

■ An Introduction to
Kant's Moral
Philosophy

Jennifer K. Uleman



CAMBRIDGE

CAMBRIDGE

www.cambridge.org/9780521199629

This page intentionally left blank

AN INTRODUCTION TO KANT'S MORAL PHILOSOPHY

Immanuel Kant's moral philosophy is one of the most distinctive achievements of the European Enlightenment. At its heart lies what Kant called the "strange thing": the free rational human will. This introduction explores the basis of Kant's anti naturalist, secular, moral vision. Moving from a sketch of the Kantian will, with all its component parts and attributes, to Kant's canonical arguments for his categorical imperative, it shows why Kant thought his moral law the best summary expression of both his own philosophical work on morality and his readers' deepest shared convictions about the good. Kant's central tenets, key arguments, and core values are presented in an accessible and engaging way, making this book ideal for anyone eager to explore the fundamentals of Kant's moral philosophy.

JENNIFER K. ULEMAN is Assistant Professor of Philosophy, Purchase College (State University of New York). She is the author of numerous articles and reviews.

AN INTRODUCTION
TO KANT'S MORAL
PHILOSOPHY

JENNIFER K. ULEMAN



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/9780521199629

© Jennifer K. Uleman 2010

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2010

ISBN 13 978 0 511 67664 2 eBook (NetLibrary)

ISBN 13 978 0 521 19962 9 Hardback

ISBN 13 978 0 521 13644 0 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

*For Ethan James and Sahara Rose,
with all the love in the world.*

Contents

<i>Acknowledgements</i>	<i>page</i> ix
1 Introduction: the strange thing	I
The strange thing	I
The free rational will	9
The value of free rational will	15
The importance of the strange thing for moral philosophy	19
2 A sketch of Kantian will: desire and the human subject	23
Desire, choice, will	25
Desire and the human subject	34
3 A sketch continued: the structure of practical reason	39
Will as practical reason: practical rules, laws, and principles	39
Maxims, or subjective practical principles	41
Grounds for action: the representation in a principle of something as good	48
Imperatives	51
Pure practical reason, or the possibility of a categorical imperative	56
4 A sketch completed: freedom	63
An overview of the free Kantian will	63
The free Kantian will in more detail	65
Rational freedom	72
5 Against nature: Kant's argumentative strategy	75
The problem	75
Kant's understanding of nature	79
Kant's common-sense case against a natural foundation for morality	80
Kant against nature	107

6	The categorical imperative: free will willing itself	111
	Kant's formalism	112
	Kant's categorical imperative: its form and its content	121
	Free will willing itself	140
7	What's so good about the good Kantian will? The appeals of the strange thing	145
	Introduction	145
	The good of free rational willing	149
8	Conclusion: Kant and the goodness of the good will	175
	<i>Bibliography</i>	180
	<i>Works by Kant</i>	180
	<i>Newer titles of note</i>	181
	<i>Works cited</i>	181
	<i>Index</i>	187

Acknowledgements

This book has been with me a long time, and I am grateful to many who helped, in many different ways. Thanks to Rachel Zuckert, Yasmin Yildiz, Sophia Wong, Elizabeth Wells, Jami Weinstein, Helga Varden, Corina Vaida, Kate Uleman, Kara Uleman, Jim Uleman, Lucas Thorpe, Amie Thomasson, Aviva Taubenfeld, Michele Sviridoff, Karen Struening, Michelle Stewart, Matt Statler, Maggie Smith, Michael Slotte, Alison Simmons, Harvey Siegel, Sally Sedgwick, Carolina Sanin, Michael Rothberg, Marilyn Rifkin, Giovanna Pompele, Thomas Pogge, Cara Perlman, Peter Ohlin, Darrell Moore, Marjorie Miller, Eric Mendenhall, Elias Markolefas, Peter Lewis, Kostis Kourelis, Morris Kaplan, Andrew Janiak, Risto Hilpinen, Sally Haslanger, Casey Haskins, Sean Greenberg, Lanny Goldman, Rudi Gaudio, Tracy Fitzpatrick, Frank Farrell, Simon Evnine, Jo Dexter, Marilyn Cleveland, Taylor Carman, Stephanie Camp, Marc Brudzinski, John Beusterien, and Tessa Addison. Thanks also to those I have failed to thank by name, but who have nonetheless helped me think, talk, and write about Kant.

I am grateful for the material support this project has received. Thanks to several Max Orovitz research grants from the University of Miami, to Paolo Tuzzi for the house in Cossogno that summer, and to the Learning Center at Purchase College for help funding my brilliant undergraduate research assistant, Felix Pichardo.

A student's thanks to Paul Guyer for being an outstanding teacher and mentor. He will surely disagree with parts of this book, but it is also thanks to his teaching that I wrote it. A writer's thanks to Hilary Gaskin and two anonymous reviewers at Cambridge University Press for all their care and dispatch. Finally, a teacher's very special thanks to all my Kant students, whose willingness to talk, listen, read, and argue has, more than anything else, helped me figure out what I wanted this book to say.

CHAPTER I

Introduction: the strange thing

THE STRANGE THING

“The thing is strange enough and has no parallel in the remainder of practical knowledge” (KpV 5:31). So writes Kant about the activity of human will. According to Kant, human will authors an ultimate action-guiding principle – a moral law – that tells what matters most and how to act accordingly. It binds itself to this law, experiencing the law’s commands as absolute and expecting as reward neither happiness nor heaven, eschewing both sensuous and divine incentives. According to Kant, human will understands the moral law it has authored as holding not only for itself but universally. The strange activity of this strange thing is strange for many reasons. It is free in a determined world; it subjects itself to itself, despite the seeming paradox of this; in the end, and strangest of all, the will that authors and can bind itself to moral law is *itself* what matters most, is itself the aim of morality. The strange will is thus its own object: at the heart of Kant’s moral theory is, to use Hegel’s words, “the free will which wills the free will.”¹ The moral law that Kantian free will authors is, to put it another way, strangely and ingeniously self-serving. This book is about all these strange things, and especially about why, for Kant, the strange, free, law-giving will is its own ultimate aim.

This book is about these things in order to offer an introduction to, as well as an interpretation of, Kant’s moral theory. It therefore surveys the foundations of Kant’s moral thought, laying out basics and making clear what Kant values, why he values it, and why he thought his famous “categorical

¹ Or in Hegel’s German, “*der freie Wille, der den freien Willen will.*” Hegel is here describing ‘the abstract concept of the idea of will in general,’ and though he does not name Kant in the passage, Hegel makes clear elsewhere that he admires Kant for identifying and attending to the will so conceived, even though he thinks Kant’s final moral theory comes up short (G. W. F. Hegel, *Grundlinien der Philosophie des Rechts* [Werke 7] [1821] [Frankfurt: Suhrkamp Verlag, 1970], §27; translation: *Elements of the Philosophy of Right*, trans. H. B. Nisbet, ed. Allen Wood [Cambridge: Cambridge University Press, 1991], §27, p. 57).

imperative"² the best summary expression of both his philosophical work on morality and of his readers' deepest shared convictions about morality and value. It aims to show that Kant's moral theory is driven by respect and awe for the specifically human capacity to act in the world in ways that are, as Kant understood the terms, free and rational. It aims to show that the core of Kantian moral theory is indeed the free rational will that wills itself. And it aims to show what a theory driven by respect and awe for free rational will asks of us – what forms of life, what micro-commitments, what conceptions of self, what collective arrangements it requires us to embrace, and what it requires us to reject.

