 593–619.

[13] Kant expresses a commitment to a teleological conception of action at, e.g., *CpV* 5:9; *MdS* 6:384–5; *G* 4:427–8. However, matters are more complicated if one distinguishes between whether we act teleologically and whether pure practical reason acts teleologically. Even if one grants the former, the latter does not immediately follow.

[14] Kant addresses this kind of possibility in the first paragraph of Chapter II, 'On the Concept of an Object of Pure Practical Reason' (*CpV* 5:57–8).

[15] Kant seems to rest this common-sense distinction on a further distinction between form and matter, since at times he wants to characterize the moral law as in some sense formal (e.g. universalizable) and to distinguish it from all 'matter' or 'material principles'. As he notes, since the moral law is 'merely formal', it abstracts 'from all matter and thus from every object of volition' (*CpV* 5:109), which suggests that any object of volition would be included in the matter of volition. However, Kant's form–matter distinction is widely employed, but never actually defined, much less argued for. It is thus not clear that this feature of Kant's system can be used to any clear advantage here.

However, one might imagine a position according to which an agent simply follows a certain set of rules without regard to the result – perhaps an uncharitable caricature of an extreme deontologist position – and then it is less obvious how to respond. On such a view, one could still maintain that the will would be a kind of causal power to bring about an effect, but it would not be required that the effect be represented as its intention or even as in any sense good (or evil). Granted, that is simply not how we understand action for finite rational agents, but to respond in this way can sound dogmatic.[16] At this point, the best answer to this first question seems ultimately to require a retreat to the level of descriptive metaphysics. Just as our experience happens to be spatio-temporal and we take its spatio-temporality to be definitive of our theoretical experience, so too the fact that all of our actions have an object is, though not logically necessary, a defining feature of our practical standpoint.[17]

This question is also difficult to answer if one takes it to be asking not whether our actions have objects as ends, but rather if the objects that we intend to bring about through the exercise of our will are objects of practical *reason* which, given Kant's characterization of reason, must be either a conditioned good or an unconditioned good.[18] One might think that the pair is exhaustive of the options – any good object is either conditioned or not and if not conditioned, then unconditioned – and therefore that the object or objects of practical reason must be one or the other as well, but one could call the applicability of the concept pair into question.[19]

A similar question can arise in theoretical philosophy. Are appearances conditioned or unconditioned? In this case Kant provides a direct and principled (even if limited) answer in the Second and Third Analogies of Experience. For in so far as, for example, the Second Analogy requires a cause (as a condition) to explain a change of determinate states in the effect (as what is conditioned), one can infer that all changes of state are

[16] In the Analytic of Practical Reason one might see Kant as employing this strategy in so far as he is simply analysing a certain concept of the object of practical reason, without arguing that this concept has applicability.

[17] In a footnote in *Religion within the Boundaries of Mere Reason*, Kant remarks: 'it is one of the inescapable limitations of human beings and of their practical faculty of reason (perhaps of that faculty in all other worldly beings as well) to be concerned in every action with its result, seeking something in it that might serve them as an end' (*R* 6:7).

[18] For reasons of space, I abstract here from the complexities involved in sorting out the different senses in which a good object might be conditioned or unconditioned.

[19] Hume calls the applicability of the substance-accident concept pair into question, e.g., by claiming that there is nothing that could possibly fall under the concept of an accident, making the attribution of substantiality effectively meaningless.

conditioned.[20] Now Kant goes further and claims that *every* appearance is conditioned, and the Second and Third Analogies may not be strong enough to establish this further claim. However, they at least provide an example of one possible answer to the question about the applicability of these concepts to a certain class of appearances in theoretical philosophy.

Returning now to the practical domain, Kant's position on this point earlier in the second *Critique* seems not to be wholly satisfactory, since he infers without any explicit argument that anything that our *pure* practical reason represents as good must also be 'good absolutely, good in every respect, and the supreme condition of everything good' (*CpV* 5:62), and that any object of volition that is based on a feeling of pleasure or displeasure is 'not a good, but a well-being' (*CpV* 5:62). It is true that one might interpret the first few pages of Section I of the *Groundwork* as arguing that the objects we typically desire will be conditioned or unconditioned, depending on their circumstances. However, even in this case, Kant does not so much give a principled argument as rely on our common-sense intuitions. This is not to say that Kant's position is in any way indefensible, but rather only to remark that he provides little explicit argument for his position on these points.

(**B**) If there must be, at least for us, some object that we try to bring about whenever we act, why must there be a single object? For starters, Kant is not denying that, e.g., the object that I intend to bring about in helping someone in need (o_1) is different from the object I pursue in telling the truth (o_2). Instead, he is asserting that these objects either have something in common or can be unified into a single coherent object of volition (an O that somehow includes o_1, o_2 … o_n), where the task of finding the common concept or the unifying principle is one that arises at the level of philosophical reflection. This single object is also not entirely distinct from the particular objects of volition one might have on a day-to-day basis (say, o_3) such that it might either compete with them or not be realized even if they all were. Rather, it is a question of the coherence or systematic unity of a plurality of finite objects that we have as an end.[21] Must there be a single object that unifies or includes all of the particular objects that one might have as ends?[22] Again, one can imagine someone who accepts Kant's

[20] See chapters 3 and 4 of my *Kant and the Metaphysics of Causality* (Cambridge and New York: Cambridge University Press, 2005), for discussion of Kant's views on causality.

[21] In addition to the distinction between the various notions of conditioning pointed out in note 3 above, Kant does not clearly distinguish between the kinds of conditioning relations among finite objects and the whole that might unite them and the kinds of conditioning relations among the ideas that constitute the highest good.

[22] While the question may have a theoretical dimension (in so far as the object is one that we represent to ourselves), it must also have a practical side, i.e. must be an object of *volition*.

moral theory and a teleological conception of action too, but who still denies that there is any overarching object that incorporates all of the particular objects of volition we might have.

There are two obvious strategies for answering this question (or this kind of 'sceptic'). One might simply define a concept that necessarily encompasses the different particular objects of volition there are, which then guarantees that there is only one such 'object'. One would be pursuing this strategy, e.g. if one took a teleological conception of action as primitive and then defined the single overarching object of volition as 'the good' (or perhaps as the 'material' of our volition). While this kind of strategy may be able to deliver the proper result, it does not capture Kant's actual intentions and procedure, since no real integration of particular objects into a coherent conception is required and there is no feature that objects would have to have such that they fall under a single well-defined concept. This strategy would thus seem to come up empty.

One could instead try to exploit the fact that Kant calls the object in question an object of pure practical *reason*. The line of argument might be developed as follows. As we have seen above, Kant holds that reason searches not only for conditions, but also for the totality of conditions and, he infers, the unconditioned as well. However, if reason has a representation of the *totality* of conditions that guides it in incorporating the conditioned into a complex of condition–conditioned relations with the unconditioned, so to speak, at the top, then the representation of the unconditioned would have to be of a singular object. For if there were a plurality, one would not have captured the *totality* of conditions or the unconditioned that is supposed to unite *all* conditioned objects underneath it.[23]

However, this line of argument is not ultimately fully convincing. One way of seeing the weakness in the argument is to note that there is no guarantee in advance (or at the level of reason's basic structure) that there is only *one* unconditioned. In principle it is possible we could pursue a plurality of conditioned objects back to two independent unconditioned objects. In fact, this is the case in the theoretical sphere, with God, the world as a totality, the soul, and so forth all as obviously different unconditioned objects. As a result, if Kant is to have an argument for the claim that there is only one overarching object of pure *practical* reason, it will have to be on the

[23] Allen Wood seems tempted by this kind of explanation. 'Any idea of an unconditioned for Kant is an idea formed by reason in its office of securing a unity under principles' (*Kant's Moral Religion*, 91). Specifically, he argues that the unconditioned, and the totality of conditions along with it, delivers a systematic unity. Wood is not explicit about whether it is the idea of the unconditioned or that of the totality of conditions that requires the singularity of the object.

basis of the particular objects of volition that come under consideration. To put the point in Kant's own terms, the object of pure practical reason necessarily has a particular form, determined by the moral law, but its matter has not yet been specified such that this issue can be settled.[24] Again, the point here is not to show that Kant's position is indefensible, but only that his best argument, based solely on practical reason's search for the unconditioned, is not sufficient to answer this question definitively. (We shall have reason to return to this issue below in (E).)

(C) In light of this, the important question at present is not so much whether the highest good is the object of pure practical reason, but rather on what basis Kant maintains this point. The standard answer is to claim that it is based on the moral law.[25] The basic motivation for this position is that Kant argues at length in the Analytic of Practical Reason that the moral law must be the determining ground of the will and that any attempt to supplant it by invoking some other object would amount to a material principle that would be heteronomous and therefore ought to be rejected. However, if the moral law must determine the will, then it might seem that the moral law must also determine the object of the will. For if it did not determine the object of the will, then something else must be determining it, in which case it could seem as if something else would be supplanting the authority of the moral law. This line of argument seems to have solid textual evidence in its favour as well: 'the moral law must be viewed as the ground for making the highest good and its realization or promotion the object … For … if one assumes any object under the name of the good as a determining ground of the will prior to the moral law and then derives from it the supreme practical principle, this would always produce heteronomy and supplant the moral principle' (*CpV* 5:109).

However, it is not clear that the moral law can determine the object of the will in the sense in question.[26] After all, as Kant emphasizes, the moral law is

[24] Another way to see this point is to note that this argument does not establish that the object of pure practical reason must be the unconditioned at the top. For all that Kant has said about the object of volition, it could well be conditioned. This would certainly be an odd state of affairs, but if the argument for a singular object of volition is supposed to be a conceptual one, it would have to be in a position to exclude this possibility, but there is no conceptual requirement that the object of volition be unconditioned.

[25] See Allen Wood (*Kant's Moral Religion*, chapters 2–3); Christine Korsgaard (*Creating the Kingdom of Ends* (Cambridge and New York: Cambridge University Press, 1996), 106–32); and, especially, Stephen Engstrom ('The Concept of the Highest Good in Kant's Moral Theory', *Philosophy and Phenomenological Research* 52 (1992), 747–80) for different versions of this line of argument.

[26] Even if the moral law is a criterion for determining which actions we should or should not undertake and thus separates actions into good and evil, that does not mean that it is also a criterion for the highest good as an object of pure practical reason that serves as a unifying ideal for the various actions we undertake.

in some sense an exclusively formal principle, whereas the concept of the highest good that is the 'object' of pure practical reason would naturally involve some 'material' aspects that the moral law, as such, cannot account for. Or, to put it more precisely, even if the moral law does determine the object of practical reason in some sense, it may not determine it as the highest good. Moreover, the textual evidence cited above is not unambiguous. Rather than taking it to establish that the moral law must determine the object of the will, one might read it as making a very different point, namely the moral law determines what *motivates* the will, not its object, because the moral law is a purely formal principle. In fact, the sentences just prior to the one quoted above support this point: 'The moral law is the sole determining ground of the pure will. However, since this is merely formal (that is to say, it requires only that the form of a maxim be universally lawgiving), it abstracts as determining ground from all matter and so from every object of volition' (*CpV* 5:109). So the point of the entire paragraph is not to claim that the moral law must determine the object of the will, but rather to insist that the moral law rather than the highest good must determine the motivation for the will, and that one should not think that introducing the highest good as the object of the will changes this point.

If the moral law does not determine the object of the pure will, what does? Two points are suggestive. First, the object of the will here is also the object of pure practical reason, so one would expect reason to be involved in some way (even if not in the form of the moral law). Second, Kant defines the highest good, which is the object in question, in terms of the unconditioned. Specifically, 'as pure practical reason it similarly searches for the unconditioned for the practically conditioned (which relies on inclinations and natural needs), and not as the determining ground of the will, but rather, if this has also been given (in the moral law), as the unconditioned totality of the object of pure practical reason, under the name of the *highest good*' (*CpV* 5:108). Taken together, what these two points suggest is that practical reason itself determines that the highest good is its object, and it does so by searching for the totality of conditions, which must therefore be unconditioned, for what is given to it as conditioned.[27] The result of its search, the unconditioned, is then defined in terms of the highest good. So even if reason cannot determine that there is only one such object, it can determine that whatever the object is, it will be the highest good.

[27] While Wood sometimes seems to endorse the previous explanation, he also makes repeated remarks that are strongly suggestive of this interpretation.

Now in so far as the moral law is defined or determined by practical reason, one might think that these two interpretations are not ultimately distinct. However, it is important to distinguish between reason in its formal employment in the determination of the will's motive and reason in its 'material' employment in the determination of the will's object. Contemporary ethicists often find Kant's moral philosophy enticing when it is viewed as attempting to deliver substantive (i.e. non-empty) normative principles but without assuming anything more than a formal notion of rationality (e.g. such that universality in scope is required to count something as a reason). Yet it is important to be clear that Kant's own conception of reason is, and must be, more robust than contemporary ethicists realize, and this is one context in which the robust conception is required. For positing the *totality* of conditions and thus the *unconditioned* goes well beyond simply requiring mere formal consistency (or even universality in scope).

As we saw above, in Kant's theoretical philosophy there is a difference between the condition–conditioned concept pair and the conditioned–unconditioned concept pair, a difference that is reflected in Kant's claim that there is an analytic connection between the conditioned and its conditions, but a synthetic connection between the conditioned and the unconditioned. Indeed, the Antinomies arise precisely because one is unsuccessful at finding a justification for synthetic claims about the unconditioned (in the form of certain determinate features of the world as a totality), with the consequence that the unconditioned is not an object of knowledge for us. Does the same consequence arise here, namely that the highest good is not an object of the will, despite initial appearances? To put the question the other way around: what justifies positing the unconditioned, i.e. the highest good, given practically conditioned objects (given that reason lacks any objective justification for such a move on the theoretical side)? Why must practical reason succeed in attaining the unconditioned, where theoretical reason fails?

The difference here lies not so much in the priority of practical reason over theoretical reason, but rather in the differences between theoretical and practical reason.[28] One major difference is that the theoretical requirement that conditioned objects be *given* to us in intuition is different from the practical requirement that our will *bring about* the unconditioned. Whether objects are given to us is a limitation of our epistemic capacities and it is,

[28] Korsgaard (*Creating the Kingdom of Ends*, 119) emphasizes the priority of practical reason in her explanation of this point.

Kant thinks, simply a fundamental limitation that the unconditioned cannot be given to us and we can therefore not know that it exists. Practical reason has no comparable limitation, since it can, indeed must, strive to bring about the highest good to the extent that it is possible for a single agent.[29] Granted, our causal powers are limited in being able to bring about the highest good, but that limitation does not attach to our rational will as such and thus does not affect our conception of the highest good, which makes it different from the theoretical case. However, despite the difference in the cognitive status of the unconditioned, it does function as an ideal, or regulative principle, in both cases, that is, it serves as a guide for our action, whether it be in theoretical inquiry or practical matters.[30] So even with the differences, there is a fundamental similarity.[31]

(**D**) If reason thus generates the idea of the highest good, a question still arises about the exact content of this idea, corresponding to the ambiguity that Kant points out between the notions of the supreme and the complete highest good. Given these two concepts of the highest good, which is the proper one in the context of the second *Critique*? That reason generates the idea of the highest good cannot be used to decide this issue, since reason is perfectly able to generate both of these more specific concepts. We must thus look elsewhere for an adequate response.

One might quite reasonably answer this question by appealing to the goals that Kant intends for the Antinomy of Practical Reason to achieve. If Kant's fundamental aims in the second *Critique* include establishing rational belief in God's existence and the immortality of the soul, then he should feel free to appeal to whatever means are necessary to demonstrate the postulates. The Antinomy of Practical Reason is relevant in this context because Kant argues that only the notion of the complete highest good provides him with the means to establish both of these beliefs as rational. In particular, it is the inclusion of the notion of happiness, which is contained only in the notion of the complete highest good, that is indispensable for rational belief in God's existence, given that only God, it is argued, is in a position to proportion happiness to virtue.

While this answer is perfectly acceptable, there are, I suggest, two further reasons for Kant to focus on the complete rather than the supreme highest

[29] Practical reason has a limitation, namely that it depends on desires that derive from sensibility, but that is a different kind of limitation, since it does not restrict what can serve as the object of our will.

[30] The highest good is something more than a merely regulative principle in so far as Kant does want to establish its real possibility, a status that is not necessary for merely regulative principles.

[31] For discussion of the regulative status of the highest good, see John Silber, 'Kant's Conception of the Highest Good as Immanent and Transcendent', *Philosophical Review* 68 (1959), 469–92.

good. First, in so far as the highest good is supposed to result from starting not with virtue (which is unconditioned), but rather with what are for us conditioned goods and tracing the *totality* of their conditions back to the unconditioned, one would not in fact have captured all conditions if one did not include those relations that obtain among the various objects that contribute to our happiness.[32] That is, to be assured that the totality of conditions is actually represented in the highest good, the highest good would need to include happiness as well, since it too includes some condition–conditioned relations.

Second, by including the idea of happiness and thus having to focus on the relation between happiness and virtue, Kant has the opportunity to distinguish more clearly between two ideas of happiness that we have and to be explicit about which one belongs in the concept of the highest good. Reason, which is the source of all of our ideas, must generate our idea of happiness as well, but it can do so in two ways. With respect to the concept of happiness that is perhaps more familiar to us, reason simply takes the *totality* of our desires and is charged with determining what largest set of desires could be consistently satisfied. Since my specific desires at any moment in time sometimes conflict, prudential reason must determine which ones should not be included in the set to be acted on, such that the greatest possible total satisfaction of desires (including not just the number, but also the weight of desires) can occur. Kant clearly has this notion in mind when he criticizes our idea of happiness as being too indeterminate to be genuinely action-guiding (given, e.g., the difficulty of foreseeing what the long-term consequences of one's actions will be on desire satisfaction).

In the complete highest good, however, reason forms a different idea of happiness, one that is conditioned by virtue. That is, our happiness, the totality of desires, is limited not only by formal consistency, but also by the further, potentially very significant, condition that the action leading to the satisfaction of the desire be one that a virtuous person could perform. As a result, this concept of happiness, which involves more than just prudential reasoning, could well have a very different structure and systematic ordering relation among the desires it includes. The systematic principle it embodies, however, can be made explicit only by considering how virtue conditions it, so one can form a proper (moral) conception of happiness – and thus perhaps be able to form an idea that is determinate enough to be properly

[32] This does not imply that the highest good is the unconditioned. Only virtue (or the good will) is unconditioned.

action-guiding – only by focusing on the complete highest good. In so far as it is important to Kant that these different ideas of happiness and their divergent structures be laid out clearly in his practical philosophy, the Antinomy of Practical Reason provides an ideal systematic opportunity to do so, but only if he takes advantage of it by discussing the complete highest good.[33]

(**E**) Finally, why must virtue and happiness be related by means of a one-way causal relation in the complete highest good? Even if one accepts that virtue and happiness are not identical, one can, it seems, imagine a number of different kinds of still properly synthetic relations between them. For example, one might think that virtue and happiness compete with each other or stand in some kind of real opposition to each other. Or, to give an example that explicitly utilizes a Kantian concept, one might suggest that virtue and happiness stand in mutual interaction with each other (such that, e.g., an increase of the one will bring about an increase in the other and vice versa). How can Kant rule out these alternatives (especially if they invoke a concept that he explicitly acknowledges to be synthetic)?