For Kant, the complex that is 'free rational willing' or 'free rational activity of the will' cannot really be taken apart and still make sense. Each of its terms – 'free,' 'rational,' and 'will' – is ultimately defined in ways that implicate and depend on each other. Intuitions and ordinary usage thus cannot provide the guidance we need to understand the thing I claim is at the heart of Kant's moral theory, since intuitions and ordinary usage would let us separate these three. Indeed, intuitions and ordinary usage would sometimes oppose them.

Is there a term that names the Kantian complex? In early work on this book, I found myself using 'autonomy' to describe free rational willing. Autonomous activity *is* more or less the same as free rational activity of the will; 'autonomy' is characteristic of a will that (freely) gives itself a (rational) action-guiding law. 'Autonomy' thus has the advantage that it encompasses and inextricably relates, in one word, Kantian freedom, rationality, and will. But I have decided not to use the term here, at least not very often, despite its being, in some contexts, a key term for Kant himself. Not unlike 'freedom,' 'rationality,' and 'will,' the term 'autonomy' is so freighted, its accreted connotations so thick, its post-Kantian adventures so various and storied, that I prefer less felicitous terms and phrases, like 'free rational practical activity,' 'free rational willing,' and 'free rational activity of the will.' Besides triggering fewer associations for readers, these also have the advantage, when they come as phrases, of reminding us just what Kant *is* seeking to encompass and inextricably relate. They may thus be worthwhile in helping to keep Kant's conceptions strange and interesting.

² Kant's 'categorical imperative' is formulated in several ways. The most familiar are these: (1) act only on maxims that you can at the same time will as universal laws, and (2) treat others never merely as means but always also as ends in themselves. See G 4:421 and 429. Much more will be said about Kant's categorical imperative in subsequent chapters.

The interpretation of Kant's moral theory that I offer cuts against the grain of interpretations that emphasize Kant's commitments to formal rules and rationalism. Such interpretations have deservedly influential proponents³ and, despite recent scholarship that pulls toward feeling and content,⁴ formalist, rationalist views of Kant still circulate widely in lecture halls and college corridors and in the collective intellectual imagination more generally.⁵ Not without reason: Kant *was* deeply committed to a kind of formalism, and was deeply committed to rationality. But if these commitments are overemphasized, or emphasized in the wrong ways, we are left with a view that is less engaging and more academic than Kant's. Kant's Aristotelian, Humean, Hegelian, and other foes have taken note: overly formalist, rationalist interpretations have the capacity to drain the life out of Kant's views, and accordingly have been offered as often by Kant's enemies as by his friends.⁶

³ Important work by Christine Korsgaard, Onora O'Neill, and John Rawls pulls Kant in what I think of as formalist, rationalist directions. See Christine Korsgaard, *The Sources of Normativity*, with G.A. Cohen, Raymond Geuss, Thomas Nagel, and Bernard Williams, ed. Onora O'Neill (Cambridge and New York: Cambridge University Press, 1996) and the essays collected in Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge and New York: Cambridge University Press, 1996), esp. "Kant's Formula of Universal Law," 77–105. See Onora O'Neill, *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge and New York: Cambridge University Press, 1989), esp. "Consistency in Action," 81–104. See John Rawls, "Themes in Kant's Moral Philosophy," in *Kant's Transcendental Deductions*, ed. Eckart Förster (Stanford: Stanford University Press, 1989), 81–113.

⁴ I am thinking especially of work by Paul Guyer, Barbara Herman, and Allen Wood. See the essays in Paul Guyer, *Kant and the Experience of Freedom* (Cambridge and New York: Cambridge University Press, 1993), esp. "Duty and Inclination," 335–93, and the essays in Paul Guyer, *Kant on Freedom, Law, and Happiness* (Cambridge and New York: Cambridge University Press, 2000), esp. "Freedom As the Inner Value of the World," 96–125, and "Kant's Morality of Law and Morality of Freedom," 129–71; the essays in Barbara Herman, *The Practice of Moral Judgment* (Cambridge, Mass.: Harvard University Press, 1993), esp. "Leaving Deontology Behind," 208–40; and Allen Wood, *Kant's Ethical Thought* (Cambridge and New York: Cambridge University Press, 1999). Marcia Baron's *Kantian Ethics (Almost) without Apology* (Ithaca: Cornell University Press, 1995) may also be seen to pull in this direction.

⁵ A representative textbook account of Kant as invested in reason and form can be found in Samuel Enoch Stumpf and James Fieser, *Socrates to Sartre and Beyond: A History of Philosophy*, 8th edn. (New York: McGraw-Hill, 2008). The account, which is admirably clear and fair, ends by observing that Kant's categorical imperative "speaks of the universality of the moral law, affirms the supreme worth of each rational person, and assigns freedom or autonomy to the will," but does not try to explain how these are connected to each other, or why any of them is morally attractive (Stumpf and Fieser, *Socrates to Sartre*, p. 289). A textbook account that is congenial to the view I am advocating can be found in James and Stuart Rachels' widely used *The Elements of Moral Philosophy*, 5th edn. (New York: McGraw-Hill, 2006), 130–40.

⁶ Annette Baier and Bernard Williams both portray Kant in a rational formalist light, and critique him for over-reliance on reason and formal procedures. See Annette Baier, *Moral Prejudices: Essays on Ethics* (Cambridge, Mass.: Harvard University Press, 1994); and Bernard Williams, *Moral Luck: Philosophical Papers 1973–1980* (Cambridge and New York: Cambridge University Press, 1981) and *Ethics and the Limits of Philosophy* (Cambridge, Mass.: Harvard University Press, 1985).

What do I mean by 'formalism,' and how can it be overemphasized? Kant does insist that diverse human aims and ends, to be morally acceptable, must conform to the 'form of universalizability' – that is, that they may be pursued only if they could be universally endorsed. Described thus, Kantian morality does not advance *any particular* aim or end – it just insists that we pursue our aims and ends only if they pass a formal test. Because it does not dictate particular aims or ends, Kantian morality seems able to accommodate good human lives lived across circumstance and historical time and place: it seems pluralistic and inclusive. Because it insists on universal acceptability, it seems to respect the value of hearing from everyone, or at least of imaginatively trying to, by putting yourself in other people's shoes (would it be OK with everyone?): it seems deeply democratic. Kant's view thus comes across as a natural predecessor to the sort of contemporary procedural liberalism advocated by thinkers like John Rawls, which claims a strong commitment to neutrality between competing conceptions of the good.⁷

However, as Rawls himself knew, and as I argue here, Kant is not so neutral, and demands much more than accord between ends and aims and a certain form: he demands that we embrace, as intrinsically and ultimately good, the free rational human will itself.⁸ Embracing the free rational will as good means organizing our individual and collective lives in ways that actively honor this good. As a consequence, Kantian morality rejects moral projects the ultimate object of which is to serve God, or to alleviate material suffering; these projects, for Kant, unacceptably subjugate free rational will to other ends. As we will see, Kantian morality also rejects projects, such as Nietzsche's, of radical self-invention, congenial as self-invention may sound to a project that values free will (especially construed as 'autonomy'). The free will Kant values is one that is fundamentally legible to others, and committed to a radically shared rationality. In emphasizing the aim or end – free rational willing – that Kant requires us to embrace, my presentation seeks to draw out the specific shape of the moral life Kant demands we live.

⁷ Rawls' basic thought is that human beings can arrive at ground rules ('principles of justice') capable of fairly governing social and political institutions without prejudice between particular cultural, religious, or other conceptions of the good (John Rawls, *A Theory of Justice* [Cambridge, Mass.: Harvard University Press, 1971], e.g., 11–16 and 446–52). Many see Rawls as 'softening' this view in his later work, via acknowledgement that his principles of justice *are* hostile to some historical and contemporary ways of life (whether 'traditional,' religious, or strongly communitarian) (John Rawls, *Political Liberalism* [New York: Columbia University Press, 1996] esp. xv–xxxii, 174–6, and 243–4).

⁸ See John Rawls, "Justice as Fairness: Political, Not Metaphysical," *Philosophy and Public Affairs* 14:3 (1985), 223–51. In this essay, Rawls distances his own 'Kantian' view from the 'metaphysical' commitments about the nature of the soul and its vocation that he finds in Kant.

How does my presentation cut against overly ‘rationalist’ interpretations? Such interpretations emphasize the moral law’s origins in Kantian pure reason. Now, the moral law *does* have such origins for Kant. And such origins *do*, as Kant intended, ensure that morality is not subject to local emotional whims or physical exigencies; such origins also ensure that the moral law holds universally for ‘all rational creatures’ (including all human beings).⁹ But because rationalist interpretations too often employ (or at least let stand) a soulless, calculative conception of reason, they can fail to make palpable reason’s own strong commitments, including its interests in and reverential respect for its own strivings. If the very idea of reason *having* commitments and interests of its own seems strange, that is because we have become accustomed to thinking of reason in precisely this soulless, calculative way. But we need not – think, for instance, of the commitments to and interests in things like accuracy, consistency, scope, simplicity, and fruitfulness that Thomas Kuhn has persuasively shown us are part and parcel of rational scientific investigation.¹⁰ These are interests internal to reason itself. The interests just named are internal to, in Kant’s language, ‘theoretical’ – or theory-building – reason. But there are also interests internal to Kantian ‘practical’ – or action-guiding – reason, chief among them an interest in free rational striving itself (or so I will argue here). Overly rationalist interpretations thus also often do violence to the emotional investments Kant thinks we have in freedom, rationality, and willing. ‘Awe’ and ‘respect’ are just two of the key terms Kant frequently uses to describe our reactions to free rational willing, whether our own or other people’s. In emphasizing the interest and reverential respect that animates Kant’s reason-grounded morality, I break with the calculative, emotionally neutral spirit of much ‘rationalist’ interpretation of Kant.

I said above that some of Kant’s foes have favored formalist and rationalist interpretations. And indeed, formalism and rationalism, especially taken together, can be harnessed to cast Kant in a very unappealing light. The reader has perhaps been introduced to this Kant: he cares more about rules than about ends, he is wedded to impersonal calculation, he is unwilling to acknowledge his own particularity, he eschews all feeling, even (if not

⁹ Barring only those who for some reason lack the capacities that constitute reason – for example, infants, young children, and those with severe mental impairments. Lest this seem to exclude too many, notice that, for Kant, even “children of moderate age,” who presumably lack fully developed reason, nonetheless have sound moral intuitions and feel respect for duty (G 4:411n; see also KpV 6:155–7).

¹⁰ Thomas Kuhn, “Objectivity, Value Judgment, and Theory Choice,” in his *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977) 320–39.

especially) kind and warm feelings.¹¹ All the while, he insists on ‘duty’ for its own (incomprehensible) sake, and he generally comes off as a very cold fish. A significant part of my aim here is to move away from interpretations that focus on this Kant and emphasize instead the ends, interests, and feelings (such as reverential respect), that drive Kant’s theory. The effect, I hope, will be to make his view less vulnerable to charges of motivational opacity, of emotional coldness, and of a stultifying obsession with rules. If the aim of moral thought and action is free rational activity itself, if we understand ourselves as deeply committed to this activity, and if rules are just the best way to express strategies for honoring this activity, Kant’s view seems sturdier. It gains a kind of lived-life plausibility, feels more ennobling and less punishing, less rigid and more productive than critics have often charged. Moreover, when the interests, feelings, and ends internal to Kant’s project are made more apparent, the temptation to read Kant as imperiously and hubristically insisting that ‘all rational creatures’ must embrace his view (on pain of being banished from the realm of the rational) is diminished. If we read Kant’s as a view that must court our allegiance by identifying interests, feelings, and ends we share, the view becomes both more interesting and more satisfying to entertain. Charges of false and condescending universalism must give way to argument about the substance of Kant’s view, and the value of the interests, feelings, and ends he identifies.

The price of inviting argument about the value of Kantian interests, feelings, and ends is, of course, that argument will be offered. Kantian

¹¹ As in Friedrich Schiller’s satirical verse, meant to mock Kant’s commitment to duty over feeling:

Gewissenskrupel
 Gerne dien’ ich den Freunden, doch tu’ ich es leider mit Neigung.
 Und so wurmt es mir oft, daß ich nicht tugendhaft bin.
 Decisum
 Da ist kein anderer Rat, du mußt suchen, sie zu verachten
 Und mit Abscheu alsdann tun, wie die Pflicht dir gebet
 [Scruple of Conscience]
 Gladly I serve my friends, but alas I do it with pleasure.
 Hence I am plagued with doubt that I am not a virtuous person.
 [Ruling]
 Surely, your only resource is to try to despise them entirely,
 And then with aversion to do what your duty enjoins you.

From Friedrich Schiller, *Xenien* [1797], collected in Goethe, *Werke I*, ed. Erich Trunz (Hamburg: Christian Wegner Verlag, 1949), 221. This translation (apart from headings, which I’ve added) appears in H. J. Paton, *The Categorical Imperative: A Study in Kant's Moral Philosophy* (London: Hutchinson & Co., 1947), 48; Paton notes, “the translation, which I take from Rashdall, *The Theory of Good and Evil*, Vol. I, p. 120, is by A. B. Bullock.” (Thanks to Anne Margaret Baxley for helping me track down these sources.)

morality *does*, as I have just suggested, reject some historically real human moral projects, and this opens it to direct challenge. Without the pretense that it is neutral between competing conceptions of the good, its own conception of the good emerges more clearly and becomes more vulnerable to attack. Kant has to show what is wrong with following God's will, or minimizing pain, at least as ultimate aims. Charges that Kantian views overstate human independence and self-sufficiency, or that they unjustly denigrate the body and nature, must also be met head-on. The task left for the Kantian is to defend the Kantian conception of what matters, not as incorporating or accommodating all other plausible conceptions of the good, but as in fact superior to them.