First, Kant takes it that he has already established three important results. (1) Virtue is an unconditioned good. (2) Happiness is not an unconditioned good. (3) Happiness is conditioned by virtue. As a result, whatever the highest good ultimately determines with respect to the relation between happiness and virtue, it must accommodate these three claims. What is significant here is that the first claim immediately rules out the possibility that virtue and happiness might stand in mutual interaction (i.e. reciprocal causal relations), since virtue cannot be conditioned by happiness, given that it is not conditioned by anything. (It is at this level of specificity that the question discussed above concerning whether there must be only one object of pure practical reason can be answered. What claims (2) and (3) make clear is that among the potential objects of practical reason, there cannot be two unconditioned objects.)

Second, if the concept of the highest good is formed by reason and reason is interested in the relations between the conditioned and its conditions, then it is natural that the concept of the highest good must involve such relations and no others. In so far as the alternative accounts of the relation of virtue and happiness employ concepts that cannot be

[33] Pat Kain has suggested (without endorsing) another reason for thinking that the focus should be on the complete rather than the supreme highest good, namely that an impartial spectator would not be pleased were a virtuous person not to be rewarded with happiness. Kant does suggest (at *CpV* 5:110) that virtue makes one worthy of happiness, which one might take to entail that virtue should be rewarded with happiness.

reduced to conditioned–condition relations, then Kant can simply note that these relations are possible in reality, but are not relevant to our conceptual grasp of the rational relations that obtain between virtue and happiness. In other words, in so far as we take seriously that the idea of the highest good is generated by reason, then it follows naturally and immediately that the relations between the elements of the highest good, namely happiness and virtue, will be rational relations and thus captured in conditioned–condition relations. Other relations may obtain between virtue and happiness, but there is no reason to include them in the concept of the highest good.

In the end, however, even if we can determine in this way that virtue and happiness must stand in a synthetic condition–conditioned relation, there is still another step to the claim that this relation must be causal. For there are clearly synthetic conditioning relations other than causality. Kant's position must ultimately be that only a causal relation is appropriate to the specific rational relation that obtains between the goodness of happiness and virtue, though he provides no explicit justification for this point.

IV. CONCLUSION

What has become clear from our consideration of the notion of the highest good is how crucial Kant's conception of reason is to determining a series of specific features of the content of our idea of the highest good. It is because reason searches for the totality of conditions and thus the unconditioned that Kant is interested in the complete highest good and can understand the relation between virtue and happiness as a relation of unconditioned condition to conditioned, which has a fundamental effect on how a properly moral concept of happiness is to be understood. This is not to say that reason alone determines every aspect of the highest good. For Kant also relies on claims about the status of virtue and happiness, specifically about whether they are conditioned or unconditioned, in constructing the concept of the highest good. At the same time, the fact that these features of virtue and happiness are relevant is due to Kant's distinctive conception of reason. So it is ultimately fundamental to his entire conception of the highest good, even if it does require further claims established elsewhere.

Within the literature on Kant's moral philosophy overall, considerably more attention has been devoted to a much less robust conception of reason. The reason for this attention typically derives from the contemporary project of attempting to derive truly substantive normative moral principles from a purely formal conception of reason that non-Kantians

would be committed to as well. One significant difficulty that those who pursue this project often seem to encounter, however, is the following dilemma: either it is difficult to derive substantive principles from such a meagre starting point in a way that is both clear and valid or one can derive substantive principles in the appropriate way, but only by appealing to a richer concept of reason.[34] Understanding more clearly what Kant's actual conception of reason is in the Antinomy of Practical Reason puts one in a position to consider whether this very same conception of reason might be what Kant is drawing on in his attempts at both describing and deriving the moral law.[35]

[34] For those who are staunchly committed to a purely formal (e.g. instrumentalist) conception of reason, what is now clear is that one can dismiss the Antinomy of Practical Reason and the entire Dialectic of Pure Practical Reason along with it, its interest in the immortality of the soul and the existence of God included, because it is based in fundamental ways on a much more robust conception of reason.

[35] I thank Stefano Bacin, Steve Engstrom, Pierre Keller, Heiner Klemme, Pauline Kleingeld, Andy Reath, Jens Timmermann, Marcus Willaschek as well as Dana Nelkin and Clinton Tolley for helpful comments on earlier versions of this paper. All remaining errors are my own.

CHAPTER 8

The primacy of practical reason and the idea of a practical postulate

Marcus Willaschek

1. INTRODUCTION: RATIONAL BELIEF WITHOUT THEORETICAL EVIDENCE

Peter is secretly in love with his colleague Mary but too shy to disclose his feelings to her. Mary is nice enough to Peter, but never has she given any indication that her feelings for him go beyond those for a colleague and friend. Nevertheless, Peter firmly believes that Mary is in love with him, too. Most people would agree with Kant's contemporary Wizenmann (on whose objection to Kant the example is based) that Peter's belief is irrational. But why is it? A possible answer might be that the fact that one wishes something to be the case in itself is no evidence for its being the case. This is the answer given by Wizenmann: 'reasons [for a belief] that do not bear at all on the credibility of a thing do not determine anything with respect to its very credibility'; therefore, according to Wizenmann, a belief like Peter's is not 'rational' (*vernünftig*).[1] In order for this answer to work, however, we must presuppose that for a belief to be rational, the subject needs sufficient evidence in favour of its truth ('reasons' that bear on the 'credibility' of the belief). In Peter's case, no such evidence is available; hence, his belief is irrational.

Immanuel Kant would have agreed that Peter's belief is irrational.[2] But in several places in his work, most prominently in the Dialectic of his *Critique of Practical Reason*, Kant has denied the unrestricted validity of the principle that rational belief requires evidence in favour of its truth. Rather, we can be rationally warranted in a belief even in the complete

[1] Thomas Wizenmann, 'An den Herrn Professor Kant von dem Verfasser der Resultate der Jacobischen und Mendelssohnschen Philosophie', *Deutsches Museum* [Leipzig] 1 (1787), 116–56.

[2] See his reply to Wizenmann in the footnote close to the end of the Dialectic of the second *Critique* (5:143 f.), where Kant alters Wizenmann's example by changing the content of the ill-founded belief from the love of a beloved person to the existence of an ideal beauty.

absence of evidence for it.[3] For this, two conditions must be given. (1) There can be no possible empirical evidence nor a conclusive argument a priori for or against the belief in question; let's call such a belief 'theoretically undecidable'. (2) Someone who acknowledges the moral law as binding must, by a kind of subjective, but still rational necessity, hold the belief in question; we may say that such a belief is 'practically necessary'. (Note that there are two 'necessities' involved here: the necessity, i.e. the unconditional bindingness, of the moral law, and the necessity that links the belief in question to the moral law.)

If Kant is right in that there can be rational belief without evidence, he needs to give a different answer to our question of what, if anything, is irrational in Peter's belief that Mary loves him. Kant's answer, not surprisingly, is that in this case neither of the two conditions for rational belief without evidence is satisfied, since Peter's belief is neither theoretically undecidable nor practically necessary.[4]

A theoretical proposition (a proposition meant to capture 'what is', aiming at being true; see *Logic* 9:86) that is both theoretically undecidable and practically necessary Kant calls a 'postulate of pure practical reason' ('a *theoretical* proposition, though one not demonstrable as such, insofar as it is attached inseparably to an a priori unconditionally valid *practical* law', 5:122). And he argues that there are exactly three such postulates, which concern (1) the existence of God, (2) our own transcendental freedom and (3) our immortality. With respect to these postulates, Kant holds that it is rational for us to *believe* in their truth even though they lie completely outside the reach of human knowledge.

In this chapter, I engage only briefly with the content of Kant's three postulates and with his arguments in their favour. Instead, I want to concentrate on the general idea of a postulate of pure practical reason and Kant's arguments for their very possibility. Kant's arguments and reflections concerning the postulates in general are laid out in several sections of the second chapter of the Dialectic, the most important of which include

[3] Note that by 'evidence' for some proposition *p* I mean reasons and considerations that tell in favour of *p* (i.e. in favour of the *truth* of that proposition). As Kant claims, there can be reasons for believing something which are not evidence in this sense, since these reasons do not bear on the truth of the proposition in question, but on the belief's relation to the subject's moral commitments. Hence, to deny that there is evidence for the propositions in question is not to deny that there are reasons for holding them. Note further that these non-evidential reasons need not be sufficient for acquiring the relevant beliefs, but only to show that a given belief one already holds is rationally warranted. I return to this point at the end of the chapter.

[4] In his reply to Wizenmann (see footnote above), Kant mentions only the second point.

sections 3 (*CpV* 5:119–21) and 6 through 8 (*CpV* 5:132–46).[5] These reflections concern nothing less than the fundamental structure of human reason – in Kantian terms, the question of the 'unity' of practical and theoretical reason and which side has primacy. I start by commenting on Kant's argument in section 3, 'On the Primacy of Practical Reason', with a special emphasis on the logical structure of that argument (section II). Next, I critically discuss – and ultimately defend – that argument (section III). Then, I look more closely at the idea of a postulate of pure practical reason and at the epistemological status Kant assigns to it (section IV). I close by suggesting that the main idea behind Kant's argument does not depend on Kant's own, demanding conception of morality (section V).

II. THE STRUCTURE OF KANT'S ARGUMENT FOR PRIMACY OF PRACTICAL REASON

In the section 'On the Primacy of Pure Practical Reason in Its Connection with Speculative Reason', Kant in effect argues for the claim that, under certain conditions, belief without evidence can be rational. His argument is difficult to follow, and even more difficult to assess, among other things because it is framed in terms of the respective 'interests' of two cognitive faculties. That a practically necessary belief can be rational without evidence is taken to be equivalent to the claim that pure practical reason has primacy over pure speculative reason. The section consists of only four paragraphs. The first introduces the idea of a primacy of one faculty over another and the concept of the 'interest' of reason in its various employments. The second lays out and explains three options: either speculative reason has primacy over practical reason, or vice versa, or neither has primacy over the other. The third and fourth paragraphs then contain Kant's argument for the primacy of practical reason. I briefly discuss each paragraph in turn; in the next section, I then ask whether Kant's argument for the primacy of practical reason (and thus for the possibility of rational belief without evidence) is convincing.

Kant starts out with a general definition of 'primacy among two or more things connected by reason' (*CpV* 5:119), but immediately turns to a more

[5] Earlier versions of Kant's arguments concerning the postulates can be found in the 'Canon' of the *Critique of Pure Reason* (A804/B832–A819/B847) and the 1786 essay 'What is Orientation in Thinking?' (in Kant, *Religion and Rational Theology*, ed. Allen W. Wood and George di Giovanni (Cambridge and New York: Cambridge University Press, 1996), 8:131–47), later restatements in the *Critique of Judgment* (section 87) and, very briefly, in *R* (6:6–7fn.) and in the 'Proclamation of a Treaty' from 1796 (8:409–22).

specific, namely 'practical' sense of that term.[6] What is practical about this sense of 'primacy' is that it concerns the subordination of 'interests' (all of which, as Kant will point out later, are ultimately practical). Although Kant at first, somewhat oddly, speaks of the interest of 'things' in general, it soon becomes clear that what he has in mind is the interest of 'faculties' or cognitive capacities. Such an interest is the fundamental principle according to which a specific 'faculty' is functioning and which therefore contains 'the condition under which alone its exercise is promoted. Reason, as the faculty of principles, determines the interest of all the powers of the mind but itself determines its own' (*CpV* 5:119–20). As Kant points out some lines later, 'consistency with itself' is no object of the interest of reason, since it is a 'condition of having reason at all.' An interest, by contrast, is not directed at constitutive conditions, but only at an 'extension'. (That consistency is constitutive of reason will turn out to be a cornerstone of Kant's argument for the primacy of practical reason.)

Next, Kant distinguishes between the interest of the 'speculative' and the 'practical use of reason'. (As we will see later, the distinction between practical and speculative reason in the title of the section is only shorthand for the distinction between the respective 'uses' of reason; in fact, Kant employs both distinctions interchangeably.) The 'interest' of the speculative use of reason 'consists in the *cognition* of the object up to the highest a priori principles' (*CpV* 5:120). Kant here builds on his diagnosis of the fallacies of pure reason in the Dialectic of the first *Critique*, where he had argued that reason, by looking for a condition for everything conditioned, can find rest only in something unconditioned (in '*highest* a priori principles'). However, since all empirical objects are conditioned in many ways (causally, spatially, temporally, and so forth), and since, as Kant had argued in the Analytic of the first *Critique*, all our knowledge (and even all our thinking about particular objects) is limited to the realm of empirical objects, we can never have knowledge of something unconditioned. And since it is through the use of pure reason itself that we can detect these limitations of our epistemic situation, Kant, later in the section on the primacy of practical reason, includes in the interest of speculative reason the respect for the 'bounds' of our knowledge (*CpV* 5:120) and 'the restriction of speculative mischief' (*CpV* 5:121) – that is, avoidance of claims about the unconditioned.

To Kant's rather broad interpretation of the interest of speculative reason it might be objected that, according to the Dialectic of the first *Critique*, it is

[6] Since it is this sense in which practical reason will turn out to have primacy, this seems to leave open the possibility that it fails to have primacy in the more general sense.

precisely reason's interest in 'cognition of the object up to its highest a priori principles' that brings about 'speculative mischief' in the first place. How can the avoidance of the latter be in the interest of speculative reason? But there is no real problem here, since after all it is '*cognition*' (a kind of knowledge) that speculative reason is interested in and not fallacies and empty speculation. Hence, avoidance of the latter is in the interest of speculative reason even though, somewhat tragically, its very interest in the cognition of ultimate principles brings about the mischief that is to be avoided.

By contrast, the interest of reason in 'its practical use consists in the determination of the will with respect to the final and complete end'.[7] The final and complete end of the will, at least in so far as it is determined by (pure) practical reason, is the highest good – a world whose inhabitants are completely virtuous and, because of their virtue, completely happy.[8] Kant had argued in the Antinomy sections (*CpV* 5:113 ff.) that, as rational beings bound by the moral law, we *must*, as a matter of rational necessity, try to realize the highest good. However, as finite beings, we have to acknowledge that it is beyond our powers to do so, since (i) we cannot achieve full virtue – which Kant now identifies with 'holiness' – but can only approach it asymptotically; and (ii) even to the degree we are successful in approaching holiness, this does not necessarily contribute to our happiness. Only if we assume that our souls are immortal and that there is an almighty and benevolent God can we conceive of the highest good as a possible end of our will. And since the highest good is a necessary end of the will of every rational agent, as rational agents we must accept that there is a God and that we are immortal.

For the purposes of this chapter, I will assume that the antinomy of the highest good is a real one and that Kant's resolution of it is convincing. (As will emerge, however, Kant's argument for the primacy of practical reason is quite independent from this assumption and thus from the content of the

[7] It seems that Kant's general definition of the 'interest' of a faculty as 'a principle that contains the condition under which alone its exercise is promoted' (*CpV* 5:119) does not fit his more specific talk of the 'interest' of the speculative and practical use of reason, respectively. After all, 'cognition of the object' and 'determination of the will' do not seem to be principles that contain 'the condition under which alone [the] exercise [of reason] is promoted'.

[8] On Kant's conception of the highest good, see Klaus Düsing, 'Das Problem des höchsten Gutes in Kants praktischer Philosophie', *Kant-Studien* 62 (1971), 5–42; Andrews Reath, 'Two Conceptions of the Highest Good in Kant', *Journal of the History of Philosophy* 26 (1988), 593–619; Stephen Engstrom, 'The Concept of the Highest Good in Kant's Moral Theory', *Philosophy and Phenomenological Research* 5 (1992), 747–80; Pauline Kleingeld, 'What Do the Virtuous Hope For? Re-reading Kant's Doctrine of the Highest Good', in H. Robinson (ed.), *Proceedings of the Eighth International Kant Congress*, vol. I.i (Milwaukee: Marquette University Press, 1995), 91–112.

Kantian postulates.) This means that the 'interest' of practical reason in the 'determination of the will with respect to the final and complete end' can only be realized in a being who believes in God and immortality. But as Kant had argued in the first *Critique*, our representations of God and immortal souls are transcendent ideas – representations of something unconditioned that can never be an object of human knowledge. This means that, as far as *speculative* reason is concerned, belief in their existence can only be the result of 'speculative mischief' and hence must be avoided.

With this, we have in place the potential conflict between speculative and practical reason that calls for a solution by granting primacy to one of them: practical reason requires us to believe in God and immortality, while speculative reason requires us to reject these propositions. (Note that to *reject* a proposition, in the relevant sense, does not require that one takes the proposition to be false, but merely that one does not accept it as true. Hence, acceptance and rejection of a given proposition are mutually exclusive and exhaustive alternatives – a feature on which Kant's argument will depend.)

In the second paragraph, Kant lays out three options for how to respond to this conflict between practical and speculative reason. I believe that Kant's argument for the primacy of practical reason cannot be adequately understood without close attention to the logical structure of this paragraph. As becomes clear from the way in which Kant introduces the options he considers, he constructs his argument as a complex 'disjunctive syllogism', that is, an argument of the general form: either *A* or *non-A*; if *non-A*, either *B* or *non-B*; neither *A* nor *B*; thus, *non-B*.

As Kant explains in his *Lectures on Logic*, the major premise of a disjunctive syllogism must always be a dichotomy (a disjunction of *two* contradictory statements, *A* and *non-A*) (see *Logic* 9:130, 147). If we need three disjuncts for our argument (*A*, *B*, and *C*), then *B* and *C* must be derivable by a dichotomical distinction within *non-A*: *A* or *non-A*; if *non-A*, then either *B* or *non-B* (= *C*). Only in this way, Kant points out, can we be sure of the exhaustiveness of the disjunction and hence of the validity of our syllogism (*Logic* 9:130).

Now if we look at the second paragraph of the section 'On the Primacy of Practical Reason', we find that this is precisely how Kant proceeds. This fact, however, is obscured by two features of Kant's presentation: first, the fundamental dichotomy (*A* or *non-A*) is not easily recognizable as a dichotomy at all. Its first disjunct (*A*) is that practical reason may accept only claims authenticated by speculative reason, in which case speculative reason has primacy. The second disjunct (*non-A*) is that practical reason has its own

'original a priori principles' (i.e. the moral law) to which certain 'theoretical positions' (that go beyond all knowledge possible for speculative reason, i.e. belief in God and immortality) are 'inseparably connected'. (We will see later why Kant's phrasing is somewhat vague when it comes to the relation between the practical principles and the theoretical positions.) In order to see that these two options are contradictory negations of each other, we can rephrase the second disjunct (in the terminology introduced above) by saying that there are postulates of pure practical reason – theoretical propositions that are theoretically undecidable but practically necessary. (Note, however, that Kant himself introduces the term 'postulate' in this context only in the next section; *CpV* 5:122.) We now can reformulate the dichotomy as follows: either there aren't (*A*) or there are (*non-A*) postulates of pure practical reason. If there aren't, practical reason must accept only propositions validated by speculative reason, which then has primacy.

Notice that as Kant sets things up, it is impossible for there to be postulates of pure practical reason if, because speculative reason has primacy, practical reason is denied the right to these propositions. Recall that the postulates (or, as they are called in the paragraph under consideration, the 'theoretical positions') are '*inseparably*' connected to a necessary principle of practical reason. Hence, practical reason, on pain of self-contradiction, *cannot* reject these propositions (if there are any). But then, if there are postulates of pure practical reason, speculative reason cannot have primacy, since this would mean that practical reason, *per impossibile*, was required to reject the postulates. Conversely, if speculative reason has primacy, there can be no postulates. So it emerges that Kant indeed builds his argument on an exhaustive dichotomy.