In fact, the ultimate aim of this book is to show both *that* and *why* Kant thought his conception of the good superior – to show that Kant thought the strange, moral-law-authoring, free human will more valuable than anything else, and to show why. In the remainder of this Introduction, I will say more about this strange thing, about the general thought that Kant's moral theory is set up to honor and revere it, and about my strategy in pursuing this thought.

The three short chapters that immediately follow this Introduction – “A sketch of the Kantian will: desire and the human subject” (Chapter 2), “A sketch continued: the structure of practical reason” (Chapter 3), and “A sketch completed: Freedom” (Chapter 4) – are intended to provide readers with a portrait of the Kantian will, which is, as must be clear, the book's central character. The process of sketching this portrait allows me to establish the claims that the Kantian will cannot act without an end, and to show how Kant understands both the will's rationality and its freedom. Chapter 5 (“Against nature: Kant's argumentative strategy”) argues that Kant's preference for formal principles issues not from an in-principle desire to deprive morality of a substantive end, but instead from Kant's low view of nature, and so also argues for the impossibility of meaningfully ‘naturalizing’ Kant's view. This chapter also introduces readers to the basic intuitions Kant thinks his readers share about morality, including the key thoughts that moral value lies in the quality of an agent's intention, and that moral praise is never merited by action undertaken on ‘ulterior motives’ – intuitions that will ultimately be satisfied by a moral theory based on the value of a certain kind of will. Chapter 6 (“The categorical imperative: free will willing itself”) makes the case for free rational willing as the ultimate Kantian value via a close reading of canonical texts (the *Groundwork of the Metaphysics of Morals* and the *Critique of Pure Reason*). I show how my reading makes sense of a host of Kantian moves, including the claim,

perplexing on its face, that three central formulations of the famous Kantian categorical imperative are 'at bottom' the same.

The book then turns away from arguments *that* Kantian moral theory aims at free rational will, and asks *why*, for Kant, free rational will is so valuable. What qualifies free Kantian will for the reverential respect Kant thinks it occasions? What makes it intrinsically good, an end in itself, indeed the ultimate end of moral thought and action? [Chapter 7](#), "What's so good about the good Kantian will? The appeals of the strange thing," takes these questions up, drawing on Kant's many scattered comments about the merits of free rational willing. The task of this chapter is delicate. The question of why Kant thought that free rational will is intrinsically and ultimately valuable cannot be addressed by showing what he might have thought it good for, since the claim is that it is *good in itself*, not for some other reason or purpose; comparing it to other goods is also not so helpful, since it is meant to be better than anything else. The task is thus one of unpacking or making explicit the 'goods' carried by something intrinsically and ultimately valuable. The goods to be discovered are intellectual, psychological, emotional, interpersonal, social, political, and arguably even physical and spiritual. To ask why free rational willing is valuable, for Kant, is to ask for more information about the package we adopt, in terms of self-conceptions and the hoped-for overall shape of our individual and collective lives, if we endorse free rational willing as the ultimate end of our own wills. For this reason, [Chapter 7](#) is offered as an account of the lived self-conceptions and experiences of Kantian subjects committed to and acting in accord with free rational will. I defend such an approach in more detail below, but the reasons for wanting an account of lived Kantian free rational willing should be clear: I want to make the fundamental motivations for Kant's view apparent, and saying he values free rational will, without saying more about what this means, about how this looks on the ground, and hence about why it might appeal, leaves too much unspoken. [Chapter 8](#), "Kant and the goodness of the good will," reviews the argument of the whole, revisiting the strangeness and the accomplishment that is a moral system based on the value of free rational will willing itself.

Before moving to a more extended overview of the basic terms of the project, a few remarks about things I will *not* do here, and a note about interpretation. First, I make no effort to survey the extensive and very excellent literature on Kant's moral theory, though I try to acknowledge debts and conscious disagreements when I can. That literature offers an embarrassment of riches, and my aim here is more introductory than comprehensive. Second, although they are very interesting, I do not worry

deeply here about the metaphysical issues – chiefly about freedom – raised by my account. They have been thoughtfully addressed by others;¹² some I try to address elsewhere;¹³ some I would like to think about another time. Finally, this about my approach to interpreting Kant: interpretations are always interpretations, and the many complicated factors that make different people interpret the same text differently are well known to all who have taken hermeneutics seriously.¹⁴ I was initially drawn to Kant because of an interest in the devotion so many people, including me, seem to have to the moral value of freedom per se. Kant has been a compelling interlocutor in trying to understand (at least one version of) this devotion. In arguing, implicitly and explicitly, that the interpretation of Kant's moral philosophy I offer is a good one, and is better than some others, I mean to argue that, in hard-to-quantify proportions, it does a good job accommodating and elucidating a range of central texts and delivers up a reading that is philosophically, morally, and psychologically plausible and powerful. This is, of course, what most interpreters try to argue; whether I succeed is for the reader to decide.

What follows here, as promised above, is a more extended overview of the basic terms of the project as it will unfold in the rest of the book.

THE FREE RATIONAL WILL

To begin, what makes a will *free*? A will is free, for Kant, if it determines itself and is not determined by anything else. A will is free, in other words, if it chooses ends, and pursues courses of action aimed at realizing those ends, on grounds that are its own, and not on grounds given to it by something or someone external to it. There are, of course, high philosophical stakes in any account of free will, and there are many theorists who would gloss 'free will'

¹² I am thinking especially of Henry Allison's work on Kant's theory of freedom, and of the many responses to Allison's work. See Henry Allison, *Kant's Transcendental Idealism: An Interpretation and Defense* (New Haven: Yale University Press, 1983/2004) and *Kant's Theory of Freedom* (Cambridge and New York: Cambridge University Press, 1990). For replies and comments, see Karl Ameriks, "Kant and Hegel on Freedom: Two New Interpretations," *Inquiry: An Interdisciplinary Journal of Philosophy* 35:2 (1992), 219–32.; Stephen Engstrom, "Allison on Rational Agency," *Inquiry: An Interdisciplinary Journal of Philosophy* 36:4 (1993), 405–18; Paul Guyer, "Review of Allison's *Kant's Theory of Freedom*," *The Journal of Philosophy* 89:2 (Feb. 1992), 99–110; and Andrews Reath, "Intelligible Character and the Reciprocity Thesis," *Inquiry: An Interdisciplinary Journal of Philosophy* 36:4 (1993), 419–29.

¹³ Jennifer Uleman, "External Freedom in Kant's *Rechtslehre*: Political, Metaphysical," *Philosophy and Phenomenological Research* 68:3 (May 2004), 578–601.

¹⁴ Like so many others, I have been guided in thinking about interpretation by Hans-Georg Gadamer, *Truth and Method* [1960], trans. and rev. Joel Weinsheimer and Donald G. Marshall (New York: Crossroad Publishing, 1989) and by Charles Taylor's essays, particularly those in *Human Agency and Language: Philosophical Papers, 1* (Cambridge and New York: Cambridge University Press, 1985).

differently, even at this level of generality. Descartes, for instance, understands freedom of the will as a mental ability to endorse and set oneself on a course of action or not (or, more simply, to *assent* to something or not), rather than as a function of the grounds determining action (or assent). Importantly for Descartes, nothing can limit this ability (though our actual efforts may be thwarted): our freedom (as mental endorsement) is ‘infinite.’¹⁵ Hobbes, in stark contrast, denies that wills can be free at all, arguing that only *bodies* can be free or unfree, where ‘free’ just means ‘unimpeded.’¹⁶ For our present purposes, the important point is this: for Kant, unlike for some others, the idea of choosing on grounds that are our own is foundational to his account of freedom and the free will.