The second feature that obscures the logical structure of Kant's argument is that he introduces the two 'subdivisions' (see *Logic* 9:130) *B* and *non-B* by saying that given *non-A* (there are postulates of pure practical reason), 'the question is, which interest is supreme' (*CpV* 5:120). Since this question is equivalent to the question of which side has primacy (*CpV* 5:119) and the next option introduced (*B*) is that of the primacy of practical reason, this suggests that the third option (*non-B*) consists in the primacy of speculative reason (and thus would be identical to the first option, *A*). The paragraph would then have the form: either *A* or *non-A*; if *non-A*, either *B* or *A*, which involves an outright contradiction. This cannot possibly have been what Kant had in mind. And in fact, it is not how Kant continues. Kant next introduces the possibility (*B*) that, given there are postulates of pure practical reason, practical reason has primacy, which would mean that speculative reason must accept the postulates and, 'although they are

transcendent for it, try to unite them, as a foreign possession handed over to it, with its own concepts' (*CpV* 5:120). This is contrasted with the possibility (*non-B*) that even though there are postulates of pure practical reason, speculative reason is justified in rejecting the postulates because their acceptance goes against its own interest (which interest requires it to respect the 'bounds' it has set for itself – i.e. to avoid 'speculative mischief'). This would mean that, on the one hand, practical reason follows its interest in determining the will, with the consequence that belief in God and immortality is rationally required, while, on the other hand, speculative reason clings to its own interest and rejects the very beliefs rationally prescribed by practical reason. In this case, *no side would have primacy*; both uses of reason would be, as Kant puts it in the final paragraph of the section, 'merely juxtaposed (coordinate)' (5:121).[9]

The upshot of all this is that Kant, in the second paragraph, argues as follows: either there are no postulates, in which case speculative reason has primacy (A) or there are postulates (*non-A*), in which case there are two possibilities: either speculative reason must accept these postulates, in which case practical reason has primacy (B), or speculative reason does not have to accept these postulates, which means that no side has primacy (non-B). In brief, Kant introduces three options: either (*A*) speculative reason has primacy or (*B*) practical reason has primacy or (*non-B*) no side has primacy. I will return to the question of whether this triad is exhaustive in the next section.

It is in introducing the second of the three options, (*B*), that Kant explicitly links the idea of a primacy of practical reason with the claim that there can be rational belief without evidence. Kant stresses that the 'positions' (or, as Kant calls them some lines later, the 'propositions') in question, although they derive their rational warrant from being connected to an a priori principle of *practical* reason, are '*theoretical*' positions. (As we will see later, the attitude we are meant to take towards them is a kind of 'taking to be true' – *Fürwahrhalten* – namely 'belief'.) On the other hand,

[9] If this interpretation is correct, then why does Kant introduce options *B* and *non-B* by saying that 'the question is, which interest is supreme' (*das oberste*, literally 'the highest')? One way to read this passage is to emphasize that the question comes up only under a certain condition: if there are no postulates of practical reason, then there is no question that speculative reason has primacy. 'Supposing, however', that there are such postulates ('certain theoretical positions' connected to practical a priori principles), *then* there is a real question of 'which interest is supreme'. Since the possibility that the interest of speculative reason is supreme has already been mentioned, Kant then goes on to mention the possibility (*B*) that the interest of practical reason is supreme and the possibility (*non-B*) that neither interest is supreme. It remains slightly odd, however, that the latter possibility should be introduced with the question of 'which interest is supreme'.

we do not have the kind of evidence for them required by theoretical reason – no evidence that they in fact are true. They are 'withdrawn from any possible insight of speculative reason', which therefore 'knows nothing' about them (*CpV* 5:120). This means that if practical reason has primacy, then there are propositions – the postulates – we are rationally warranted to believe even though we lack any evidence for their truth. At the same time, Kant stresses that this lack of knowledge must go both ways, since the propositions in question 'must not contradict' any insight of speculative reason (*CpV* 5:120). Not only do we lack evidence that the propositions are true; neither must there be any evidence that they are false – indeed, not even *possible* evidence, since we are talking about synthetic propositions a priori, for which, according to Kant, we either have or don't have sufficient a priori reasons. This is what I had called 'theoretical undecidability': there is no possible empirical evidence and no argument a priori sufficient for establishing either the truth or the falsity of the propositions in question.

Kant starts the third paragraph by granting that if practical reason were only concerned with 'regulating the inclinations by the sensible principle of happiness' – in other words, if there were no pure, but only empirical practical reason – then speculative reason would have primacy. In this case, reason would not have the 'original a priori principles' with which 'certain theoretical positions' could then be 'inseparably connected'. To be sure, empirical practical reason may well lead to principles connected with theoretical commitments. Kant mentions 'Mohammed's paradise' and a mystical 'fusion with the Deity' in this connection: the mystic, for instance, believes that being united with the all-encompassing Deity means eternal bliss; hence, his reason directs him to work his way up to fusion with the Deity, which he must therefore believe to exist. What is missing in cases like this, however, is the necessity of the practical principle (here the principle 'Strive for unification with the Deity') on which the theoretical proposition ('There exists an all-encompassing Deity') is based, since the principle cannot be shown to be valid for every human being and hence is not 'necessary' in Kant's sense. But if the principle that commits us, say, to a belief in the existence of the Deity is not necessary, then it is *possible* to solve the conflict with speculative reason (which, following its own interest, requires us to reject that belief) by giving up the practical principle in question. In other words: if my empirically based practical principle commits me to believe things that cannot possibly be known, the rational course to take, according to Kant, is not to believe such 'monstrosities' (*CpV* 5:121) as a mystic Deity, but rather to give up the practical principle. (I will return to this issue in the final section.)

Things look completely different, according to Kant, 'if pure reason of itself can be and really is practical', as Kant takes himself to have shown in the Analytic. In this case 'it is clear' that theoretically undecidable propositions have to be accepted by speculative reason 'as soon as these same propositions belong inseparably to the practical interest of pure reason' (*CpV* 5:121). The reason Kant gives for this claim is that 'it is still only one and the same reason which, whether from a theoretical or a practical perspective, judges according to a priori principles'. Obviously, this brief remark as such is no convincing argument for the primacy of practical reason and I don't think that it is meant to be. Rather, Kant's thought seems to be something like this: if pure reason is practical, this means that there are unconditionally binding practical principles. In this case, the practical interest of reason is to determine the will in accordance with these principles. If for that purpose it is necessary to believe in certain theoretical propositions (the postulates), then speculative reason cannot reject these principles, but has to accept them 'as something offered to it from another source, which has not grown on its own land but yet is sufficiently authenticated' (*CpV* 5:121). Speculative reason cannot reject these propositions, because they are rationally warranted by practical reason, and it is 'still only one and the same reason which, whether from a theoretical or a practical perspective, judges according to a priori principles'. Read this way, the impact of Kant's insistence on the unity of reason is to make clear that it is impossible rationally to accept some proposition (from a practical point of view) but at the same time to reject it (from a theoretical point of view).

That this is indeed the way Kant's argument is meant to proceed is confirmed if we now turn to the fourth paragraph. It offers a highly condensed, but formally complete and self-sufficient argument for the primacy of pure practical reason. Kant begins with an explicit statement of the primacy thesis and the condition under which it is meant to hold: 'Thus, in the union of pure speculative with pure practical reason in one cognition, the latter has primacy, assuming that this union is not *contingent* and discretionary but based a priori on reason itself and therefore *necessary*' (*CpV* 5:121). Kant then introduces two considerations, each of which is designed to exclude one of the three options mentioned in the second paragraph. First, Kant points out that the two uses of reason cannot be 'merely juxtaposed' (option *non-B*), because this would result in a 'conflict of reason with itself', since the same propositions would be accepted by practical reason and rejected by speculative reason. Second, speculative reason cannot have primacy (option *A*), since 'all interest is ultimately

practical and even that of speculative reason is only conditional and is complete in practical use alone' (5:121). If these dense remarks can be unpacked into convincing arguments, and if the choice of three options from the second paragraph is exhaustive, then Kant has argued successfully for the primacy of practical reason (option *B*).

III. KANT'S ARGUMENT FOR THE PRIMACY OF PRACTICAL REASON ASSESSED

In a recent paper on the primacy of practical reason, Sebastian Gardner has suggested that in the section under consideration Kant offers three different arguments for the primacy of practical reason.[10] If the above reading of the way in which Kant structures the section is correct, however, it seems much more plausible to assume that he wants to present one extended argument, the logical form of which is that of a 'disjunctive syllogism' (see *Logic* 9:130):

(1) Either speculative reason has primacy over practical reason (*A*) or practical reason has primacy over speculative reason (*B*) or neither has primacy over the other (*non-B*).
(2) Either speculative reason or practical reason must have primacy (not *non-B*).
(3) Speculative reason does not have primacy (not *A*).
(4) Thus, practical reason has primacy (*B*).

This is a logically valid argument. So let us turn to the premises. Since Kant considers the subordination of the respective interests of only two faculties, namely speculative and practical reason, I think we must grant Kant the first premise, according to which there are only three possibilities. Logically speaking, 'having primacy' functions as a transitive, non-reflexive and asymmetrical relation like 'longer than'. Given two distances *D1* and *D2*, for instance, there are only three possibilities: either *D1* is longer than *D2*, or *D2* is longer than *D1* or none is longer than the other, i.e. both are of the same length. Similarly, either speculative reason has primacy, or practical reason has primacy or neither has primacy, i.e. both are on an equal footing.

We can see that this is not just an artefact of logic, but a philosophically substantial truth, if we consider what the primacy of the one use of reason over the other is meant to consist in. The issue at stake is how reason is to proceed with respect to propositions that are both theoretically undecidable

[10] Sebastian Gardner, 'The Primacy of Practical Reason', in Graham Bird (ed.), *A Companion to Kant* (Oxford: Blackwell, 2006), 259–73, here 265.

('withdrawn from any possible insight of speculative reason') and practically necessary ('inseparably connected' with 'original a priori principles' of pure practical reason). Considered merely from the perspective of speculative reason, these propositions (the postulates) must be rejected (because we can't tell whether they are true). Considered from the perspective of practical reason, they must be accepted (because they are practically necessary).[11] If speculative reason had primacy, then the interest of speculative reason would prevail and *both* speculative and practical reason would have to *reject* these propositions. If practical reason had primacy, by contrast, then they would have to be *accepted* by *both* speculative and practical reason. And if neither has primacy, then the propositions in question would have to be *rejected* by speculative reason and *accepted* by practical reason. A little reflection shows that indeed there are only these three possibilities.[12]

So much for the first premise of Kant's disjunctive syllogism. Let us now turn to Kant's argument for his second premise, which denies that speculative reason and practical reason could coexist on equal footing, without any primacy on either side. Kant writes that 'without this subordination a conflict of reason with itself would arise', since speculative reason would 'admit nothing' from practical reason 'into its domain, while the latter would extend its boundaries over everything' (*CpV* 5:121). In other words, without primacy on either side, speculative reason would reject the postulates while practical reason would accept them. It might be asked what is so bad about such a situation.[13] One might insist that for theoretical purposes, the postulates, being theoretically undecidable, are not sufficiently warranted. But this need not exclude relying on them for practical purposes. As long as we keep the speculative and the practical uses of reason apart, no conflict need ensue.

What blocks this solution, however, is the conjunction of three points Kant had already mentioned in the previous passages of the section (and thus may expect his readers to have kept in mind). The first is that the postulates are *theoretical* propositions, as Kant had emphasized in the second paragraph (*theoretische Positionen*, *CpV* 5:120) and as he repeats in the next section in his definition of a 'postulate of pure practical reason'

[11] Remember that 'rejection' here does not mean 'taking to be false', but merely 'not taking to be true'; hence, rejection of a proposition is equivalent to its non-acceptance.

[12] The two pairs of mutually exclusive disjuncts (speculative/practical; reject/accept) allow for four combinations. Kant does not consider the logically possible fourth combination – the propositions are accepted by speculative reason and rejected by practical reason – since it is philosophically absurd. This absurd possibility corresponds, in the distance analogy, to the logical impossibility of both distances being longer than the other (both speculative reason and practical reason having primacy).

[13] See Gardner, 'Primacy of Practical Reason', 266.

(*CpV* 5:122). I think that it is a major insight on Kant's part that there can be theoretical propositions that follow from (or, at least, are 'inseparably connected' to) practical attitudes – ontological commitments we undertake not by accepting a theory, but rather by wanting or willing something and by accepting practical commands and imperatives as binding. As such, the imperative 'Open the window in this room!' does not conflict with the theoretical proposition that there is no window in the room. However, *accepting* the imperative 'Open the window in this room!' as normatively binding commits one to accepting the claim that there is a window in the room. Hence, a rational person cannot consistently act on that imperative and at the same time deny that there is a window in the room. If such an ontological commitment is based on a command that is not merely contingent, but rationally necessary (see *CpV* 5:121), its truth cannot rationally be denied. For instance, if Kant is correct to claim that we are obligated to realize the highest good, and that this is possible only if God exists, then we are rationally committed to believing that God exists.

Against this, one might object that it might be sufficient to *hope* that the ontological presuppositions of our moral duties are fulfilled,[14] where 'hope' is a doxastic state weaker than belief, so that hoping that *p* is incompatible with believing that *p* and compatible with not believing that *p*. This is a very serious challenge to which I cannot do justice here. A Kantian response, I think, has to begin from the practical necessity of realizing the highest good. Kant's argument for the postulates of God and immortality, very roughly, proceeds as follows. (1) It is morally necessary (i.e. obligatory) to realize the highest good. Therefore, (2) it is possible to realize the highest good. But (3) it is possible to realize the highest good only if God exists and if our souls are immortal. Therefore, God exists and our souls are immortal. The step from (1) to (2) relies on the principle *ultra posse nemo obligatur* (no one is obligated to do what it is impossible for him to do). Kant argues extensively for premises (1) and (3); to evaluate the strength of these arguments lies beyond the scope of this essay. If we grant (1), (3) and the *ultra posse* principle, then (4) follows by elementary logic. Now let us embed this argument in suitable doxastic states of a person who accepts that argument and its premises. (1′) S accepts that it is practically necessary to realize the highest good. (2′) S believes that it is possible to realize the highest good. (3′) S believes that it is possible to realize the highest good only if God exists and our souls are immortal. (4′) S believes that God exists and our souls are immortal. If we were to replace (4′) with (4″) S hopes that God exists and

[14] See Gardner, 'Primacy of Practical Reason', 266.

our souls are immortal, a contradiction, or inconsistency, would follow (given that 'hope that *p*' is weaker than, and incompatible with, 'believe that *p*'), since from (4″) it follows that (5) S does not believe (but only hopes) that God exists and that our souls are immortal. But then, it seems, S cannot consistently believe (2) that it is possible to realize the highest good, given that (3′) S believes that it is possible to realize the highest good only if God exists and our souls are immortal. S could only hope, but not believe, that it is possible to realize the highest good. But if S accepts that it is practically *necessary* (obligatory) to realize the highest good, she must *believe* that this is possible.[15]

The second point relevant to Kant's argument against a coordination of speculative and practical reason concerns the unity of reason invoked in the third paragraph: that 'it is still only one and the same reason which, whether from a theoretical or a practical perspective, judges according to a priori principles' (*CpV* 5:121). It is notoriously obscure what precisely Kant means by this claim. One aspect, at least, is that in distinguishing between speculative (or theoretical) and practical reason, Kant is not talking about two different faculties, but about two 'uses' of the same faculty. And talk of 'uses' brings to mind that there must be *someone* who uses this faculty – namely the person whose rational faculty it is. Kant's talk of speculative and practical reason, of their interests, their potential conflicts and the primacy of the one over the other may often suggest that he is talking about so many agents, each with its own agenda. But this is not what Kant wants to say. Rather, the only kinds of agents he acknowledges are human beings, who, if

[15] Unfortunately, this doesn't settle the matter, since now one might question whether (1) gives adequate expression to our obligation with respect to the highest good. Shouldn't it rather read: (1″) It is practically necessary to *work towards* realizing the highest good? It seems that someone who accepts (1″) is committed to *believing* that it is possible to *work towards* realizing the highest good, for which it may be sufficient to *hope* that it is possible to realize the highest good. Some of Kant's own formulations are ambiguous in this respect, for instance when he writes: 'It is a duty to realize the highest good to the utmost of our capacity; therefore it must be possible [*daher muss es doch auch möglich sein*]; hence it is also unavoidable for every rational being in the world to assume what is necessary for its objective possibility. The assumption is as necessary as the moral law, in relation to which alone it is valid' (*CpV* 5:143fn.). If Kant wants to say that it must be possible to realize the highest good *to the utmost of our capacity* (which, after all, would be compatible with its being impossible to realize the highest good), then this indeed follows from the duty Kant mentions. But then it is doubtful whether it follows that every rational being must believe that the necessary conditions for realizing the highest good are given. If Kant, on the other hand, wants to say that it must be possible to *realize* the highest good, it may indeed follow that every rational being must believe that the necessary conditions for its realization are given. But then it is questionable whether it follows from accepting the duty to realize the highest good *to the utmost of our capacity* that we must believe the highest good to be *realizable*. It therefore seems that Kant's argument for the postulates of God and immortality requires an unconditional duty to *realize* the highest good and not just a duty to realize it to the utmost of our capacity. (I am grateful to Sebastian Gardner for pressing me on this point.)

everything goes well, are capable of rational thinking. This capability may be employed for two different (although ultimately connected) purposes: explaining the multitude of empirical phenomena by as few fundamental principles as possible and determining the will according to universally valid principles of action.

It is a deep and difficult question how far, according to Kant, these are two employments of the same 'faculty'.[16] But for our purposes, we can set this question aside, since what matters here is not so much the unity of the *faculty* of reason in its various employments (i.e. the unity of its 'principles'), but rather the unity of the rational being who employs his reason for different purposes. Thus, Kant says that 'in the union of pure speculative with pure practical reason in one cognition [*in einer Erkenntnis*], the latter has primacy' (*CpV* 5:121). I take this to mean that ultimately the specific propositions of speculative and practical reason, respectively, must be incorporated into *one cognitive system*, since they are only so many uses of the same rational faculty by the same rational being. If speculative reason would issue only theoretical propositions and practical reason only practical ones (imperatives, maxims and so on), there would be no problem of uniting them into 'one cognition'. But as Kant noticed (that was the first point above), practical reason brings with it the commitment to *theoretical* propositions which, in case of a conflict with commitments of speculative reason, would render the subject's belief system inconsistent.

This takes us to the third point Kant relies on in his argument against the unhierarchical 'mere juxtaposition' of speculative and practical reason, that consistency (or at least the attempt to avoid inconstancies) is a constitutive condition of having reason at all (*CpV* 5:120). This means that if practical reason is required by its interest to accept certain *theoretical* propositions, and speculative reason is required by its interest to reject the very propositions accepted by practical reason, then the result is something not recognizable as reason at all. After all, it would be the same person whose practical reason commits her to accepting propositions (holding them to be true) she must reject on behalf of her speculative reason. This person would not so much contradict herself but rather defy rational description: she would believe and not believe one and the same thing. We would not know what to say of that person, except that we cannot make any rational sense of her.

I think that this is Kant's argument against the 'mere juxtaposition (coordination)' of speculative and practical reason: if there was no primacy

[16] On this topic, see Pauline Kleingeld, 'Kant on the Unity of Theoretical and Practical Reason', *Review of Metaphysics* 52 (1998), 311–39.

on either side, then the same theoretical propositions would have to be both accepted as true and rejected by the same rational being (within the same system of rational beliefs); but this would lead into inconsistency and thus undermine that person's claim to rationality.