We can bring this foundation into sharper focus by noticing two challenges faced by Kant’s conception of free will. The first challenge is posed by nature, as Kant calls the physical world; the second is posed by reason. The challenge posed by nature stems from the fact that we, who have wills, are embodied finite physical beings. As long as Newtonian laws of physics are operating – universal and necessary laws of mechanical cause and effect which, for Kant, govern everything in the natural world – it seems that everything we do must, if traced just a little way, have causal roots in forces and events outside ourselves. But if this is so, then our wills are merely conduits for external causes, particular kinds of locations in a causal net that stretches out infinitely in all directions. Under these conditions, we could not meet Kant’s criterion for being free: we would not determine ourselves, but would be externally determined. The challenge posed by reason is analogous. Just as a will determined by external mechanical causes is not free, neither is a will determined by the demands of reason. To the extent, for instance, that I cannot reject a conclusion (of, say, a chain of mathematical reasoning), I am not free and my acceptance of the conclusion is not properly a choice. It must be *up to me* what I chose, in some ultimate sense – choices, if they are to be real choices, cannot be dictated by external rules or standards. This point was made often enough by scholastic and early modern voluntarists, who insisted that in order to be truly free, God’s will (or ‘power of volition’) had to be free from answerability to reason or, for

¹⁵ This is at least an important piece of Descartes’ view. See, for example, René Descartes, *Meditations on First Philosophy with Selections from the Objections and Replies* [1641], trans. John Cottingham (Cambridge and New York: Cambridge University Press, 1986), 39–42 (AT 56–61). For a nice discussion of the complexities of Descartes’ view, see Gary Hatfield, *Routledge Philosophy Guidebook to Descartes and the Meditations* (New York: Routledge, 2003), 183–202.

¹⁶ Hobbes writes, “the Liberty of the man ... consisteth in this, that he finds no stop.” Quotation and discussion both Thomas Hobbes, *Leviathan* [1651], ed. Richard Tuck (Cambridge and New York: Cambridge University Press, 1991), 146 (Ch. 21).

that matter, to any set of independent standards, whether of truth or goodness or beauty or whatever. If reason constrains me, forcing me to recognize six as the sum of three and three, I do not determine myself.

Faced with both these challenges, Kant's predecessor Locke concluded that human will is never free, but is always determined, by either nature or reason: for Locke, 'free will' is a nonsensical phrase. *Human beings*, Locke thought, could properly speaking be free or not, depending on whether they, as whole beings, were free to pursue what they willed.¹⁷ But wills are always determined by something other than will itself, and so are never *themselves* free. But for a theorist like Kant, who wants to keep a meaningful sense of 'free will,' another route must be found. There must, for Kant, be a way to insist on the possibility of the will's *own self*-determination, that is, on will's being itself definitively pushed neither by laws of nature nor by laws of reason.

Kant's solution to the problem of determination by natural law is well known, if not widely accepted. Kant argues that there are two sorts of things, one of which is subject to natural causality, one of which is not. Spatio-temporal objects and events of the sorts available to the senses – *phenomena*, things that *appear* (φαίνω or *pheno*) – are subject to natural causal law; things that are not spatio-temporal or available to the senses, things like God and the immortal soul – *noumena*, which can only be grasped through the intellect or mind (νοῦς or *nous*) – are exempt from the natural causal order. Human wills, Kant argues, are, in significant if limited ways, non-spatio-temporal non-sensible things, and so are exempt (in significant if limited ways) from natural determination. Will thus joins God, the immortal soul, and possibly many other things, in 'the noumenal realm,' a realm [*Reich*] of objects and concepts related to each other not by mechanical cause and effect but by a different set of logical and conceptual relations and a different set of laws.¹⁸ How to conceive the relationship between the noumenal and the phenomenal – a relationship there must be since our noumenal wills (like God's) produce results in the spatio-temporal sensory world – remains vexed. But what matters here is that Kant, in seeking to preserve will's freedom, insists on envisioning will as something other than a mechanism determined from without and instead as another sort of thing, belonging to an order outside nature, outside space and time, which can begin its own unconditioned beginnings.

¹⁷ John Locke, *An Essay Concerning Human Understanding* [1689], ed. Peter H. Nidditch (Oxford: Oxford University Press, 1975), Book II, Ch. 21.

¹⁸ At least some of which Kant will, not insignificantly, call 'laws of freedom.' See, for example, G 4:387, or KpV 5:65, or MS 6:214.

What about the challenge from reason? Reason threatens freedom by placing *intellectual* demands. This threat may be less familiar to us, but was well known to Kant. The threat as Kant faced it can be expressed thus: "If I will what I will for reasons, is it not fair to say that reasons determine my will?" Or, more specifically, "If it is inevitable that I will whatever I judge best, and if judgment is a process of applying standards and rules, am I not therefore bound to will according to rules and standards?" Kant's solution to the challenge posed by reason in fact echoes Leibniz's solution to the scholastic and early modern problem of voluntarism. For the strict voluntarist, as noted above, God's will is absolutely free, unconstrained by any independent nature of things or by rules or standards. Neither the nature of space, time, and matter, nor the rules of math and logic, nor standards of beauty and goodness can constrain God; on the contrary, all must be at his pleasure. Otherwise, God's freedom would not be complete or perfect (which, like God's goodness and justice and knowledge, it must be). God must be able to choose and make real or true whatever he wills, even if it is that two plus three equals six, or that something be and not be at the same time.

Now one of the deepest problems facing those with a strong commitment to voluntarism is that it renders God both arbitrary and incomprehensible. If we have before us the thought that God can change the natures of things and the rules that govern them, willy-nilly, we cannot hope to understand God or his creation, or to make sense of his aims and purposes for us and for the world he created. God becomes a fearsome and arbitrary despot, not a just and wise creator and sustainer. And as Leibniz argued, the praise and worship God expects, and which we offer, make no sense unless we have good independent grounds for judging him praiseworthy.¹⁹

Leibniz offers an ingenious solution to the dilemma presented by voluntarism, maintaining God's freedom to follow rules or not, and to make rules as he chooses, without making him arbitrary or incomprehensible. For Leibniz, God creates rules and laws and standards according to what in some all-told, general sense seems best to him, though he doesn't have to; he could author sub-optimal rules. Even having authored optimal rules, God does not have to follow them. Instead, God *chooses* to follow those rules, and, as it turns out, does so over and over again. But it is still true that he doesn't have to author the best rules, or to follow them: we worship him because he reliably does so, because he chooses to do what is best, given his

¹⁹ G. W. Leibniz, *Discourse on Metaphysics* (1686), in *Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett Publishing, 1989), 36–7, §§2–3.

own overall background all-told sense of ‘best.’ The key, of course, lies in there being some such overall background all-told sense, one that belongs to God as God, and that is also comprehensible to us. Leibniz nominates God’s interest in maximal harmony with maximal diversity as providing such an overall background all-told global criterion of ‘bestness’: let there be as many things as possible organized as elegantly as possible, thinks God. (It’s an appropriate most general interest for a creator and sustainer, isn’t it?) God’s very general interest in maximal diversity joined with maximal harmony cannot be properly said to *constrain* his will, but instead represents his will’s most basic contents – and he could, in any case, choose to ignore it.