Against this, it may be objected that local inconsistencies are a widespread phenomenon – no belief system of any human being is entirely consistent. But this does not mean that no human being possesses reason. Hence, contrary to what Kant assumes, consistency is not a constitutive condition of 'having reason at all' (*CpV* 5:120). But even if this is granted, Kant's argument remains intact, since what he discusses is not just some minor and local inconsistency, but rather a 'conflict of reason with itself' (*CpV* 5:121): a conflict concerning the relation between what we have to accept as rational *agents* and what we must reject as rational *thinkers*. Moreover, we would have to acknowledge that the conflict and the inconsistencies that result from it are insurmountable. Whereas local inconsistencies *can be* avoided once they are detected, no such route would be open to us if propositions 'inseparably connected' with a priori practical principles had to be rejected by speculative reason. Our own reason would lead us into inconsistencies that, in contrast to the fallacies of pure speculative reason diagnosed in the Transcendental Dialectic, could not even be overcome by a 'critique of pure reason'. I think that Kant has given us good reason to reject this possibility.

This leaves the third premise of Kant's syllogism, which denies that speculative reason has primacy. Kant's explicit argument again is very brief: 'one cannot require pure practical reason to be subordinate to speculative reason … since all interest is ultimately practical and even that of speculative reason is only conditional and is complete in practical use alone' (*CpV* 5:121). Against this argument, Sebastian Gardner has objected that it 'seems too quick'.[17] He reads the phrase 'all interest is ultimately practical' as saying that 'there is spontaneity in the theoretical employment of reason' and that 'theoretical reasoning is a case of doing something' and he objects that from this it does not follow 'that the conditions (principles, end, etc.) of practical reason are what explain, or what should determine, theoretical reasoning in any important or interesting sense'.[18] Admittedly, Kant's brief remark is less than an explicit argument. But I think that it can be given a much more interesting reading than the one considered by Gardner – a reading, indeed, in which the argument turns out to be quite convincing.

The first thing to keep in mind is that the argument has two parts, the first consisting of the claim that 'all interest is ultimately practical', the

[17] Gardner, 'Primacy of Practical Reason', 267. [18] Gardner, ibid., 267.

second of the claim that even the interest of speculative reason 'is only conditional and is complete in practical use alone.' I think it is the second claim, rather than the first, that is meant to carry the burden of the argument. That the interest of speculative reason is 'only conditional' (*nur bedingt*) means that reason is interested in the 'cognition of the object up to the highest a priori principles' only under some condition. The context suggests that the condition in question is that practical reason is able to determine the will according to the moral law. This may be taken in one of two ways.

On the one hand, it may mean that reason is interested in speculative knowledge about God and immortality not for its own sake, but rather for practical reasons, having to do with thinking of ourselves and the world in such a way that realizing the highest good becomes possible. This is clearly a view in close keeping with Kant's general philosophical outlook and with his conception of reason. Unfortunately, it sounds very much like a mere restatement of the primacy of practical reason, in which case the argument would be viciously circular.

But there is a different way of taking Kant's claim that the interest of speculative reason is 'conditional', namely as contrasting the conditional interest of speculative reason with the *unconditional* interest of practical reason based on the moral law. To see what this could mean, let us briefly look at a later section from the Dialectic, 'On Assent from a Need of Pure Reason', where Kant contrasts the 'hypotheses' of speculative reason with the 'postulates' of pure practical reason (*CpV* 5:142). Both are 'presuppositions' answering to a 'need of reason' and, indeed, their possible contents seem to be the same: God, freedom and immortality. But according to Kant, they differ in their epistemological status. A hypothesis is merely a 'most reasonable opinion' (*CpV* 5:142), a postulate is a kind of 'belief' (*CpV* 5:143), namely 'pure practical rational belief' (*CpV* 5:144). (I will return to the epistemological status of the postulates in the next section.) In order to explain this difference in epistemological standing, Kant mentions two major differences between hypotheses and postulates. First, while acceptance of the former depends on my will in so far as I may, but need not, 'satisfy completely my inquiring reason' (*CpV* 5:142), the latter are based on an 'inflexible command of reason' (i.e. the moral law) with respect to which we have no choice (*CpV* 5:143). For this reason, second, the hypotheses of speculative reason are merely 'permitted' (*CpV* 5:143), while the postulates of practical reason are based on a '*need from an absolutely necessary point of view*', namely that of the moral law (*CpV* 5:143). Kant does *not* claim that (while the hypotheses are permissible) the postulates are obligatory, since 'a

belief that is commanded is an absurdity' (*CpV* 5:144). However, it is clear that Kant allows for some leeway when it comes to our attitude to the hypotheses of speculative reason, whereas we have no such leeway when it comes to the postulates. The reason for this is that *only* pure practical reason issues categorical imperatives – unconditionally binding principles. It may be in the interest of speculative reason to cognize 'the object up to the highest a priori principles'. But still, it is up to us, as rational beings, how far to indulge our reason in this respect. Speculative reason can issue only hypothetical imperatives: 'If you want to satisfy your speculative interest, don't stop inquiring until you have gained knowledge of the highest a priori principles.' But obviously, it is not irrational to terminate inquiry (or even never to start it) if your livelihood is at stake or if moral considerations require you to employ your mental and physical powers otherwise. In this sense, the interest of speculative reason is 'only conditional': it directs us to do something (inquire into a priori principles) only under the condition that no other, overriding interests tell against it. The interest of pure practical reason, by contrast, is not conditional, since it is based on the unconditionally binding moral law. It cannot be overridden by other interests. And for this reason, speculative reason cannot have primacy over practical reason.[19]

The point of the famous remark that 'all interest is ultimately practical' would then be twofold: first, it secures the *commensurability* of the respective interests of speculative and practical reason. It is not as if the speculative interest was a completely different *kind of interest* whose rational standing was independent of that of the interest of practical reason. Rather, even the interest of speculative reason is practical in so far as it is based on a *need* of reason and also in so far as it directs us to *do* something (namely to inquire into a priori principles). Second, the remark prepares the way for the claim that the speculative interest is only conditional. Given that both interests are commensurable and thus may possibly conflict, and given further that the interest of practical reason is unconditional (based on a categorical imperative), the speculative interest can only be conditional.

[19] That this is what Kant means by the remark that the interest of speculative reason is 'only conditional' is confirmed by a parallel passage from 'What is Orientation in Thinking?': we can see that the 'need of reason … in its speculative employment' 'is only conditional, i.e. we must assume God's existence if we *want to judge* upon the first causes of the contingent' (8:139; Kant's emphasis). Kant proceeds: 'Far more important is the need of reason in its practical employment, since it is unconditional, and we are necessitated to assume God's existence not only if we *want* to judge, but because we *must* judge' (8:139). See also Kant's distinction in the Canon between belief based on a contingent end and belief based on a necessary end (*CrV* A823/B851).

In sum, by saying that all interest is practical and the speculative interest is only conditional Kant wants to outline the following argument: the moral law is absolutely binding for every rational being. It requires us to realize the highest good, which presupposes that we can conceive of the highest good as an end for our will. This we can do only if we accept the postulates. The interest of speculative reason, unless it is subordinated under the interest of practical reason, would require us to reject the postulates. But the interest of speculative reason in rejecting the postulates is only conditional, while the interest of practical reason in accepting the postulates is unconditional. Since, in case of a conflict, the conditional interest cannot prevail over the unconditional one, we cannot reject the postulates. Hence speculative reason cannot have primacy. Since Kant had already argued that practical and speculative reason cannot be merely coordinated, it follows that practical reason must have primacy.

I think that Kant, at least given his moral philosophy presented in the Analytic of the second *Critique*, offers us excellent reasons to accept the premises of his disjunctive syllogism and thus excellent reasons to accept the primacy of practical reason.[20]

[20] In fact, the argument for the claim that there can be rational belief without evidence rests only on two substantial premises. (1) There is an unconditional practical principle a priori (the moral law) that issues categorical imperatives no agent can rationally renounce. (2) This principle rationally commits us to accept as true certain theoretical propositions that are theoretically undecidable. From these two premises it follows that: (3) We are rationally committed to accept as true certain theoretically undecidable propositions. If things are that straightforward, why does Kant not present them this way? Because things get complicated by the addition of two further premises, each of which is highly plausible. (4) The principles of speculative reason require us not to accept as true theoretically undecidable propositions. (5) It is rationally impossible to both accept and not to accept as true the same proposition at the same time. But from (4) and (5), the negation of (3) follows: (6) We cannot be rationally committed to accept as true theoretically undecidable propositions. If we take for granted premises (1) and (2), then either (4) or (5) must be false. Since (5) looks fine, (4) has to be given up. And indeed, this is the premise Kant abandons by granting primacy to practical reason, which is equivalent to changing (4) into (4′). The principles of speculative reason require us not to accept as true theoretically undecidable propositions unless these are authenticated as postulates of pure practical reason. With this amendment, the inference from (1) and (2) to (3) meets no resistance, which means that the claim that we may rationally accept the postulates as true even though they are theoretically undecidable rests only on the two premises (1) and (2). In fact, this is implicitly acknowledged by Kant in the way he sets up the fundamental dichotomy in the second paragraph of the section on the primacy. There, Kant had claimed that *either* speculative reason has primacy *or* there are (1) original a priori principles of practical reason (2) to which there are inseparably connected certain theoretical propositions. Hence, given (1) and (2), speculative reason does not have primacy, which means that practical reason may accept the postulates as true. This is not the primacy of practical reason, yet, since there remains the possibility that the postulates are accepted by practical reason but rejected by speculative reason (option *non-B*, above). This possibility is excluded by (5), so that the primacy of practical reason rests on the three premises (1), (2) and (5), whereas the somewhat weaker claim that we are rationally committed to accepting the postulates as true requires only (1) and (2).

IV. THE DOXASTIC STATUS AND EPISTEMOLOGICAL STANDING OF THE POSTULATES

We have already heard that a postulate of pure practical reason, according to Kant, is 'a *theoretical* proposition, though one not demonstrable as such, insofar as it is attached inseparably to an a priori unconditionally valid *practical* law' (*CpV* 5:122). In sections 6 to 8 Kant's aim is to specify the doxastic status and epistemic standing of the postulates – what kind of attitude are we supposed to take to them and what degree of confidence in them is warranted. In the course of this reflection, Kant places the postulates in a complex net of considerations concerning, among other things, their relation to theoretical reason in general and to speculative reason in particular, their being based on a 'need of pure reason' and the question why, on Kant's account, belief in God (and immortality) is not morally obligatory. I cannot discuss Kant's finely nuanced views in all their detail, but must restrict myself to those aspects that are central to understanding his conception of a postulate.

The first thing to note is that belief in the postulates, according to Kant, is knowledge neither in the sense of being part of a systematic and unified scientific world-view (cognition, *Erkenntnis*) nor in the sense of being justified true belief (*Wissen*). Concerning the former, Kant devotes section 6 and the first part of section 8 to the question how far the postulates can be an 'extension of pure reason' and its knowledge in a *practical* respect without extending our speculative knowledge. The problem arises because the postulates (among which Kant now includes freedom as intelligible causality, *CpV* 5:132) seem to give us *some* kind of cognitive access to the very ideas of God, freedom and immortality about which Kant in the first *Critique* had shown that no knowledge about them is humanly possible (see *CpV* 5:133). If practical reason has primacy over speculative reason, then speculative reason must accept the postulates as true (see section II above). But this, Kant now insists, does not mean that we gain speculative knowledge of God, freedom and immortality (*CpV* 5:133). Nevertheless, Kant claims that our cognition is extended by the postulates – not our speculative knowledge about things in themselves, but our theoretical cognition of reason in general (*CpV* 5:135). Moreover, this is an extension only in a practical, not in a speculative respect.

There are three issues here that I would like to address. First, what is the epistemic status of the postulates and the kind of attitude we are supposed to take to them? Second, given that we do not *know* that the postulates are true, how can they nevertheless extend our cognition 'in a practical

respect'? And third, how does Kant avoid the consequence that belief in the postulates is morally obligatory?

Concerning the first question, our epistemic standing with respect to the postulates is not one of knowing that they are true; this is part of the reason why Kant calls them 'postulates'. They are 'not theoretical dogmas but *presuppositions* having a necessarily practical reference' (*CpV* 5:132). Hence, Kant does not contrast theoretical knowledge of the postulates, which we don't have, with some kind of 'practical knowledge' of them that we do have, but rather with rational *belief* or *faith* in their truth. (The German word Kant uses here, *Glauben*, can mean both 'belief' and 'faith'. In Kant, in the context of the postulates, it often has overtones of both, which is why I will translate *Glauben* as 'belief/faith'.) Kant doesn't draw on the distinction between knowledge and belief/faith (*Wissen* and *Glauben*), prominent in the B Introduction to the *Critique of Pure Reason*,[21] explicitly in the sections under consideration.[22] But implicitly Kant refers back to the section on '*Meinen*, *Wissen und Glauben*' ('Opinion, Knowledge and Belief/Faith') as the three kinds of 'taking something to be true' (*Fürwahrhalten*) in the Canon of the first *Critique* (A820/B848–A831/B59), a section that immediately follows the 1781 precursor to the argument for the postulates (*CrV* A804/B832–A819/B847). Since in the second *Critique*, Kant calls the attitude to the postulates one of *Fürwahrhalten* (taking to be true) (*CpV* 5:142), since he distinguishes the postulates from 'hypotheses' which are the object of our 'most reasonable opinion' (*CpV* 5:142), and since he calls our attitude to the postulates 'a pure practical rational faith/belief' (*reiner praktischer Vernunftglaube*; *CpV* 5:144), it will be worth taking a look at Kant's distinction between opinion, knowledge and belief/faith (which, incidentally, he draws not only in the first *Critique*, but repeatedly in his *Lectures on Logic*; see *Logic* 9:66–8).

In the section from the Canon, Kant distinguishes between three levels of 'taking something to be true': while 'opinion' (*Meinung*) is a case of taking something to be true that is 'consciously both subjectively and objectively insufficient', 'belief/faith' (*Glauben*) is 'only subjectively sufficient and at the same time is taken to be objectively insufficient' (*CrV* A822/B850). Knowledge (*Wissen*), by contrast, is a 'both subjectively and objectively

[21] 'I had to suspend knowledge in order to make room for faith' (*CrV* Bxxx).

[22] Kant uses the term *Wissen* in a relevant context within the *CpV* only once (*CpV* 5:57) in the section of the Analytic entitled 'On the Warrant of Pure Reason in Its Practical Use to an Extension Which Is Not Possible to It in Its Speculative Use' (*CpV* 5:50–7). That section mirrors section 7 of Book Two of the Dialectic in so far as Kant is concerned there with the question of how our cognitive access to freedom through the moral law does not contradict his denial of speculative knowledge.

sufficient taking to be true' (*CrV* A822/B850). This passage raises a number of exegetical problems, primarily because it seems to contradict Kant's initial definition of 'conviction' and 'persuasion' as two kinds of 'taking to be true' at the beginning of the section in question (*CrV* A820/B848).[23] Ignoring these issues here, we can say that the 'subjective sufficiency' Kant is talking about concerns the question how firmly the subject is convinced that what she takes to be true in fact is true, whereas the 'objective sufficiency' concerns the question whether the 'objective reasons' (*CrV* A820/B848) on which the 'taking to be true' is based are sufficient to guarantee the truth in question. (Kant identifies objective sufficiency with 'certainty'; *CrV* A822/B850.)

Thus, in a case of an opinion, the subject is conscious of the fact that her reasons for taking something to be true do not guarantee that it is in fact true, and hence she places only a limited degree of confidence in its being true. In a case of knowledge, by contrast, she is aware that her taking something to be true is based on reasons that guarantee the truth in question, so that she will be fully and wholeheartedly convinced of that truth. In a case of belief/faith, however, the subject is fully confident of the truth of what she takes to be true, *although* she is aware that her reasons fall short of guaranteeing that truth. (In this respect, 'faith' seems to be more appropriate than 'belief'.) Both in the *Critique of Pure Reason* and in the *Critique of Practical Reason*, belief/faith is the attitude Kant recommends us to take to the postulates.

In the Canon, Kant limits belief/faith explicitly and exclusively to practical issues, since for theoretical or speculative purposes, in cases where we don't have sufficient objective reasons, we have to suspend judgment (*CrV* A824/B852). In a practical respect, however, we often cannot wait for a proof before we take something to be true. Here, it is legitimate to put full confidence in something that, theoretically speaking, we cannot be certain about. Kant distinguishes between two cases: 'pragmatic belief/faith', which is based on hypothetical means–ends considerations and hence is only 'accidental' (*CrV* A824/B852), and 'moral belief/faith', which is based on moral considerations and hence is 'necessary' (*CrV* A828/B856). Since all this is in close keeping with Kant's views in the Dialectic of the second *Critique*, we may safely assume that by calling the attitude to the postulates a 'pure practical rational faith/belief' (*reiner praktischer Vernunftglaube*, *CpV* 5:144), Kant wants to distinguish that attitude from knowledge by insisting

[23] For an illuminating discussion of the third section of the Canon, see Andrew Chignell, 'Belief in Kant', *Philosophical Review* 116 (2007), 323–60.

that it is, although subjectively sufficient, objectively insufficient. This includes the awareness that the reasons we have for taking the postulates to be true fall short of guaranteeing their truth.

I now turn to the second issue mentioned above. What then does the extension of our cognition (*Erkenntnis*) by the postulates consist in, if our attitude to the postulates is not one of knowledge (*Wissen*)? As noted above, the extension does not concern 'speculative' cognition, which Kant here seems to identify with an insight into the inner make-up of things in themselves, a kind of knowledge that would express itself in 'synthetic judgments' about these objects (see *CpV* 5:133, 135, 137).[24] By the postulates we 'cognize neither the nature of our souls, nor the intelligible world, nor the supreme being as to what they are in themselves' (*CpV* 5:133). We do not understand how these things are possible, but only postulate that they exist (see *CpV* 5:133). Moreover, as we have just seen, we do not *know*, but only *believe* that they exist. Since *knowledge* (*Wissen*) about the supersensible, if only we could reach it, is the *only* kind of attitude suitable to speculation (see *CrV* A823/B851), the postulates do not constitute an extension of *speculative* knowledge (*Erkenntnis*).

However, they do constitute an extension of 'theoretical cognition, *not indeed of these objects* [God, freedom, immortality] but of reason in general … insofar as *objects were given* to those ideas by the practical postulates, a merely problematical thought having by this means first received objective reality' (*CpV* 5:135; cf. 5:136). This claim has to be read in the light of Kant's central doctrine in the first *Critique*, briefly summarized in section 7 (*CpV* 5:136), according to which, of the two fundamental kinds of human representations, only sensible intuitions refer to objects immediately, whereas concepts represent determinate objects only mediately through intuitions. Our representations of God, freedom and an immortal soul, however, do not have objects that can be given in experience, which means that they lack the necessary link to sensible intuition (see *CpV* 5:136). They are 'transcendental ideas' – representations our reason necessarily arrives at in metaphysical speculation, but which lack any determinate relation to objects in the world. They are 'merely problematical thoughts', which

[24] Note that Kant, even though he often seems to use the terms 'speculative' and 'theoretical' interchangeably, sometimes draws an explicit distinction between the two. In his *Lectures on Logic*, Kant defines 'speculative sentences' as those 'from which no rules for behaviour can be derived and which do not contain grounds for possible imperatives' (*Logic* 9:86). 'Theoretical cognitions', by contrast, 'are those that do not say what ought to be but what is and that thus do not have some action, but some existence as their object' (*Logic* 9:86). All speculative sentences are theoretical, but not vice versa.

means that they are logically possible, but that we do not know whether they have 'real possibility' or, which is the same, 'objective reality' (*CpV* 5:134).

Hence, when Kant says that these *ideas* of God, freedom and immortality 'receive objective reality' (e.g. *CpV* 5:135), this does not mean that their *objects* really exist (even though that is what the postulates postulate), but rather that the content of these ideas is specific enough in order to refer to determinate objects. Otherwise, they would be mere 'forms of thought' without 'sense and reference' (see *CrV* B148 f.). Only with respect to representations that have 'objective reality' does the question whether their objects exist or not make any sense, since otherwise it would be indeterminate what would count as such an object. Thus, although Kant's arguments for the postulates are meant to show that we must *believe* in the *existence* of God, freedom and immortality, the practical extension of our *cognition* by the postulates does not concern the existence of God, freedom and immortality, but only the 'objective reality' of our *concepts* of God, freedom and immortality. All that is gained by the postulates for the purposes of theoretical cognition (*Erkenntnis*) is that otherwise merely 'logically possible' and thus 'problematical' concepts now turn out to have 'real possibility'.[25]

Finally, let us turn to the question whether the 'inseparable connection' between the moral law and the postulates implies that belief in the postulates is morally obligatory. Kant, of course, wants to deny this. In section 8, after distinguishing between the 'hypotheses' of speculative reason and the 'postulates' of pure practical reason, Kant points out that the 'pure practical rational faith/belief' is not 'announced as itself a command', since 'a belief/faith that is commanded is an absurdity' (*CpV* 5:144). Although Kant doesn't explain here why belief/faith cannot be commanded, we may assume that this is because we cannot 'take something to be true' at will. But the sensible admission that belief/faith in the postulates is not morally obligatory lands Kant in a difficult position, since now he has to combine three claims that do not sit together well: (1) realizing (promoting) the highest good is morally obligatory (see *CpV* 5:113; 144fn.); (2) it is impossible to acknowledge this obligation without accepting the postulates (see *CpV* 5:134, 145); and (3) accepting the postulates is not morally obligatory (see *CpV* 5:144, 146). If we can do as the moral law requires only if we accept the

[25] This is why Kant, in one place, says that what practical reason contributes to an extension of theoretical reason is that 'these concepts are real and really have their (*possible*) objects' (*CpV* 5:134; my emphasis).

postulates, why is belief in them not obligatory? Kant's strategy seems to be to restrict (2) to the 'subjective conditions of our reason' (*CpV* 5:145). As Kant reminds us, there is nothing contradictory in the idea of the highest good (*CpV* 5:144). Hence, realizing the highest good without divine aid, relying merely on the laws of nature, is not *objectively*, but only *subjectively* impossible, 'that is, our reason finds it *impossible for it* to conceive, in the mere course of nature, a connection [between virtue and happiness] so exactly proportioned and so thoroughly purposive between events occurring in the world in accordance with such different laws [i.e. the laws of nature and those of morality], although, as with everything else in nature that is purposive, it nevertheless cannot prove – that is, set forth sufficiently on objective grounds – the impossibility of it in accordance with universal laws of nature' (*CpV* 5:145). So Kant claims that although we cannot prove that it is impossible to realize the highest good according to the laws of nature, we can conceive how it *is* possible to realize the highest good only by appeal to God. In this situation, although we don't have a 'choice' as to whether we want to realize the highest good (*CpV* 5:143), since the latter is morally obligatory, we do have a 'choice' as to the way in which we acknowledge the possibility of the highest good (*CpV* 5:145) – either by restricting ourselves to its mere logical possibility (according to the course of nature) or by being able to conceive of its real possibility (through divine intervention). 'Reason cannot decide this objectively', Kant claims (*CpV* 5:145). Therefore, the 'free interest of pure practical reason decides for the assumption of a wise author of the world' (*CpV* 5:146). Since belief/faith in the existence of God thus does not follow *directly* from the moral law, it is not morally obligatory. Since, on the other hand, practical reason has an interest in a positive conception of the possibility of the highest good, the only rational option with respect to the choice in question is accepting the postulates. And since the basis for this decision is merely subjective, arising from the 'subjective conditions of our reason', the resulting attitude is one of 'pure practical rational faith/belief' (*CpV* 5:146).[26]

[26] One may wonder how this solution coheres with Kant's claim that the only way to resolve the antinomy of pure practical reason is to deny that existence in the sensible world is the only kind of existence possible for rational beings, since only this denial opens the way to allow for an indirect causal relation between virtue and happiness through a divine author of the word (*CpV* 5:115). If it is logically possible to realize the highest good 'according to universal laws of nature', it may seem that there was no antinomy in the first place. A possible response would be that the antinomy concerns not the 'logical', but the 'practical possibility' of the highest good (see *CpV* 5:115, e.g., where Kant says that the highest good is 'practically possible' (*praktisch möglich*)). For the highest good to be practically possible, we must be able to make it an end of our actions, which in turn requires that we can

V. CONCLUSION

In Book Two of the Dialectic of the *Critique of Practical Reason*, Kant offers an argument for the claim that there can be rational belief without evidence. This argument rests only on two substantial premises,[27] namely that the moral law is unconditionally binding for all rational beings and that acceptance of the moral law carries ontological commitments (the 'postulates') the truth of which cannot be decided on purely theoretical grounds. From this it follows that it is rational to accept these ontological commitments even though we don't have any evidence that they are true. Note that this argument is completely independent of the content of the ontological commitments in question. In fact, in the section on the 'Primacy of Pure Practical Reason' Kant mentions neither God nor immortality, but rather puts the argument in completely general terms, speaking only of 'theoretical propositions' and 'positions'[28]. So even someone who is unconvinced by Kant's arguments for the highest good as a necessary end of our will and for God and immortality as necessary preconditions for the highest good being 'practically possible', can accept the argument for the primacy of practical reason.[29] For instance, someone who believes with Kant that we cannot know on purely theoretical grounds that we are free might argue that since freedom is a necessary precondition for being bound by the moral law, it is rational to believe in one's freedom even when lacking theoretical evidence. But there are other candidates for relevant ontological commitments besides freedom: diachronic personal identity, for instance, or the existence of other minds. Both may plausibly be considered to be theoretically undecidable, but practically necessary in the Kantian sense.

However, what Kant's argument does presuppose is his rationalist conception of ethics. If the moral law is not a law of reason that is unconditionally valid for all rational beings, then there is nothing to back Kant's

understand how we might be able to realize that end through our own actions. As Kant argues, the only way in which we can understand how we can realize the highest good is by appeal to God and immortality. A related response, one that refers to Kant's concept of 'real possibility' instead of that of 'practical possibility', is discussed extensively in Paul Guyer, 'From a Practical Point of View: Kant's Conception of a Postulate of Pure Practical Reason', in Paul Guyer, *Kant on Freedom, Law, and Happiness* (Cambridge: Cambridge University Press, 2000), 333–71. See also Eric Watkins's contribution to this volume.

[27] See footnote 21, above.

[28] As Stefano Bacin has pointed out to me, Kant's use of the term 'positions' (*theoretische Positionen*, *CpV* 5:120) coheres nicely with talk of ontological commitments, since for Kant, the original meaning of *Position* (from Latin *positio*) is the *positing* of something. (Cf. 2:73: 'Existence is the absolute position of a thing … The concept of position or *Setzung* is absolutely simple and identical to that of existence'; also see 17:269: 'Existentia est positio absoluta.')

[29] See Gardner, 'Primacy of Practical Reason', 264.

claim that reason cannot reject the postulates in question. If a moral principle commits us to some theoretical belief and we find the belief to lack evidence in its favour, we are rationally free to reject the principle as long as it is not unconditionally binding for all rational beings. However, there remains something to be learned from Kant's argument even for those who don't accept his rationalist conception of ethics.

Recall Wizenmann's objection that 'if it is rational for a thinking human being to presuppose, and to believe in, a God for whose existence he has no reasons merely because there is a need for him to do so in his practical use of reason, then it must also be rational for a lover to presuppose, and to believe in, the love of a creature for the reality of whose love he has no reasons, because this belief answers his need'.[30] Wizenmann's implication clearly is that both kinds of belief are equally irrational. Kant, by contrast, admits the irrationality of the second kind of belief, but denies that of the former.

Now compare this to an observation by William James in his essay 'The Will to Believe' about 'personal relations': '*Do you like me or not?* … Whether you do or not depends, in countless instances, on whether I meet you half-way, am willing to assume that you must like me, and show you trust and expectation. The previous faith on my part in your liking's existence is in such cases what makes the liking come … How many women's hearts are vanquished by the mere sanguine insistence of some men that they must love him!'[31] We may object for ethical and political reasons to the idea of a woman's heart being 'vanquished' by a man who insists on being loved. But the general lesson James draws is surely correct: in 'innumerable cases', the 'desire for a certain kind of truth … brings about that special truth's existence'[32]. Now imagine that Wizenmann's lover knows that this is the case – is it so clear that it would be irrational for him to 'presuppose, and to believe in' the love of his beloved? To be sure, if there is evidence to the contrary, if the beloved has made it clear that she does not love him, his insistence may well be irrational. But if we assume (as we must if the analogy with the postulates is to work) that there is no sufficient[33] evidence either way – that the beloved is friendly to her shy

[30] Wizenmann, 'An den Herrn Professor Kant', 137.

[31] William James, 'The Will to Believe', in W. James, *The Will to Believe and Other Essays in Popular Philosophy* (New York: Longman, Green, & Co., 1897), 1–31, here 23 f.

[32] James, 'The Will to Believe', 24.

[33] Since we are talking about empirical propositions here, we cannot require that there is *no* theoretical evidence whatsoever for or against the proposition in question. Rather, the point is that the evidence is such that, on purely theoretical grounds, it would require us to withhold judgment. That is what is meant by saying that there is no 'sufficient' evidence either way. Note that Kant, even though he claims that speculative reason 'knows nothing' about the postulates (*CpV* 5:120), allows that there are

lover, but has given no indication either that she loves him or that she doesn't – it may seem rationally admissible to believe in that person's love, given that one knows that this belief may well contribute to bring about that love in the first place. The case can be made even stronger if (a) the belief does not concern the present, but the future (the lover believes that he *will* win the heart of his beloved) and if we assume (b) that the lover is convinced (not without some justification, perhaps from reading James) that he will *not* win her heart if he does not believe that she will love him. In this case, one might hold that it would be irrational *not* to believe that she will love him, even if he lacks sufficient evidence for the truth of this belief (provided that there is no sufficient evidence against it). Something similar could be said about many other kinds of cases (such as athletic success or musical performance, for instance) in which, because of the complicated ways of the human psyche, belief in the coming about of some desired fact is necessary for the fact to come about. Thus, rational belief without evidence may seem to be more widespread than Kant assumes.

There may seem to be something odd about these cases, which we can bring to light by asking for what reason the person in question is supposed to hold the beneficial belief. Isn't it irrational to hold a belief simply because it is beneficial to do so, even though there is no evidence in its favour? Here we have to keep in mind Kant's definition of belief/faith as a case of taking something to be true that is 'only subjectively sufficient and at the same time is taken to be objectively insufficient' (*CrV* A822/B850). That it is 'subjectively sufficient' means that we find ourselves convinced of its truth. So the point is not the acquisition of new beliefs, but rather whether it can be rational to continue to believe something one already believes even though one is aware that one cannot produce (sufficient) evidence for it.[34] As Kant makes clear, this kind of attitude is inappropriate with respect to questions of mere theoretical interest. But when it comes to practically relevant questions, we often cannot wait for the evidence to allow for a clear verdict, since finally we will *have* to act one way or another (this is what it means for a question to be practically relevant). The kind of necessity involved here may either be, in Kant's terminology from the Canon, pragmatic or moral. Even if we don't accept that moral principles are unconditionally valid for

purely speculative reasons for assuming God's existence, which reasons make that assumption 'the most reasonable opinion' (*CpV* 5:142). With the terminology adopted here, we can say that even though this opinion is most reasonable, the evidence is not sufficient, so that a full-fledged acceptance is unwarranted. (Remember that opinion is both 'subjectively' and 'objectively' insufficient acceptance.)

[34] William James makes a similar point against Pascal; see James, 'The Will to Believe', 5 f. Also see Guyer, 'From a Practical Point of View', 369.

all rational beings, the moral principles and pragmatic ends one finds oneself to be committed to very often carry ontological commitments we cannot rationally renounce without giving up the principles or ends in question. Not in every such case will it be rational to cling to the principle or end even though their ontological commitments are theoretically undecidable, but in many cases it will be.[35]

In between the Kantian moral law and James's wish to be loved, there is a range of practical attitudes that are not based on pure practical reason in Kant's sense, but are not completely contingent, either. Take an example from epistemology: the belief that I am not a brain in a vat. This belief is practically relevant in so far as all my actions would be pointless if the belief were false. Now it is the very point of sceptical scenarios such as the brain in a vat scenario that we can have no evidence either way. All evidence we may think we have that we are not brains in vats would be available even if we were brains in vats. Hence, the question whether we are brains in vats is theoretically undecidable. But denying that one is a brain in a vat is a condition not just of agency, but of sanity. Even to consider seriously (that is with *practical* consequences) that one might be a brain in a vat is bordering on insanity. Hence, it clearly is *rational* to accept that one is not a brain in a vat (not deluded by an evil demon, and so forth) even in the absence of theoretical evidence (if the latter is understood as something known or rationally believed independently of the belief it is meant to be evidence for). Therefore, it seems safe to go beyond Kant's own claim and to admit that there can be rational belief without evidence even outside the sphere of what is postulated in Kantian morality.[36]

[35] Examples of morally relevant beliefs it may be rational to adopt without sufficient evidence for their truth, even independently from a commitment to a specifically Kantian ethics, might include belief in the possibility of overcoming poverty and belief in a moral psychology that allows for the possibility of moral improvement and just social relations. (I am grateful to Andy Reath for suggesting these examples to me.)

[36] I would like to thank the participants in the preparatory conferences in St Andrews and Leiden, including the contributors to this volume, as well as Claudia Blöser, Thomas Höwing and Steffi Schadow for extremely helpful comments on earlier versions of this chapter. I am particularly grateful to Andy Reath and Sebastian Gardner for written comments on the penultimate version.

CHAPTER 9

The meaning of the Critique of Practical Reason *for moral beings: the Doctrine of Method of Pure Practical Reason*

Stefano Bacin

1. THE ROLE OF KANT'S DOCTRINES OF METHOD

Philosophical works are not always read to the end. The 'Doctrine of Method of Pure Practical Reason', presented as the second main part of the *Critique of Practical Reason*, may be subject to this liability, since it apparently plays no role in the overall argumentative structure of the work. After all the subtle argumentation in the first part of the book, it is difficult to reach page 269 of the original edition and still take into account the possibility that the remaining few pages may add something relevant to the *Critique*. After such intricate sections as those on the fact of reason, the moral motive or the postulates, the statements contained in the Doctrine of Method sound less profound, even conventional, as they apparently provide mere suggestions for moral education in what may appear as an unnecessary appendix. Although the Doctrine of Method is not the most innovative part of the second *Critique*, it is important, at least from an exegetical point of view, to try to comprehend why Kant made it the final part of such an ambitious work and what its actual role may be.[1]

Though it may be tempting to see the Doctrine of Method simply as a product of Kant's excessive – some may even say obsessive – fidelity to a rigid constructive pattern, one has only to recall the several occasions in the *Critique of Practical Reason* where Kant stresses that the subdivisions of the new work correspond only imperfectly to those of the earlier one (see, for

[1] Unfortunately, L. W. Beck did not devote much attention to the Doctrine of Method in his *Commentary on Kant's Critique of Practical Reason* (Chicago and London: University of Chicago Press, 1960), 283 f. Specific analyses are G. F. Munzel's '"Doctrine of Method" and "Closing"', in O. Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft* (Berlin: Akademie Verlag, 2002), 203–17, and M. Cohen-Halimi's '"Sic volo sic jubeo", méthodiquement. Une lecture de la méthodologie de la raison pure pratique', in M. Cohen-Halimi (ed.), *Kant. La rationalité pratique* (Paris: PUF, 2003), 93–118.

example, *CpV* 5:16, 90 f.). He had in fact no intention of reusing the pattern set up in 1781, as if that were the only authorized outline for an exposition of the critical philosophy. He wanted rather to adapt to the new context some concepts which had already allowed him to bestow an appropriate structure on the first *Critique*, including the main division between Doctrine of Elements and Doctrine of Method.

This pair of concepts was not a relic of scholastic terminology, employed in order to follow the dictates of academic exposition; on the contrary, it was one of Kant's attempts to innovate and correct philosophical language, albeit one which was not adopted by later philosophers. Kant distinguishes these two steps with reference to the internal subdivision of logic. According to the distinction which was customary in his times, theoretical logic should supply a theory of the components of reasoning, while practical logic should show how rules must be correctly applied. Kant maintains that such a division was not appropriate to logic – to 'general logic', as he says (see *CrV* A708/B736) – but could be applied to other philosophical disciplines to mark the main distinction between a first part, devoted to developing the core of the analysis, and a second, which should state the conditions whereby the fundamental concepts individuated in the preceding analysis could be properly applied.

That Kant believed he had thus established a useful distinction and (at the same time) a proper connection between the two main steps of a philosophical inquiry – and not only of a critique of reason – is suggested by the fact that he returned to this issue several times and in different contexts. Having made the distinction in the first *Critique*, he not only reused it in the *Critique of Practical Reason*, but carefully considered whether it might also suit his subsequent major works, each time coming to a different decision. The second part of the *Metaphysics of Morals*, the 'Doctrine of Virtue', ends with a Doctrine of Method, whereas the 'Doctrine of Right' has none (see *MdS* 6:411); similarly, Kant considers the possibility of adopting the division again in both parts of the *Critique of Judgment* and in the *Anthropology*, but eventually judges that it is suitable only for the Critique of Teleological Judgment (see *CU* 5:354 ff. and 416 f., and *Anth* 7:400 and 412; see also 15:661).[2]

Therefore, in Kant's terminology, a doctrine of method is by no means a mandatory chapter in a philosophical exposition, which should recall the

[2] On this see J. Kopper, 'Die Bedeutung der Methodenlehren', in N. Fischer (ed.), *Kants Metaphysik und Religionsphilosophie* (Hamburg: Felix Meiner Verlag, 2004), 391–407; C. La Rocca, 'Methodenlehre, transzendentale', in G. Mohr, J. Stolzenberg and M. Willaschek (eds.), *Kant-Lexikon* (Berlin and New York: de Gruyter, forthcoming).

range of themes discussed in the Transcendental Doctrine of Method. Instead, the meaning of the notion is adapted to the specific context in which it is applied. A method, on Kant's general understanding of the term, is the *final* step in any discipline, and supplies instructions on how the discipline's aim is to be achieved. Kant's doctrines of method are always presented as the conclusions of works, not as their beginnings, as one might expect according to the current view, which requires that a method should preliminarily establish a manner of proceeding in an inquiry. On the contrary, a Kantian doctrine of method must explain how the results of the inquiry are to be put into practice.