Kant’s effort to conceive the human will and its freedom bears deep similarities to the solution Leibniz offers to the dilemma presented by voluntarism. How so? God’s will, for Leibniz, and human will, for Kant, are both powers to choose things and to initiate action aimed at making those things real or true.²⁰ God’s will for Leibniz, and human will for Kant, are both free when this power to choose is not determined by external forces, but instead by the will itself, on grounds that are its own.²¹ Just as Leibniz’s God has a fundamental desire that befits him, *qua* God, so free human will, for Kant, has a fitting interest, namely an interest in its own continued free agency. For Kant, this interest is both fully mine and fully rational – it is not fundamentally external to me. In contrast, grounds of action that stem ultimately from my physical body, with its physically determined needs and desires, *are* seen by Kant as foreign or external to me; they are the equivalent of rules or standards not my own. So, for Kant, are determining grounds attributable to ‘the will of God’ – for Kant, a person who puts God’s will before her own forsakes her own freedom.²²

Like God’s will for Leibniz, Kantian free will *can* depart from reason, perversely deciding against what in some most general sense seems best to it: for both, will remains radically free.²³ But when it is responsive to reason, this responsiveness precisely honors its own deepest internal structure and commitments. For Kant, the will’s responsiveness to reason honors will’s own deep interest in free rational agency, that is, in itself. (God, being

²⁰ See Kant, KpV 5:9n and MS 6:211–13; see G. W. Leibniz, “On Freedom and Possibility” [1680–82?], in *Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett Publishing, 1989), 19–23.

²¹ See Kant, G 4:446, for example; see Leibniz, *Discourse on Metaphysics*, §30, pp. 60–1.

²² If, however, an agent decides, for reasons of her own, that God’s will is good, and that she therefore endorses its demands, she maintains her freedom. At stake here is whether the ultimate standard of goodness, the standard guiding judgment, is understood to lie within (maintaining freedom) or without (impairing it). See REL 6:177.

²³ See, for example, Kant, KpV 5:20; see Leibniz, “On Freedom and Possibility,” 19–23.

perfect, will always honor his deepest internal structure and commitments – we, on the other hand, often mess this one up.) A will, on this ‘Leibnizo-Kantian’ view, does not therefore have to act in ways that are arbitrary or indifferent to be free. A free will can be responsive to (though not wholly determined by) reasons for acting, without any diminution of freedom, given that the choice to be responsive was its own. Moreover, insofar as the will chooses to be responsive to grounds that are truly its own, that is, insofar as the will chooses in ways that ‘make sense’ (are rational) given its own interests, it maintains its full freedom at the same time that it completely *self-determines*.

We can summarize these thoughts by noticing how they fit with Kant’s helpful distinction between *Willkür* – the capacity to choose – and *Wille* – the faculty of practical reason as a whole. Both, as the words themselves suggest, are part of will.²⁴ And both contribute distinctive elements or components to human freedom. The first, *Willkür*, is the capacity for free choice itself, the capacity to choose ‘at will’ between alternatives – alternative ends, alternative courses of action, alternative guiding principles of action. For Kant, *Willkür* is metaphysically necessary for morality since without it praise and blame and responsibility-holding would not make sense: to be held responsible, to be considered the author of an action, an agent must be the ultimate source of her choices.

Wille, the second term, is the capacity to formulate ends, and to formulate action-guiding principles aimed at serving those ends. Thus does Kant call *Wille* ‘practical reason itself’:²⁵ *Wille* conceptualizes and formulates in ways that actually guide practice, or intentional action. For Kant, ends and action-guiding principles formulated by *Wille* insofar as it seeks grounds within itself and not in external sources, that is, ends and action-guiding principles formulated by *pure* practical reason, count, not surprisingly, as ends and action-guiding principles that are deeply *mine*. Such ends and principles are grounded in interests internal, for Kant, to my deepest self, my free rational self. And by choosing to act in accordance with such purely rational ends and principles, I choose action that is given aim and shape by this self. Of course, once I choose a course of action, I am determined – I am no longer exercising a capacity to go this way or that – but if I have chosen to act toward ends and on principles that are truly my

²⁴ As is noted below, and as will be discussed at greater length in Chapter 4, ‘*Wille*’ is both the Kantian term for that part of will that is specifically rational, and for will as a whole – just as ‘reason’ for Kant is sometimes the term for pure reason (in contrast to sensibility or understanding) and sometimes for the whole faculty (including sensibility and understanding).

²⁵ MS 6:213. See also MS 6:226; G 4:412.

own, I am still free in the crucial sense that I am *self*-determined. These two components of Kantian freedom – a capacity for choice (*Willkür*) and a capacity to furnish ends and principles that are my own (*Wille*) – are not reducible to each other, but are both essential components of will, as Kant understands it. Together, they make Kantian sense of the possibility of a free will.

The claim here (and the claim I will try to make good on in the next three chapters) is that Kantian will must always be understood as a complex, standing possibility of action that is simultaneously free and rational. As a power to choose, Kantian will is radically free from determination by anything, whether natural or rational: it *can* be arbitrary, can decide in favor of anything, ‘at will.’ But this freedom is connected, is hung together – is synthesized a priori, as it were, for those already fluent in Kantian terminology – with reason. As a free power to choose ends and practical principles – as a free power, that is, to envision, make, and change things – Kantian will recognizes a fundamental interest in its own continued free envisioning, making, and changing, in its own free agency. This interest links freedom to reason: reason identifies this interest, and shows us how to pursue it. When we act in ways that are responsive to our own rational discoveries about how best to respect and pursue this interest, we remain free (since the aims of action are deeply our own) and are at the same time rational (since the source of action is not blind or arbitrary). Indeed, freedom and reason enhance each other: for Kant, the more rational our actions, the more they preserve and promote freedom, and the more freedom we exercise, the more we expand the arena in which reason, rather than blind force, operates. We might then conclude here by noticing that free will, for Kant, is a possibility that intimately involves reason even as it sets itself worlds apart from nature.