II. THE FUNCTION OF A DOCTRINE OF METHOD IN THE *CRITIQUE OF PRACTICAL REASON*

The Doctrine of Method of Pure Practical Reason shows how carefully Kant avoids being constrained by the pattern of the first *Critique*. If he had wanted to imitate the earlier work, this is where he would have mapped out the plan of his practical philosophy, and where he would have explained how he envisioned its further developments: he would have mentioned, in the first place, his plan of a metaphysics of morals, to which the *Critique of Practical Reason* does not refer (except indirectly, by reminding the reader that the *Critique* as such is a 'preparatory' work to a 'system': *CpV* 5:161; cf. 5:8).[3] The Transcendental Doctrine of Method discusses the discipline of pure reason and the architecture of the system because its function is to explain how a priori knowledge can be correctly developed, according to the analysis of its basic components in the Transcendental Doctrine of Elements:

> if we look upon the sum of all knowledge of pure speculative reason as an edifice for which we have at least the idea within ourselves, it can be said that in the Transcendental Doctrine of Elements we have made an estimate of the materials, and have determined for what sort of edifice and for what height and strength of building they suffice … At present [that is, in the Transcendental Doctrine of Method], however, we are concerned not so much with the materials as with the plan. (*CrV* A797/B735)

[3] On the contrary, the opening sentence of the *Metaphysics of Morals* (*MdS* 6:205) states (not unambiguously) the connection with the *Critique*. On Kant's idea of a metaphysics of morals see J. Timmermann, *Kant's 'Groundwork of the Metaphysics of Morals'. A Commentary* (Cambridge: Cambridge University Press, 2007), 168 ff., and S. Bacin, 'Sulla genesi della "Metafisica dei costumi" di Kant', *Studi settecenteschi* 25 (2005/6), 253–79.

The topics and the task of the Doctrine of Method must change in the second *Critique*, as this deals with 'materials' of a different nature. These are not elements with which to construct a theoretical edifice: since they are *practical* concepts, it is here of no relevance how a system of practical philosophy should be structured. The opening sentences of the Doctrine of Method of Pure Practical Reason stress precisely this point, namely that the notion of a 'doctrine of method' has been adapted to the present context: the aim, now, is *not* to explain 'a procedure in accordance with principles of reason by which alone the manifold of a cognition can become a system' (*CpV* 5:151). Instead, Kant explains by means of three terminologically different, yet equivalent, formulations how *practical* concepts and principles are to be put into practice. (The care with which this is done also reveals the importance he attributed to this part of the *Critique*.) The Doctrine of Method of Pure Practical Reason should show:

(1) 'the way in which one can provide the laws of pure practical reason with access to the human mind and influence on its maxims' (*CpV* 5:151);[4]
(2) 'the way in which one can make objectively practical reason subjectively practical as well' (*CpV* 5:151; cf. 5:153 and 157);
(3) the way in which it is possible to produce not mere 'legality of actions', but genuine 'morality of dispositions' (*CpV* 5:151).

The first two formulations clearly stress the continuity with the preceding part of the work. By referring to the influence that moral laws should acquire over maxims, Kant shows that this second part has to consider a further, crucial aspect of the relationship between moral laws and maxims, which is one of the major focuses of the work from the very first paragraph. Now that the Analytic has defined the conditions by which a subject's maxim can be valid as a moral law, it is necessary to show how these conditions are to be satisfied in the mind of an individual subject, providing an appropriate ground for his moral choices.

In order to understand the second formulation, it should be remembered that Kant defines as 'objectively practical' both the concept of duty (*CpV* 5:80; *G* 4:427) and the 'practical good' (*G* 4:413), that is, the basic normative concepts expressed by pure reason, which are objective in so far as they are

[4] In a further passage Kant speaks similarly of 'providing the moral law with influence on the human heart' (*CpV* 5:156). In the Doctrine of Method the term 'heart' occurs several times, more often than in the more technical language of the Analytic and the *Groundwork*. Kant may have intended to recapture and rectify the traditional vocabulary of religious or sentimental moralists (of contemporary writers such as Gellert, for example), by taking 'heart' in a precise sense, as 'the subjective first ground of his [a subject's] maxims' (*R* 6:51; cf. 8:825), or 'the *principium* of the moral *Gesinnung*' (27:1422). See also *MdS* 6:441, and 19:127, 181 and 291 f.

independent of any purpose and inclination, and thus universally valid. At the close of the second *Critique*, Kant's goal is to vindicate their *subjective* reality as well. The term 'subjective', then, has in this context, in contrast to most of Kant's ethical works, no negative connotation, as it stands here simply for the dimension of possible finite subjects (see, for example, *CpV* 5:38, 72, 74, 75 f., 81, 88, 117; 11.25).[5] This terminology therefore allows Kant to stress quite clearly that this second part of the work must also be understood as a step towards the solution of the overall problem of the *Critique*, that is, to show how pure reason can be practical (see *CpV* 5:42). The difference between 'subjectively' and 'objectively practical' introduced here is a further specification of that aim, and stresses that the Doctrine of Method belongs to the core of the *Critique*.

The Doctrine of Method, therefore, has the important function of connecting the philosophical inquiry to its outcome in the life of moral subjects. Kant had given no special emphasis to this transition before the *Critique*. One passage of the *Groundwork*, however, seems comparable to the first two formulations of the function of the Doctrine of Method: in order to explain the purpose of the alternative formulas of the categorical imperative, Kant observes that they are equivalent and implicitly contained in the main formula ('the very same law'); 'there is nevertheless a difference among them, which is indeed subjectively rather than objectively practical, intended namely to bring an idea of reason closer to intuition (by a certain analogy) and thereby to the feeling' (*G* 4:436).[6] The three formulas employ the concepts of a universal law of nature, of humanity as an end in itself and of a kingdom of ends, all of which are less abstract than the formal criterion expressed in the main formula. Kant observes, therefore, that 'if, however, one wants also to provide access for the moral law, it is very useful to bring one and the same action under [*durch*] the three concepts mentioned above and thereby, as far as possible, bring it closer to intuition' (*G* 4:437).

However, what in the *Groundwork* is declared to be 'the strict method' is not the procedure based on the auxiliary formulas, but that which evaluates an action on the basis of the main formula (here spelled out as: 'act in accordance with a maxim that can at the same time make itself a universal

[5] See for example *MdS* 6: 480: 'a maxim of virtue consists precisely in the subjective autonomy of each human being's practical reason'.

[6] '*Subjektiv-praktisch*' is a textual emendation introduced by K. Vorländer in his edition of the *Groundwork* (Leipzig: Meiner, 1906; Hamburg: Meiner, 1952) and accepted by B. Kraft and D. Schönecker (Hamburg: Meiner, 1999) and J. Timmermann (Göttingen: Vandenhoeck & Ruprecht, 2004) in their editions. In any case, rejecting the emendation does not seem to affect the interpretation of the passage.

law', *G* 4:436 f.). The three additional formulas can apparently be useful, first of all, in elucidating and making acceptable the meaning of the categorical imperative through expressions which Kant believes to be more easily understandable and closer to the current vocabulary of moral thought; he tries therefore to put forward a more consistent interpretation of concepts which, compromised by preceding philosophical positions, may have a negative influence on human conduct. In spite of certain similarities, the observations of the *Groundwork* must therefore be differentiated from the aim of the Doctrine of Method. Both texts stress the impossibility of communicating or proving the meaning of the moral law through empirical instances, because it is a principle which expresses an idea of reason; unlike the *Groundwork*, however, the Doctrine of Method does not simply aim to make the new formulation of the principle of morality acceptable and to clarify its meaning: the object of the last step of the *Critique* is to show how the moral law can become the basis for a subject's choices.

While moral motivation is not explicitly referred to in this passage of the *Groundwork*, this subjective aspect of pure practical reason is crucial in the Doctrine of Method, as becomes explicit in the third formulation of its function, which, then, adds this important specification to the different perspectives given by the two other formulations. At the conclusion of the work, in order to complete the preceding inquiry, Kant now aims to instruct the reader how to pursue not mere legality, but 'morality of the disposition', which can only take place if 'the immediate representation of the law and the objectively necessary observance of it' become 'the proper incentives [*Triebfedern*] to action' (*CpV* 5:151).[7] In the light of this formulation, the close connection between the Doctrine of Method and the *Triebfeder* chapter of the Analytic becomes clear. For the goal of the method is to make the individual subject aware of his own capacity both to evaluate the moral worth of actions and to determine himself to act morally.

III. PROVING THE 'RECEPTIVITY TO MORALITY'

Although the first formulation of the subject ('the way in which one can *provide the laws* of pure practical reason *with access to the human mind* and *influence on* its maxims') might prima facie suggest that the moral law

[7] See a parallel passage in *CpV* 5:118. The terminological distinction between morality and legality, absent from the *Groundwork*, is introduced in *CpV* 5:71 f. and 81. Both in the Doctrine of Method (*CpV* 5:152) and in the *Triebfeder* chapter (*CpV* 5:72, 83, 85) Kant also mentions the corresponding distinction between the spirit and the letter of the moral law.

should come to the subject, as it were, from the outside, this is of course *not* what Kant means.[8] In contrast, the whole Doctrine of Method is based on the idea that the objective grounds for the determination of the will are accessible to moral subjects as such. In a very brief outline of the text Kant explains that he will first provide a 'proof of the receptivity' to the purely moral *Triebfeder*, and then, on this basis, 'sketch in a few words the method of founding and cultivating genuine moral dispositions' (*CpV* 5:153).

The central topic of the second main part of the *Critique*, indeed, is defined by Kant as 'receptivity' (*Empfänglichkeit*) to morality. The word can be ambiguous, because it means first of all, as it does in the *Critique*, the 'receptivity to a pleasure or displeasure' (*CpV* 5:21; see for instance 5:22 and 58 and *MdS* 6:211), which allows the possibility of being conditioned by an interest in the existence of a desired object. On the contrary, Kant now uses *Empfänglichkeit* in a more positive sense, implicitly distinguishing it from the sense which refers to the bounds of sensibility, in the same way as a 'practical' interest in morality is differentiated from a 'pathological' interest in an object (*G* 4:413 f.). Generally, *Empfänglichkeit* stands for the capacity of the mind to be immediately affected by the proper objects of its faculties.[9] Therefore, by 'a receptivity to a pure moral interest' (*CpV* 5:152), Kant means the natural capacity of a human being to take an immediate interest in morality regardless of any other consideration.

Kant announces that, in order to give proof of 'receptivity' to morality, he will make use of 'observations anyone can make' (*CpV* 5:152). This experimental approach, however, starts from an even lower level than we might expect, because it refers to 'conversation in mixed companies' (*CpV* 5:153). Kant wishes to draw our attention towards a 'propensity of reason' (*CpV* 5:154), which demonstrates that common human intelligence possesses a self-evident ease in tackling and solving moral questions. More interestingly, the same facility is further shown by the fact that even very young people (Kant speaks here of a ten-year-old boy; see *TP* 8:286 and *R* 6:48) 'must necessarily judge so' and easily recognize 'the mark of pure virtue' (*CpV* 5:155), without possessing any profound knowledge of the world, which they cannot yet have acquired. This remark – which inevitably sounds anti-Aristotelian – is one of the several passages where Kant underscores the peculiar ease of moral evaluation in comparison with theoretical

[8] The same phrase occurs e.g. in *CpV* 5:86: 'Duty! Sublime and mighty name that … only holds forth a law that *of itself finds entry into the mind*.'

[9] See *CU* 5:192 and 265, where Kant speaks of a 'susceptibility [*Empfänglichkeit*] to a pleasure from the reflection on the form of things (of nature as well as art)', and of a 'receptivity … to ideas' required by 'the disposition of the mind to the feeling of the sublime'.

cognition (see e.g. *G* 4:409, with *CrV* A783/B811, and *G* 4:259 f.; *CpV* 5:92). While in the *Groundwork* such statements are also meant to stress the difference between morality and the more complicated prudential reasoning (see *G* 4:401 f.), in the Doctrine of Method Kant more explicitly underscores that this capacity involves not only an ease in expressing judgments concerning moral subjects, but also the priority of the moral motive. Unlike receptivity to pleasure in the existence of an object, this peculiar 'receptivity' to moral worth has an active side, as it provides the subject not only with the capacity of recognizing morality in every situation, but also with a motive to act accordingly; it makes the subject not merely capable of appraising the worth of actions, but also alive to the perception of the motive which genuine morality affords at any given time.

In the Doctrine of Method Kant returns to the thesis of respect for the law as the strongest motive, which he had already expounded in the Analytic (see *CpV* 5:73 ff.): 'receptivity to pure moral interest' as a 'property of our minds' has as its counterpart the fact that the 'pure representation of virtue … is the most powerful incentive to moral good' (*CpV* 5:152), which remains clear and evident in that it is separated from all other motives. This point now becomes the main focus of Kant's observations in the Doctrine of Method,[10] and is here justified by appealing to what everyone may experience, namely the fact that respect for the law is the only motive that 'teaches the human being to feel his own dignity' (*CpV* 5:152; cf. 5:88) because it shows him that he is able to cognize it and adopt it as the ground of his choices. While any other motive is necessarily limited to the goal defined by the agent's desires and purposes, the representation of the moral law, with its subjective aspect, namely respect for the law, makes the subject aware of his own status as a moral being. While other motives 'can give a person no moral worth and not even confidence in himself, without which the consciousness of one's moral disposition and of a character of this kind … cannot come to exist' (*CpV* 5:157), the peculiar simplicity of moral questions must lead everyone to recognize the 'receptivity' to morality that they possess as moral beings, and to see it as the ground of their dignity.

[10] A sketchy formulation of the same thoughts on which the Doctrine of Method is based can be found in a fragment from the *Nachlaß*: 'The motivating force of the moral concept lies in its purity and its difference from all other impulses. The original *intellectuale* concept is striking only insofar as it is compared to other analogical motivating grounds of honor, happiness, mutual love, and peace of mind and elevated above all others in the comparison. Encomia of virtue and admonitions can be of no value, rather only the development of its concept is. Examples that illuminate the purity of the concept of virtue and an immediate moral aversion are better for education', 19:200, in Kant, *Notes and Fragments*, trans. Paul Guyer, Curtins Bowman and Frederick Rauscher (Cambridge and New York: Cambridge University Press, 2005), 449.

The peculiar 'receptivity' to morality to which the Doctrine of Method of Pure Practical Reason refers is better understood in connection with Kant's observations in other works. In *Religion* (the first two essays of which are closely related to the topics and the positions of the Doctrine of Method) he discusses what he calls 'an optimistic presupposition on the part of the moralists', that is, that man is good according to his nature (*R* 6:20). He does not reject the content of the presupposition, but merely that the relation of human beings to moral goodness can be conceived of as a natural property, whereas it should rather be interpreted as an original predisposition. Every moral demand presupposes what he calls the 'predisposition to personality', defined through exactly the same term used in the Doctrine of Method: 'the susceptibility [*Empfänglichkeit*] to respect for the moral law' (*R* 6:27).[11] When in the 'Doctrine of Virtue' Kant describes further features which are distinctive to human beings as moral subjects, he again uses the same word, to specify the 'subjective conditions of receptiveness [*Empfänglichkeit*] to the concept of duty' (*MdS* 6:399). They are 'natural predispositions of the mind (*praedispositio*) for being affected by concepts of duty', which a man cannot lack entirely, since they make him 'morally alive' (see *MdS* 6:400).

By this multi-faceted 'receptivity' to morality – which can be examined at different levels – Kant means the basic condition(s) whereby morality is possible in every subject. Because these conditions are a predisposition, they cannot be lost, but ought to be recognized and developed.[12] Therefore Kant claims that his method can show 'the reality of such a feeling [for the moral good] but not any improvement brought about by it' (*CpV* 5:153). The Doctrine of Method does not show the way to 'moral progress' (the very notion is inappropriate, according to Kant: see *CpV* 5:84 and *MdS* 6:409), but tries to ensure the proper starting point for the virtuous effort which should bring about what Kant calls a 'progress *in* goodness' (*CpV* 5:157). On a number of occasions Kant contrasts Rousseau's and Lord Kames's views, always stating his agreement with the latter on the point that virtue must be 'learned', since it is not received by nature and has to be acquired.[13] Now, the 'receptivity' to morality which every subject possesses is the

[11] In *Religion*, *Empfänglichkeit* also has a slightly different meaning, related to man's worthiness to receive the grace of God (see *R* 6:75, 115, 145, 170).

[12] See e.g. 9:446: 'It is the human being himself who is supposed first to develop his predispositions toward the good. Providence has not placed them already finished in him; they are mere predispositions and without the distinction of morality. The human being shall make himself better, cultivate himself, and, if he is evil, bring forth morality in himself.'

[13] See 25:188, 387, 1107, 1334; 27:248–9, 1401 f.; 29:603 (in the last two passages 'Hume' must be corrected to 'Home').

fundamental condition by which it is possible to 'learn virtue' in this sense of the term.[14] What Kant calls the 'revolution in the disposition' (*R* 6:47), which requires the subject to restore the correct order of his motives, is the basic move by which pure reason becomes subjectively practical.[15] The edifice that this Doctrine of Method should help us to build is an individual moral consciousness, grounded on the awareness of our humanity.

The transition that the Doctrine of Method seeks to accomplish – from moral law to maxims, from objectively practical to subjectively practical pure reason – requires that the subject becomes aware of the intrinsic connection between the humanity presented through the categorical imperative and his own individuality as a subject who can (and ought to) make choices on the basis of this law. One crucial passage clarifies that, if the fundamental moral concepts

> are to become subjectively practical, they must not stop short with objective laws of morality, to admire and esteem them with reference to humanity: the representation of them must be considered in relation to the human being and his individuality. (*CpV* 5:157)

Of course, this is not intended to repeal the moral subordination of the subject's individuality to the observance of the moral law (Kant mentions the necessary 'self-denial' again a few lines later: *CpV* 5:158). Rather, each moral being should recognize the humanity expressed by the moral law as his 'proper self' (to borrow the phrase from *G* 4:457).[16] The *Critique* requires, at its close, a 'method' which shows that the individual subject can and should recognize himself as a being capable of moral judgment and moral action, since only this awareness can be the proper ground of a person's moral life (see *CpV* 5:88 and *MdS* 6:441). The Doctrine of

[14] According to Kant the first of the two possible meanings that can be given to the question 'Whether virtue can be taught?' is precisely: 'whether without moral sentiment, that is *determinability of the will by a practical rule of reason*, someone would *understand* what virtue is? Answer: No' ['Ob tugend könne gelernt werden? Kan in zwiefacher Bedeutung genommen werden. 1. ob ohne alles moralische Gefühl, d. i. Bestimbarkeit des willens durch practische Vernunftregel, jemand, was tugend sey, verstehen würde? Antwort: – Nein'] (19:266 f.). On the same question see *R* 6:24 and *MdS* 6:477.

[15] See *R* 6:48: 'if by a single and unalterable decision a human being reverses the supreme ground of his maxims by which he was an evil human being (and thereby puts on a "new man"), he is to this extent, by principle and attitude of mind, *a subject receptive to the good*'.

[16] Note that the word 'humanity' occurs in only a few further passages in the *Critique*: see 5:85, 87, 88 and 137.

Method of Pure Practical Reason differs, then, from the 'Ethical Doctrine of Method' which concludes the 'Doctrine of Virtue' (*MdS* 6:477–85), precisely because the former is focused on the status of moral beings, its basic conditions and the necessity to secure them in the consciousness of the individual, while the latter deals with specific moral demands and not only with the subjective conditions of their authority.[17]

Thus, the Doctrine of Method is not meant as a subsidiary addition to an inquiry which is in itself already complete and should be now tailored to a somehow restricted field of application; on the contrary, the Doctrine of Method is precisely intended to give completion to the project of the *Critique*. Like the work as a whole, the Doctrine of Method does not indeed refer to any specific feature of human beings, but directly addresses finite (rational and sensible) moral beings.[18] The only relevant feature, here, is their dignity and their capacity to become aware of it.