THE VALUE OF FREE RATIONAL WILL

The centrality of the complex that is freedom and reason and will, taken together, to Kant’s moral theory cannot be disputed. Kant’s 1797–8 *Metaphysics of Morals* is an extended treatise on the “laws of freedom,” or the structure of rational human action.²⁶ But what makes this complex so *valuable*? Freedom and reason and will all have inherent value for a thinker like Kant; none is merely instrumentally valuable. (If this seems surprising, reflect on the fact that freedom, reason, and creative power [the power of

²⁶ See, for example, MS 6:214.

will] are all considered ‘perfections’ or laudable virtues in the early modern conception of God.) It is true that for Kant, as for many thinkers, freedom of the will is a metaphysical precondition of morality, required if our practices of moral deliberation, giving advice, attributing responsibility, and praising and blaming are to make sense; if we were not free to choose, all of these gestures would be empty. Reason likewise is a precondition of morality: if we were not able to represent action-guiding principles to ourselves, to conceptualize and prioritize ends, etc., the demands of morality could not emerge. And if the question ‘what should I do?’ is to be asked at all, I clearly must have a will, capable of *doing*.²⁷

But beyond being preconditions of morality, freedom and rationality and will are all, for Kant, *themselves* good. Each is better, fundamentally, than its opposite: bondage, nonsense, and impotence are all, *prima facie*, things from which we want to move away. Kant is of course clear that, when detached from each other, freedom and rationality and will *can* go bad: freedom not organized by reason leads to anarchy, which Kant identifies with “savage disorder” (VE 27:344; see also REL 6:35 and 37); reason that has not recognized its commitments to freedom and to agency, and that therefore adopts ends fundamentally disrespectful of them, is worse than nonsensical bumbling (G 4:394, on the dangerous “coolness of a scoundrel” see also VE 27:366); will, when in the grip of external power, as well as when arbitrary, can do more harm than good (VE 27:344–6; G 4:447; see also G 4:393). But this does not change the basic prejudice in favor of all three, or the very important fact, for Kant, that when freedom and reason and will work hand in hand, when the complex is knit together as it should be, any ‘falling aways’ from proper use are corrected. When freedom, rationality, and will are all fully present, and fully cooperating, we have something very good indeed. We have, in fact, the thing Kant declares, in the opening sentence of the [first chapter](#) of *Groundwork of the Metaphysics of Morals*, to be the only thing, “at all in the world, or indeed even beyond it,” that can be “considered good without limitation,” that is, a good (free rational) will (G 4:393).

Freedom, rationality, and will are so jointly constitutive of goodness for Kant that they appear, in his texts, as a package more often than not. They are conceptually and sometimes functionally distinct, but each so ineluctably

²⁷ See the following from the *Metaphysics of Morals*: “An action is called a *deed* insofar as it comes under obligatory laws and hence insofar as the subject, in doing it, is considered in terms of the freedom of his choice. By such an action the agent is regarded as the *author* of its effect, and this, together with the action itself, can be *imputed* to him” (MS 6:223).

implicates the others – if we are trying to understand any of the three at its best – that Kant will often use umbrella terms to signal the whole complex: Kantian *Wille* or ‘practical reason as a whole’ functions this way (KpV 5:28–30; see also, e.g., KpV 5:72). As we suggested earlier, ‘autonomy,’ or rational self-determination, functions this way as well, characterizing a will that is at once free and rational.

So, what *is* the value of a will at once free and rational? What makes this will good? The answer depends on the exegetical arguments of the book’s first chapters: that Kant’s commitment to the value of free rational will makes most sense of the structure of Kantian willing itself; that Kant’s commitment to the value of free rational will makes most sense of the ‘common-sense’ assumptions about morality Kant accepts and seeks to accommodate philosophically; and that Kant’s commitment to the value of free rational will makes most sense of Kant’s claims about his categorical imperative. But successful as these arguments may be, they are apt to frustrate the reader who wants to know not just *that* but also *why* Kant takes free rational will to be so valuable. What does Kant think is so great about free rational will? What does he think *we* will think is so great about it that it will move us to embrace his moral theory? Any argument that Kant’s view is motivated by the attractions of free rational will must, I believe, say something full-bodied about what, after all, *is* attractive about free rational will.

To this end, I will, as noted earlier, offer the beginnings of a story about the experiences of self and other and world promised by Kantian morality. Such a story will bring the appeals and attractions of such a will into relief. This strategy will strike many as distinctly unKantian. First, the very idea of a story about the experience of Kantian free will may give many pause: Kant famously insists that the free will is not an object of experience, that it cannot appear, cannot be ‘phenomenal.’ An account of free will as it appears would therefore seem ruled out of court. Second, the suggestion that free rational Kantian willing ‘appeals,’ ‘attracts,’ or ‘motivates’ will raise flags. Kant famously argues that the moral good cannot attract or appeal or motivate in any ordinary way – that indeed, were it to do so, it would automatically disqualify itself as a truly *moral* good. Kantian morality is supposed to command a disinterested loyalty, to generate obligations without promising the moral agent satisfactions in return. Thus, an effort to make the experiential attractions of Kantian free rational willing clear may provoke qualms, precisely for looking at *experience* and for seeking *attractions*.

But in fact the qualms are misplaced, and in my view have been stumbling blocks for many as they seek to make deep sense of Kant’s view. First off, it is important to notice that when Kant tells us that free will is not an object of

experience – not something that appears, not phenomenal – he does *not* mean that we cannot have awareness of it. Kant means something more specialized, namely that a free will is not the kind of thing that can occupy a spatio-temporal position or be fit into a comprehensive theory describing spatio-temporal objects. ‘Experience’ (*Erfahrung*), for Kant, is a technical philosophical term, much narrower in its application than our ordinary English word ‘experience.’²⁸ For Kant, to experience something is to be able to conceptualize and fit it within a system of spatio-temporal events and objects governed by mechanistic causal laws and described by Newtonian science. In fact, many things of which we can have ordinary awareness cannot be conceptualized or fit in this way, and would be ruled ‘outside experience’ for Kant. Money, *qua* money, for example, has no place in this Kantian world of “experience,” though pieces of pressed metal and printed paper do. But we cannot understand the nature or movements of money, *qua* money, in terms of spatio-temporal events and objects governed by mechanistic causal laws.²⁹ Money is nonetheless certainly a part of our experience, in the ordinary English sense of the term. In claiming that the free will cannot ‘appear,’ or find a place in the world of experience, Kant’s concern is to distinguish the conceptual ‘space’ in which free will operates from the space described by Newtonian science – since there is no room for freedom in *that* space. When I call for a story about the ‘experience’ of freedom, I am using ‘experience’ in the ordinary, more expansive English sense. And in calling for such a story, I am trying to break the grip of a dogma that has, I think, made it hard for us to look to Kant for an account of the lived life of a free rational willing subject. But the account is there, in the texts, plain to see as soon as we let ourselves focus on it. I seek, in the beginnings of a story I offer, to describe the experience of self and world and other that is made available to the subject who adopts a Kantian commitment to free rational willing.

Second, Kant’s insistence that the moral good doesn’t appeal or attract or motivate in an ordinary way is not insistence that it cannot appeal or attract or motivate *at all*, but that it cannot appeal or attract or motivate in a way that could be tracked and explained by a scientific physiology of sensuous

²⁸ The German word ‘*Erlebnis*,’ which has a different meaning from ‘*Erfahrung*,’ is also translated into English as ‘experience.’ The difference between them is instructive. Where ‘*Erlebnis*’ is something one has, or suffers, or lives through, or is otherwise the passive ‘recipient’ of, ‘*Erfahrung*’ is active gaining and organizing of information, the sort of ‘experience’ one is supposed to gain through on-the-job training. One can, in German, have an uncanny or revelatory or otherwise unclassifiable ‘*Erlebnis*’; ‘*Erfahrung*’ in contrast, is always presumed to cohere with and add to the rest of what we know about the ordinary world. (Thanks to Yasmin Yildiz for help with this point.)