Therefore, the role these final observations play in the overall argumentative structure of the *Critique* is anything but irrelevant. In the *Groundwork* Kant explicitly states that 'the practical use of common human reason confirms the correctness' of the argumentation, which concludes with the statement of the possibility of the categorical imperative (*G* 4:454). Now, the Doctrine of Method seems to develop the same basic idea: the practical use of common reason has to be taken as the most significant confirmation of the results reached in the first main part of the *Critique*, by showing an actual connection between them and the experience of each moral subject. The Doctrine of Method gives instructions for 'observations that anyone can make', which should '*prove* [*beweisen*]' (*CpV* 5:153)[19] both our status as moral beings and the essential features of morality assessed in the Doctrine of Elements. If its results could not be verified in the individual experience of each subject, they would have no moral relevance, and philosophical ethics would be mere theory, lacking any intrinsic relation to practice. A moral theory which could not satisfy this demand would be useless, like those 'physicians and jurists' (*TP* 8:275) who are not able to apply their

[17] On the 'Ethical Doctrine of Method' see C. Piché, 'La méthodologie éthique de Kant', in S. Goyard-Fabre and J. Ferrari (eds.), *L'année 1797. La métaphysique des moeurs* (Paris: Vrin, 2000), 109–22, and S. Bacin, *Il senso dell'etica. Kant e la costruzione di una teoria morale* (Bologna: Il Mulino, 2006), 246 ff.

[18] The Doctrine of Method is interpreted as a step in the application of morality to human beings by G. F. Munzel, *Kant's Conception of Moral Character. The 'Critical' Link of Morality, Anthropology, and Reflective Judgment* (Chicago and London: University of Chicago Press, 1999), 307 ff., and R. B. Louden, *Kant's Impure Ethics. From Rational Beings to Human Beings* (New York and Oxford: Oxford University Press, 2000), 48 ff. A reading of the Doctrine of Method from a pedagogical point of view is L. Koch's *Kants ethische Didaktik* (Würzburg: Ergon, 2003).

[19] The Cambridge Edition has a weaker translation of *beweisen*: 'show'.

knowledge to concrete situations and cannot 'give an expert opinion' when faced with real problems, thus failing to perform their function. The main difference between those other disciplines and moral theory (or 'a theory based on the concept of duty', *TP* 8:276) is that moral theory should not justify its claims through a confrontation with actual circumstances, but has to refer to that peculiar and individual (although not 'private') 'experience that can be only inward' (*TP* 8:287), that experience whose guidelines it is the aim of the Doctrine of Method to explain.

The relevance of this second main part of the *Critique* is also highlighted by repeated polemical hints, for Kant emphasizes that the positions commonly adopted in the exposition of morals are inadequate, or even harmful. These are, in his view, the most insidious consequences of 'bad moral philosophy', which can only confuse common 'receptivity' to morality.[20] Avoiding such confusion is one of the primary tasks of a moral theory, as is also clearly stated in the essay *On the Common Saying*, the first section of which is closely connected with the Doctrine of Method.[21] Therefore, Kant emphasizes that his own proposal is entirely new (see *CpV* 5:153), and 'more necessary than ever' (*CpV* 5:157). The observations of the Doctrine of Method, then, are meant to play the strategic role of showing how moral theory should accomplish its proper guiding function by also pointing out the negative implications of erroneous philosophical positions. While philosophical ethics should not invent new principles of morality, it should pay special attention to the proper way of clarifying moral distinctions, which common reason is already able to apply.

If, in stimulating this moral self-awareness, the Doctrine of Method returns to topics already discussed in the Doctrine of Elements, apparently without adding anything new, this is not because it should supply a

[20] See *CpV* 5:155: 'only philosophers can make the decision of this question doubtful', that is, 'what … really is pure morality, by which as a touchstone one must test the content of every action'.

[21] See *TP* 8:288: 'if this attention were drawn to it more often and he became used to ridding virtue completely of all the rich booty of advantages to be amassed through the observance of duty and to representing it in all its purity; if it became a principle of private and public instruction always to make use of this (a method of inculcating duties that has almost always been neglected), human morality would soon be better off. That historical experience up to now has still not proved the success of the doctrine of virtue may well be the fault of just the false presupposition that the incentive derived from the idea of duty in itself is much too fine for the common concept whereas the coarser incentive drawn from certain advantages to be expected, in this world or even in a future one, from compliance with the law (without regard for the law itself as the incentive) would work more powerfully on the mind, and that up to now it has been made a principle of education and homiletics to give preference to the aspiration for happiness over that which reason makes the supreme condition of this, namely worthiness to be happy.' See also H. F. Klemme, 'Beobachtungen zur Kantischen Vermittlung von Theorie und Praxis in der praktischen Philosophie', in H. Robinson (ed.), *Proceedings of the Eighth International Kant Congress*, vol. II (Milwaukee: Marquette University Press, 1995), 521–31.

condensed popular version of the main contents of the work: this would not be suitable to the project of a critique of reason.[22] There is no substantial addition to the points made in the first main part of the work because the main features of moral agency, which have already been examined, must now be factually confirmed in the individual's experience of the practical use of pure reason.[23] No further thesis or theorem need be discussed, defended or elucidated. The reader must now move on to the exercises.

IV. EXERCISES AND EXAMPLES

When Kant comes explicitly to the method (see *CpV* 5:159), he sketches out some guidelines, which are divided under two main headings. These instructions sum up the preceding observations and give them a precise order:

(1) 'at first it is only a question of making appraisal of actions by moral laws a natural occupation, as it were, a habit accompanying all our free actions as well as our observations of those of others, and of sharpening it' (*CpV* 5:159);

(2) the object of 'the second exercise' is 'to draw attention, in the lively presentation of the moral disposition in examples, to the purity of will' (*CpV* 5:160).

(1) The first exercise requires the reader to pay attention to the morally salient features of every action: one has to consider both 'whether the action objectively conforms to the moral law, and with which one' (*CpV* 5:159), and 'whether the action was also done (subjectively) for the sake of moral law' (*CpV* 5:159). This attention to the morally salient features of an action has to become, 'as it were, a habit' (*CpV* 5:159; cf. 5:154), in order to consolidate the already mentioned 'propensity of reason'; this is in fact the only way to strengthen the power of judgment, which 'cannot be taught but only practised' (*CrV* A133/B172). Furthermore, Kant writes, '*as it were*,

[22] The Doctrine of Method could be seen, instead, as the 'popular' part of the *Critique* according to W. Kersting, 'Kann die "Kritik der praktischen Vernunft" populär sein? Über Kants Moralphilosophie und pragmatische Anthropologie', *Studia Leibnitiana* 25 (1983), 82–93. Furthermore, it should be remarked that the 'popular moral philosophy' addressed in the *Groundwork* is one of the main objectives of the *Critique*, too. As Kant's remarks on examples and the hints to the connection between doctrine of wisdom and science reveal (see *CpV* 5:155 f. and 163, discussed below), the Doctrine of Method is intended to show how a critique of reason can reach the results which the popular philosophy aims at, but cannot achieve.

[23] On the role of the common practical use of pure reason see Jens Timmermann's chapter in this volume.

a habit' because it is crucial that such an activity should become regular but not mechanical, and that it should remain voluntary and conscious.[24]

First, the subject has to consider the rightness of the action and assess whether its description conflicts with any moral norm or properly expresses a moral norm, thus proving that it is easily possible to recognize 'the least deviation' from the moral law (*CpV* 5:154). Kant then adds, somewhat unexpectedly, that we have indeed to consider 'which law' the action conforms to, and clarifies this statement as follows:

> by this, attention to such law as provides merely a ground of obligation [*einen Grund zur Verbindlichkeit*] is distinguished from that which is in fact obligatory (*leges obligandi a legibus obligantibus*) (e.g., the law of what the need of human beings requires of me as contrasted with what their right requires, the latter of which prescribes essential duties whereas the former prescribes only nonessential duties), and thus one teaches how to distinguish different duties coming together in an action. (*CpV* 5:159)

The so-called clash of duties, briefly discussed in the *Metaphysics of Morals* and in the lectures (see 6:409; 27:537, 1410), is here brought to the forefront not as an issue concerning the modal and normative consistency of a moral theory, but as a basic feature in the practice of the sharpness of common moral judgment which Kant recalled before. In order to judge our own actions or those of someone else, we must grasp the relation between the deed and the moral norms; but the same action can be justified or prescribed by different norms, which thus 'come together' in a single case. The evaluation must then assess which norm is more relevant and provides the stronger justification. Kant explains this discrimination by means of technical notions (namely those of 'ground of obligation' and essential or non-essential duties)[25] which he does not feel the need to define here, since the function of the Doctrine of Method is not to supply further conceptual clarifications, but to explain how to verify that the common moral understanding is per se able to apply such distinctions, even though he does not have a precise notion of them. So, instead of a definition or of a

[24] To stress this point, Kant frequently differentiates *Gewohnheit* and *Angewohnheit*, i.e. a habit in a weaker sense, which belongs to a conscious practice, and a routine based on a fixed pattern. His conception is well summarized in a fragment: 'Die Gewohnheit macht alles leicht (und starke Empfindung unmerklich). Die Angewohnheit macht alles nothwendig oder das Gegentheil schwer … Angewohnheit ist niemals, selbst nicht in guten Handlungen, vollkommen zu billigen' (15:98). See 7:148 f.; 6:383 f., 407; 9:463, 480.

[25] On the notion of *ratio* (or *motivum*) *obligandi* see also 6:224; 27:508, 537, 1410. On the topic see O. O'Neill, 'Instituting Principles: Between Duty and Action', in M. Timmons (ed.), *Kant's 'Metaphysics of Morals'. Interpretative Essays* (Oxford: Clarendon Press, 2002), 331–47. (I thank J. Timmermann for making available to me a very useful unpublished paper of his on this topic.)

philosophical discussion of the issues the alleged clash of duties raises (which are to be discussed elsewhere, in the *Metaphysics of Morals*), here, Kant mentions only two possible grounds which can make an action dutiful: the right and the need of human beings (see the distinction also in *CU* 5:273 and 7:87). In this way, he intends to point to the most evident distinction between moral considerations which are both valid grounds of duties, but apply with different strictness. For, whereas the happiness of others is a dutiful end which can be pursued and determined in many different ways, the 'right of human beings' requires directly and strictly to respect human dignity before any further moral consideration is taken into account, since the right of human beings is 'what is most sacred of all among human beings' (*MdS* 6:304; cf. 8:352; 19:308). Thus Kant not only refers to a sharp conceptual distinction which should in fact be quite easy to apply by the common moral judgment, but also, at the same time, draws the subject's attention to the fact that 'essential duties' are justified precisely by the consideration of the humanity which the subject should become aware of through these exercises.

The other aspect which deserves attention in the exercise of action evaluation is the motivation underlying the action ('whether … it has not only moral correctness as a deed but also moral worth as a disposition by its maxim'). But this seems to raise difficulties too, since it is impossible, in Kant's view, even for the agent himself to ascertain the 'incentive' of an action, because it remains opaque and empirically inaccessible (see *R* 6:20; *MdS* 6:447; *TP* 8:284; *G* 4:407). Why, then, does Kant suggest an exercise which, apparently, cannot be solved? The aim, indeed, is not to make a correct judgment about an action, but rather to practice and strengthen the natural capacity to appraise an action according to moral principles; in focusing his attention properly, the subject will become aware, first of all, that the relevant moral features of any action are its correctness and the disposition with which it is done.

The expected outcome of this exercise in evaluating actions is described in terms which anticipate the *Critique of Judgment* (whose project is closely connected with that of the second *Critique*, and was carried out immediately afterwards).[26] Kant remarks that the special attention to salient moral features of actions makes us 'feel our entire cognitive faculty (understanding and imagination) strengthened' (*CpV* 5:160). Considering actions from a moral perspective, as this first exercise requires, or as we do in conversation, is pleasant since it is in harmony with the basic structure of practical reason,

[26] See Heiner Klemme's chapter in this volume.

thus revealing its fundamental 'propensity' (see *CpV* 5:153 f.). In this manner, the virtuous action takes on 'a form of beauty' (*CpV* 5:160), since the judgment which is passed on it expresses the implicit conceptual structures of the 'reason which judges merely about the practical' (*CpV* 5:159), and the subject perceives that his faculties match their proper object.[27] The relationship with the positions of the third *Critique* becomes even more evident as Kant compares the effects of the practice of moral judgment recommended by him with the emotional aspect of inquiry in the natural sciences: the biologist feels an affection for the objects of his observations by which he may discover some inner purpose. Both in the practical use of pure reason and in the inquiry guided by judgment 'we finally come to like [*wir gewinnen … lieb*] something the contemplation of which lets us feel a more extended use of our cognitive powers', and this 'is especially furthered by that in which we find moral correctness' (*CpV* 5:160, and also later on the same page). Thus, through this exercise in action evaluation, a first step of the subject's fundamental moral self-knowledge is to be accomplished, which focuses on his capacity to apply moral distinctions.

(**2**) However, the contemplation of regularity in the correct moral evaluation of actions does not involve as yet a genuine interest in bringing about moral worth (see *CpV* 5:160). In the first exercise morality can still be viewed merely as an object which is well suited to our mental capacities, a matter which demonstrates its potentiality, but does not necessarily appear as valuable per se. If we missed this distinction, we should have to think that every conversation about morality immediately provides us with the pure motivation to act for the sake of the moral law, and this is certainly not the case. The object of the Doctrine of Method is to enable us to achieve a proper interest in moral worth, which in its turn provides the impulse towards its accomplishment, since 'an interest is that by which reason becomes practical' (*G* 4:460), namely subjectively practical, too.[28]

In addition to an accurate evaluation of actions, it is therefore necessary 'to draw attention … to the purity of will',[29] and Kant holds that this should happen 'in the lively presentation of the moral disposition in examples' (*CpV* 5:160). After the strict limitation of their role in morals in the

[27] These observations should not be confused with Kant's view of the idea of beauty as a symbol of morality, expounded in the *Critique of Judgment*. While there Kant underscores the analogies which connect the appraisal of the beautiful with moral worth (see *CU* 5:353 ff.), here he speaks of a purposiveness perceived *in* moral evaluation.

[28] On the notion of 'interest' in the second *Critique* see *CpV* 5:79 f.

[29] See *CpV* 5:155: 'the mark by which pure virtue is tested'.

Groundwork, the remarks in the Doctrine of Method, and the analogous ones in the *Religion* (*R* 6:61 ff.) and in the 'Doctrine of Virtue' (*MdS* 6:479 f.; see also *CU* 5:283), may appear as at least a partial retraction.[30] Kant's position is, however, coherent. Examples cannot be accepted in the foundations of morality because they do not supply a criterion for moral judgment nor a rule for practical reason; indeed, they merely show that such a criterion is needed (see *MdS* 6:355): 'every example of it [that is, of morality] represented to me must itself first be appraised in accordance with principles of morality' (*G* 4:408) to assess whether it can be taken as an instance of them. In fact, if we were first to examine examples, we would gather not so much models of morality as evidence of the evil of human beings (see *R* 6:20; *MdS* 6:409). Nevertheless, precisely because examples imply a reference to a criterion by which morality can be judged, they can help to show that 'the required prototype always resides only in reason' (*R* 6:63), and can be of use in making it explicit.

So, while no moral models can be presented, a proper understanding of examples can provide occasions for the subject to reflect on his own capacity to deal with moral principles. Since they should not be imitated, they must neither involve an emotional participation on the part of the subject, nor emphatically present heroic deeds as prototypes of alleged moral strength. Sentiments cause an apparently powerful, but transitory reaction in the subject, and depend on the situation which gives rise to them (see *CpV* 5:156); therefore, they do not supply a reliable ground for moral attitude.[31] Presenting 'so-called noble (supermeritorious) actions' (*CpV* 5:155; cf. 5:157) proves to be an even worse strategy, since such acts can only confuse one's consciousness of moral distinctions, because they present acts in conflict with ordinary morality and with fundamental duties, specifically with duties towards oneself (see *CpV* 5:158).[32] Kant rejects not only those attempts to motivate the subject to practise virtue through examples which stimulate

[30] On Kant's view of examples in morals see O. O'Neill, *Constructions of Reason. Explorations of Kant's Practical Philosophy* (Cambridge: Cambridge University Press, 1989), 165–86; A. Ferrarin, *Saggezza, immaginazione, giudizio pratico. Studio su Aristotele e Kant* (Pisa: ETS, 2004), 135–48; I. Heidemann, 'Die Funktion des Beispieles in der kritischen Philosophie', in F. Kaulbach and J. Ritter (eds.), *Kritik und Metaphysik. Studien. Heinz Heimsoeth zum 80. Geburtstag* (Berlin: de Gruyter, 1966), 21–39; and R. B. Louden, 'Go-Carts of Judgment: Exemplars in Kantian Moral Education', *Archiv für Geschichte der Philosophie* 74 (1992), 303–22. Some information on the historical context can be found in G. Buck, 'Beispiel, Exempel, exemplarisch', in J. Ritter and K. Gründer (eds.), *Historisches Wörterbuch der Philosophie*, vol. I (Basel: Schwabe, 1971), 818–23.

[31] On these considerations are based also Kant's restrictive remarks on the relevance of literature for morals: see e.g. *CU* 5:273; 25:1213.

[32] See also the casuistical questions in *MdS* 6:423 and 27:629. On the topic see M. Baron, *Kantian Ethics Almost without Apology* (Ithaca: Cornell University Press, 1995), 36 ff.

compassion or 'elevation of the heart', but also the common practice of emphasizing moral worth through the impact of alleged heroic acts, because both strategies only lead to partial perspectives, which merely express the (non-moral) qualities of a single individual: in other words, it is impossible for a subject to recognize on such grounds the connection between his own individuality and a shared 'humanity', so that the necessary moral self-awareness, at which the Doctrine of Method aims, cannot be reached.

To clarify the proper function of examples which illustrate morality, Kant suggests that one should tell 'the story of an honest man whom someone wants to induce to join the calumniators of an innocent but otherwise powerless person (say, Anne Boleyn, accused by Henry VIII of England)' (*CpV* 5:155), who withstands every temptation and maintains strict moral conduct. Everything in the story is intended to raise in the subject first 'approval and applause', then 'admiration' and finally 'the greatest veneration and a lively wish that he himself could be' such a good person (*CpV* 5:156). In an example of 'purity of will' one can recognize and at the same time appreciate an expression of the moral principle: the awareness of it goes hand in hand with the genuine interest in morality which such consciousness entails. The examples to be used in morals must be intended primarily as 'examples of reason judging morally' (*CpV* 5:163). According to Kant, reflection on such cases rouses in the subject the 'consciousness of his freedom' (*CpV* 5:160), as it makes him feel both the capacity and the motivation to determine his will on the basis not of inclinations but of principles. The attention paid to instances of morality makes each of us feel, with an apparent paradox, 'respect for *ourselves*' (*CpV* 5:161), which is precisely what the Doctrine of Method – and the whole *Critique* – sets out to achieve: recognizing our status as moral subjects constitutes the ground for a moral disposition, that is, for a conscious moral life. This positive consciousness of the humanity within ourselves, then, affords the only genuine access to the moral law within our own minds, and at the same time confirms the authority of the moral law through what may be seen as experimental proof.

Finally, the perspective on the *Critique* given in the Doctrine of Method is very close to that of the Conclusion. Besides giving the *Critique* a rhetorically appropriate ending, the Conclusion focuses on its status and aim in a way which appears quite similar to the observations examined so far. By defining 'science' as 'the narrow gate that leads to the doctrine of wisdom' (*CpV* 5:163; cf. 5:108), Kant emphasizes that the complex inquiry of the *Critique* achieves its proper meaning only if we are aware that the work demands a development, even a confirmation, which is not to be attained

through mere theory. This statement has, in Kant's intentions, a primarily polemic meaning (as is made clear by the final assertion that 'the public need take no interest in its subtle investigations'): in contrast with the popular philosophy of his time, Kant stresses that a deep philosophical analysis is unavoidable if ethics is to be laid on a profitable foundation. Nevertheless, the connection between the 'science' pursued in the *Critique*, and the 'doctrine of wisdom' which philosophy aims at, expresses the interdependence between the analyses of the Doctrine of Elements and the completion of these analyses suggested in the Doctrine of Method, which is where Kant tries 'to prepare well and clearly the path of wisdom which everyone should travel, and to secure others against taking the wrong way' (*CpV* 5:163). The function of the Doctrine of the Method, then, proves to belong to the very core of the project of the *Critique*, and provides an important key to the correct understanding of the work. The fact that, after a demanding and abstract inquiry, Kant proposes exercises which should confirm its results reveals that in his view the whole work is meant to strengthen the common moral consciousness of every subject, who, 'as a man … never outgrows the school of wisdom' (*TP* 8:288).

Bibliography

Abel, Jacob Friedrich. *Versuch über die Natur der speculativen Vernunft zur Prüfung des Kantischen Systems* (Frankfurt and Leipzig, 1787; reprinted Brussels: Culture et civilisation, 1968).

Albrecht, Michael. *Kants Antinomie der praktischen Vernunft* (Hildesheim: G. Olms Verlag, 1978).

Allison, Henry E. *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1996).

'Kant on Freedom: A Reply to My Critics', in *Idealism and Freedom*, 109–28.

Kant's Theory of Freedom (Cambridge and New York: Cambridge University Press, 1990).

Kant's Transcendental Idealism (New Haven: Yale University Press, 1983).

Ameriks, Karl. *Interpreting Kant's Critiques* (Oxford: Clarendon Press, 2003).

'Pure Reason of Itself Alone Suffices to Determine the Will', in Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft*, 99–114.

'The Critique of Metaphysics: The Structure and Fate of Kant's Dialectic', in Paul Guyer (ed.), *The Cambridge Companion to Kant and Modern Philosophy* (Cambridge and New York: Cambridge University Press, 2006), 269–302.

Bacin, Stefano. *Il senso dell'etica. Kant e la costruzione di una teoria morale* (Bologna: Il Mulino, 2006).

'Sulla genesi della "Metafisica dei costumi" di Kant', *Studi settecenteschi* 25 (2005/6), 253–79.

Baron, Marcia. *Kantian Ethics Almost without Apology* (Ithaca: Cornell University Press, 1995).

Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960, 1966). German translation: *Kants 'Kritik der praktischen Vernunft'. Ein Kommentar* (Munich: W. Fink, 1974).

'A Prussian Hume and a Scottish Kant', in *Essays on Kant and Hume* (New Haven: Yale University Press, 1978), 111–29.

Brandt, Reinhard. 'Der Zirkel im dritten Abschnitt von Kants *Grundlegung zur Metaphysik der Sitten*', in H. Oberer and G. Seel (eds.), *Kant. Analysen – Probleme – Kritik* (Würzburg: Königshausen & Neumann, 1988), 169–91.

Die Bestimmung des Menschen bei Kant (Hamburg: Felix Meiner Verlag, 2007).

Buck, Günther. 'Beispiel, Exempel, exemplarisch', in J. Ritter and K. Gründer (eds.), *Historisches Wörterbuch der Philosophie*, vol. I (Basel: Schwabe, 1971), 818–23.

Chignell, Andrew. 'Belief in Kant', *Philosophical Review* 116 (2007), 323–60.

Cohen-Halimi, Michèle. '"Sic volo sic jubeo", méthodiquement. Une lecture de la méthodologie de la raison pure pratique', in M. Cohen-Halimi (ed.), *Kant. La rationalité pratique* (Paris: PUF, 2003), 93–118.

Düsing, Klaus. 'Das Problem des höchsten Gutes in Kants praktischer Philosophie', *Kant-Studien* 62 (1971), 5–42.

Engstrom, Stephen. 'Introduction', in Kant, *Critique of Practical Reason*, trans. Werner Pluhar (Indianapolis: Hackett Publishing Company, 2002), xv–liv.

'The Concept of the Highest Good in Kant's Moral Theory', *Philosophy and Phenomenological Research* 52 (1992), 747–80.

The Form of Practical Knowledge (Cambridge, MA: Harvard University Press, 2009).

Ferrarin, Alfredo. *Saggezza, immaginazione, giudizio pratico. Studio su Aristotele e Kant* (Pisa: ETS, 2004).

Förster, Eckart. 'Die Dialektik der reinen praktischen Vernunft (107–121)', in Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft*, 173–86.

'Kant und Strawson über ästhetische Urteile', in J. Stolzenberg (ed.), *Kant in der Gegenwart* (Berlin and New York: de Gruyter, 2007), 269–89.

Kant's Final Synthesis. An Essay on the Opus postumum (Cambridge, MA: Harvard University Press, 2000).

(ed.). *Kant's Transcendental Deductions* (Stanford: Stanford University Press, 1989).

Franks, Paul W. *All or Nothing. Systematicity, Transcendental Arguments, and Skepticism in German Idealism* (Cambridge, MA: Harvard University Press, 2005).

Gardner, Sebastian. 'The Primacy of Practical Reason', in Graham Bird (ed.), *A Companion to Kant* (Oxford: Blackwell, 2006), 259–73.

Gesang, Bernward (ed.). *Kants vergessener Rezensent. Die Kritik der theoretischen und praktischen Philosophie Kants in fünf Rezensionen von Hermann Andreas Pistorius* (Hamburg: Felix Meiner Verlag, 2007).

Giordanetti, Piero. 'Kants Entdeckung der Apriorität des Geschmacksurteils. Zur Genese der Kritik der Urteilskraft', in H. F. Klemme *et al.* (eds.), *Aufklärung und Interpretation* (Würzburg: Königshausen & Neumann, 1999), 171–96.

Greenberg, Sean. 'From Canon to Dialectic to Antinomy: Giving Inclinations Their Due', *Inquiry* 48 (2005), 232–48.

Grimm, Jacob and Grimm, Wilhelm. *Deutsches Wörterbuch*, 16 vols. (Leipzig: Hirzel, 1854–1960).

Guyer, Paul. 'From a Practical Point of View: Kant's Conception of a Postulate of Pure Practical Reason', in *Kant on Freedom, Law, and Happiness* (Cambridge: Cambridge University Press, 2000), 333–71.

'Naturalistic and Transcendental Moments in Kant's Moral Philosophy', *Inquiry* 50 (2007), 444–64.

Hamann, Johann Georg. *Briefwechsel*, vol. VI, ed. A. Henkel (Frankfurt am Main: Insel-Verlag, 1975).

Hatfield, Gary. 'Empirical, Rational, and Transcendental Psychology: Psychology as Science and as Philosophy', in P. Guyer (ed.), *The Cambridge Companion to Kant* (Cambridge and New York: Cambridge University Press, 1992), 200–27.

Heidemann, Ingeborg. 'Die Funktion des Beispieles in der kritischen Philosophie', in Friedrich Kaulbach and Joachim Ritter (eds.), *Kritik und Metaphysik. Studien. Heinz Heimsoeth zum 80. Geburtstag* (Berlin: de Gruyter, 1966), 21–39.

Henrich, Dieter. 'Der Begriff der sittlichen Einsicht und Kants Lehre vom Faktum der Vernunft', in Dieter Henrich (ed.), *Die Gegenwart der Griechen im neueren Denken: Festschrift für Hans-Georg Gadamer zum 60. Geburtstag* (Tübingen: JCB Mohr, 1960), 77–115. Translated as 'The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason', trans. Manfred Kuehn, in Henrich, *The Unity of Reason*, 55–88.

'Ethik der Autonomie', in D. Henrich, *Selbstverhältnisse. Gedanken und Auslegungen zu den Grundlagen der klassischen deutschen Philosophie* (Stuttgart: Reclam, 1982), 6–56. Translated as 'Ethics of Autonomy', trans. Louis Hunt, in Henrich, *The Unity of Reason*, 88–122.

'Kant's Notion of a Deduction and the Methodological Background of the First *Critique*', in Förster (ed.), *Kant's Transcendental Deductions*, 29–46.

The Unity of Reason: Essays on Kant's Philosophy, ed. Richard Velkley (Cambridge, MA: Harvard University Press, 1994).

Herman, Barbara. 'Justification and Objectivity: Comments on Rawls and Allison', in Förster (ed.), *Kant's Transcendental Deductions*, 131–41.

Moral Literacy (Cambridge, MA: Harvard University Press, 2007).

The Practice of Moral Judgment (Cambridge, MA: Harvard University Press, 1993).

Hill, Thomas E., Jr. *Dignity and Practical Reason* (Ithaca: Cornell University Press, 1991).

'Personal Values and Setting Oneself Ends', in *Human Welfare and Moral Worth* (Oxford: Clarendon Press, 2002), 244–74.

Hobbes, Thomas. *Leviathan*, ed. Richard Tuck (Cambridge: Cambridge University Press, 1997).

Höffe, Otfried (ed.). *Immanuel Kant: Kritik der praktischen Vernunft* (Berlin: Akademie Verlag, 2002).

Horn, Christoph and Schönecker, Dieter (eds.). *Groundwork for the Metaphysics of Morals* (Berlin and New York: de Gruyter, 2006).

Hudson, Hud. *Kant's Compatibilism* (Ithaca: Cornell University, 1994).

Hume, David. *A Treatise of Human Nature*, 2nd edn (Oxford: Clarendon Press, 1978).

An Enquiry Concerning the Principles of Morals (Oxford: Clarendon Press, 1975).

Hutcheson, Francis. 'Illustrations on the Moral Sense', in *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*, ed. Aaron Garrett (Indianapolis: Liberty Fund, 2002), 97–144.

James, William. 'The Will to Believe', in *The Will to Believe and Other Essays in Popular Philosophy* (New York: Longman, Green, & Co., 1897), 1–31.

Keller, Pierre. *Kant and the Demands of Self-Consciousness* (Cambridge and New York: Cambridge University Press, 1997).

Kersting, Wolfgang. 'Kann die "Kritik der praktischen Vernunft" populär sein? Über Kants Moralphilosophie und pragmatische Anthropologie', *Studia Leibnitiana* 25 (1983), 82–93.

Kleingeld, Pauline. 'Kant on the Unity of Theoretical and Practical Reason', *Review of Metaphysics* 52 (1998), 311–39.

'What Do the Virtuous Hope For?: Re-reading Kant's Doctrine of the Highest Good', in Hoke Robinson (ed.), *Proceedings of the Eighth International Kant Congress*, vol. I.i (Milwaukee: Marquette University Press, 1995), 91–112.

Klemme, Heiner F. 'Beobachtungen zur Kantischen Vermittlung von Theorie und Praxis in der praktischen Philosophie', in Hoke Robinson (ed.), *Proceedings of the Eighth International Kant Congress*, vol. II (Milwaukee: Marquette University Press, 1995), 521–31.

'Einleitung' and 'Sachanmerkungen', in Immanuel Kant, *Kritik der praktischen Vernunft*, ed. H. D. Brandt and H. F. Klemme (Hamburg: Felix Meiner Verlag, 2003), ix–lxiii and 219–42.

'Einleitung', in Immanuel Kant, *Kritik der Urteilskraft. Beilage: Erste Einleitung in die Kritik der Urteilskraft*, ed. H. F. Klemme (Hamburg: Felix Meiner Verlag, 2001; 2nd edn 2006), xxvii–ci.

'Praktische Gründe und moralische Motivation. Eine deontologische Perspektive', in H. F Klemme, M. Kühn and D. Schönecker (eds.), *Moralische Motivation. Kant und die Alternativen* (Hamburg: Felix Meiner Verlag, 2006), 113–53.

Koch, Lutz. *Kants ethische Didaktik* (Würzburg: Ergon, 2003).

Kopper, Joachim. 'Die Bedeutung der Methodenlehren', in N. Fischer (ed.), *Kants Metaphysik und Religionsphilosophie* (Hamburg: Felix Meiner Verlag, 2004), 391–407.

Korsgaard, Christine. *Creating the Kingdom of Ends* (Cambridge and New York: Cambridge University Press, 1996).

The Sources of Normativity (Cambridge and New York: Cambridge University Press, 1996).

La Mettrie, Julien Offray de. *Machine Man and Other Writings* (Cambridge and New York: Cambridge University Press, 1996).

La Rocca, Claudio. 'Methodenlehre, transzendentale', in G. Mohr, J. Stolzenberg and M. Willaschek (eds.), *Kant-Lexikon* (Berlin and New York: de Gruyter, forthcoming).

Landau, Albert (ed.). *Rezensionen zur Kantischen Philosophie,* vol I. (Bebra: Landau, 1990).

Leibniz, Gottfried Wilhelm. *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil* (London: Routledge and Kegan Paul, 1951).

Louden, Robert B. 'Go-Carts of Judgment: 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 and Oxford: Oxford University Press, 2000).

Łuków, Paweł. 'The Fact of Reason: Kant's Passage to Ordinary Moral Knowledge', *Kant-Studien* 84 (1993), 203–21.

Meerbote, Ralf. 'Kant on the Nondeterminate Character of Human Actions', in W. L. Harper and R. Meerbote (eds.), *Kant on Causality, Freedom, and Objectivity* (Minneapolis: University of Minnesota Press, 1984), 138–63.

'Which Freedom?', in P. Cicovacki (ed.), *Kant's Legacy* (Rochester: University of Rochester Press, 2001), 197–225.

Milz, Bernhard. *Der gesuchte Widerstreit. Die Antinomie in Kants 'Kritik der praktischen Vernunft'* (Berlin and New York: de Gruyter, 2002).

Munzel, G. Felicitas. '"Doctrine of Method" and "Closing"', in Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft*, 203–17.

Kant's Conception of Moral Character. The 'Critical' Link of Morality, Anthropology, and Reflective Judgment (Chicago and London: University of Chicago Press, 1999).

O'Neill, Onora. 'Autonomy and the Fact of Reason in the *Kritik der praktischen Vernunft* (§§ 7–8, 30–41)', in Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft*, 81–97.

Constructions of Reason. Explorations of Kant's Practical Philosophy (Cambridge: Cambridge University Press, 1989).

'Instituting Principles: Between Duty and Action', in M. Timmons (ed.), *Kant's 'Metaphysics of Morals'. Interpretative Essays* (Oxford: Clarendon Press, 2002), 331–47.

Piché, Claude. 'La métholodogie éthique de Kant', in S. Goyard-Fabre and J. Ferrari (eds.), *L'année 1797. La métaphysique des moeurs* (Paris: Vrin, 2000), 109–22.

Priestley, Joseph and Price, Richard. *Free Discussion of the Doctrines of Materialism and Philosophical Necessity* (London: Johnson and Cadell, 1778).

Proops, Ian. 'Kant's Legal Metaphor and the Nature of a Deduction', *Journal of the History of Philosophy* 41 (2003): 209–29.

Quarfood, Marcel. 'The Circle and the Two Standpoints (GMS III, 3)', in Horn and Schönecker (eds.), *Groundwork for the Metaphysics of Morals*, 285–300.

Rawls, John. *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, MA: Harvard University Press, 2000).

'Themes in Kant's Moral Philosophy', in Förster (ed.), *Kant's Transcendental Deductions*, 81–113.

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

'Autonomy, Practical Law, and Taking One's Choices to Be Good: Replies to Critics', *Philosophical Books* 49:2 (April 2008), 125–37.

'Introduction', in Kant, *Critique of Practical Reason*, trans. and ed. Mary J. Gregor (Cambridge and New York: Cambridge University Press, 1997), vii–xxxi.

'Kant's Critical Account of Freedom', in Graham Bird (ed.), *A Companion to Kant* (Oxford: Blackwell, 2006), 275–90.

'Two Conceptions of the Highest Good in Kant', *Journal of the History of Philosophy* 26 (1988), 593–619.
Rousseau, Jean-Jacques. 'Discourse on the Origin and Foundations of Inequality among Men', in *The Discourses and Other Early Political Writings*, ed. and trans. Victor Gourevitch (Cambridge: Cambridge University Press, 1997).
Sala, Giovanni B. *Kants 'Kritik der praktischen Vernunft'. Ein Kommentar* (Darmstadt: Wissenschaftliche Buchgesellschaft, 2004).
Schönecker, Dieter. *Kant: Grundlegung III. Die Deduktion des kategorischen Imperativs* (Freiburg and Munich: Alber, 1999).
Schulz, Eberhard. *Rehbergs Opposition gegen Kants Ethik* (Berlin: de Gruyter, 1975).
Schulz, Johann Heinrich. *Versuch einer Anleitung zur Sittenlehre für alle Menschen, ohne Unterschiede der Religionen* (Berlin, 1783).
Silber, John. 'Kant's Conception of the Highest Good as Immanent and Transcendent', *Philosophical Review* 68 (1959), 469–92.
Stark, Werner, *Nachforschungen zu Briefen und Handschriften Immanuel Kants* (Berlin: Akademie Verlag, 1993).
Sussman, David. 'From Deduction to Deed: Kant's Grounding of the Moral Law', *Kantian Review* 13 (2008), 52–81.
Timmermann, Jens. 'Das Creditiv des moralischen Gesetzes', *Studi Kantiani* 20 (2007), 111–15.
Kant's 'Groundwork of the Metaphysics of Morals'. A Commentary (Cambridge: Cambridge University Press, 2007).
Tittel, Gottlob August. *Ueber Herrn Kant's Moralreform* (Frankfurt and Leipzig: Pfähler, 1786; reprinted Brussels: Culture et civilisation, 1969).
Watkins, Eric. *Kant and the Metaphysics of Causality* (Cambridge and New York: Cambridge University Press, 2005).
'Kant's Antinomies: Sections 3–8', in G. Mohr and M. Willaschek (eds.), *Immanuel Kant: Kritik der reinen Vernunft* (Berlin: Akademie Verlag, 1998), 445–62.
Wike, Victoria. *Kant's Antinomies of Reason: Their Origin and Resolution* (Washington, D.C.: University Press of America, 1982).
Willaschek, Marcus. 'Die Tat der Vernunft: Zur Bedeutung der Kantischen These vom "Factum der Vernunft"', in Gerhard Funke (ed.), *Akten des Siebenten Internationalen Kant-Kongresses* (Bonn: Bouvier, 1991), 455–66.
Praktische Vernunft. Handlungstheorie und Moralbegründung bei Kant (Stuttgart: Verlag J. B. Metzler, 1992).
Williams, Bernard. 'Internal and External Reasons', in *Moral Luck* (Cambridge and New York: Cambridge University Press, 1981), 101–13.
Wizenmann, Thomas. 'An den Herrn Professor Kant von dem Verfasser der Resultate der Jacobischen und Mendelssohnschen Philosophie', *Deutsches Museum* [Leipzig] 1 (1787), 116–56.
Wood, Allen W. *Kantian Ethics* (Cambridge: Cambridge University Press, 2008).
'Kant's Compatibilism', in A. Wood (ed.), *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), 73–101.

Kant's Ethical Thought (Cambridge and New York: Cambridge University Press, 1999).

Kant's Moral Religion (Ithaca: Cornell University Press, 1970).

'Preface and Introduction (3–16)', in Höffe (ed.), *Immanuel Kant: Kritik der praktischen Vernunft*, 25–41.

Zedler, Johann Heinrich. *Grosses vollständiges Universallexikon* (Leipzig: Johann H. Zedler, 1732–54).

Index