²⁹ See Kant’s account of money, and also his account of property as ‘noumenal possession,’ in his *Metaphysics of Morals*, MS 6:286–9 and 245–51 respectively.

desire. Kant's concern is not that morality be 'disinterested,' but that it not be driven by natural impulse, which would render it fundamentally compulsive and not truly free. Kant makes strong statements, to be sure, that oppose 'duty' and 'desire' (see, e.g., KpV 5:22ff.). But these should not be taken to suggest that Kant wants us to do our duty *for no reason*. Readers of Kant sometimes, I think, take Kantian admonitions too far, sticking to their letter, or what appears to be their letter, but losing their spirit, rushing away from anything that smacks of motive or desire, and consequently insisting on a view that we must somehow be drawn to do our duty without anything drawing us. But Kant is certainly clear that the moral law, moral duty, the good will, free rational willing, and so on, occasion strong feelings of reverential respect (see, e.g., KpV 5:73 and G 4:436). And only a determined effort to reject all suggestion of 'appeal' can deafen readers to the further hosts of ways that free rational willing, for Kant, has palpable, 'experienceable,' attractions, appeals, advantages, and satisfactions. It will do us good, I think, to recognize that experience and appeal, in their *non-Kantian* senses, play key roles in Kant's own argument for his moral theory.³⁰

THE IMPORTANCE OF THE STRANGE THING FOR MORAL PHILOSOPHY

Making the appeal of Kantian free rational willing palpable is key to showing that Kantian morality intends to present a compelling, secular, humanist moral vision, one that is able to do what competing moral systems claim to do, namely offer guidance, give lives meaning, confer dignity, and promise 'higher' satisfactions. Kant's view is meant to compete not only with the Aristotelian and utilitarian views to which it is so often contrasted in academic debates, but also with those value systems that draw on monotheistic religious teachings and visions of the moral universe. Indeed, taken seriously, Kantian free rational will reclaims for humanity

³⁰ 'Desire' and 'happiness' are also terms that mean things for Kant that are not what we *must* mean by them and that are narrower than ordinary usage. Again I think readers have often let Kant's use of these terms mislead them into concluding that, for example, there is no sense in which Kantian morality could contribute to (non-Kantian) happiness, or that there is no sense in which agents could desire (in a non-Kantian way) the moral good. Kudos here to Alenka Zupančič and Allen Wood for fearlessly writing about Kantian desire, in an ordinary sense of the term, for the moral (Alenka Zupančič, *Ethics of the Real: Kant, Lacan* [London and New York: Verso, 2000]; Allen Wood, *Kant's Ethical Thought*), and to Paul Guyer, for pursuing the contributions of morality to happiness, in both Kantian *and* ordinary senses of the term (Guyer, *Kant on Freedom*). Kudos along similar lines to David Cummiskey for pointing out ways in which Kant's view is consequentialist – resisting an overly fastidious blanket rejection of the idea that Kant might put any moral stock at all in consequences (David Cummiskey, *Kantian Consequentialism* [New York: Oxford University Press, 1996]).

many of the awe-inspiring qualities early modern philosophers invested in God: creative power (the ability to bring genuinely new things into being, to initiate change in the world), self-sufficiency (at least with respect to the task of living a morally good life), intimacy with universality and necessity (in understanding and in action). It seems to me worthwhile, in these interesting times – times when virulent religious fundamentalisms and wanton political ambition threaten to eclipse secular, egalitarian liberalism – to recall the intellectual and moral resources of humanist Enlightenment thinkers such as Kant. Kant offers a philosophical platform capable of supporting an egalitarian vision of thoughtful, engaged human beings, joined in communities that are held together by mutual respect and admiration for human strivings and capacities. Important and legitimate critiques of Enlightenment thinking notwithstanding, this vision is still potent. The conception of the free rational activity of the human will at the heart of this Enlightenment vision may perhaps inspire us in new ways.

It seems to me urgent that we not shy away from the thought that a moral theory might inspire. Moral theories stand or fall – that is, they gain adherents and influence practices, or they fail to – not on the basis of philosophical consistency or argumentative superiority (important as these things rightly are to philosophers) but because they promise a lived experience of self, of others, of community, and of the world that answers deep human longings. Perhaps some longings are universal: most moralities promise happiness, even as they articulate it differently. But the longings answered by a moral theory may also be historically and culturally specific, born of particular struggles or circumstances or even mere restlessness. From my point of view here, it does not much matter: whatever their source, and however universal or necessary, human longings for a better way of living are what make people care about what a moral theory has to offer, and whether we try to adhere to a given theory is a question of whether, at the end of the day, the theory says something that resonates enough, that excites and holds and intensifies the feeling that something better is possible enough, to hold our allegiance.

Kant tries to get and hold our allegiance in many ways. If we insist on our own free rationality and act in ways that respect the free rationality of others, some of our deepest longings, Kant thinks, will be answered. In understanding ourselves as capable of determination by our own reason, and not only by the physical laws of nature, we will come to see ourselves as more than inert cogs in the vast machine of the universe. We will experience ourselves as loosed from instinctive animality, from the brutish whims and compulsions of our sensuous natures. We will find ourselves able to occupy

a 'higher order': freely, rationally authoring moral law permits us to occupy a space characterized by universality, necessity, and infinity and free of particularity, contingency, and finitude. Free rational willing constitutes us as potent, active, and intellectually self-sufficient. In understanding ourselves as self-determining, we understand ourselves as genuinely creative agents. The value of free rational willing detaches the conditions of our moral worth and dignity from fate or luck and places it in our own hands. In understanding ourselves and others as free rational agents, we find grounds for deep respect and admiration for humanity.

Of course, some will not like these promises. Where the rational formalist interpretation of Kant's view invited charges of coldness, hyper-rationalism, and reliance on a generally implausible moral psychology, the interpretation I am offering – according to which Kantian morality delivers a substantive moral vision – invites other charges. Ideals of purely free rational human willing have been accused of promoting destructive delusions of self-sufficiency, of ignoring dependency and connectedness, and of denigrating, among other things, nature, women, community, and love. Given the focus on the individual, views premised on the value of free rational willing have also been charged with an incapacity to detect, and therefore to condemn, structural injustice.³¹

³¹ Critics of many stripes have charged that Kantian autonomy requires coldness toward loved ones, promotes delusions of self-sufficiency, simultaneously sustains and conceals racial, sexual, and economic stratification, and detaches agents from the communities and contingencies within which they are located and which therefore ought to inform their self-conceptions and actions. The coldness critique is laid out by Ermanno Bencivenga, "Kant's Sadism," *Philosophy and Literature* 20:1 (1996), 39–40. Representatives of gender-based critique are found in Eva Kittay, *Love's Labor: Essays on Women, Equality, and Dependency* (New York: Routledge, 1999); Annette Baier, *Moral Prejudices*; Carol Gilligan, *In a Different Voice: Psychological Theory and Women's Development* (Cambridge, Mass.: Harvard University Press, 1982); and in essays collected by Virginia Held, ed., *Justice and Care: Essential Readings in Feminist Ethics* (Boulder, Col.: Westview Press, 1995) and Robin May Schott, ed., *Feminist Interpretations of Kant* (University Park, Pa.: Pennsylvania State University Press, 1997). Critiques based on Kant's racist or racialist thinking can be found in Emmanuel Eze, "The Color of Reason: The Idea of 'Race' in Kant's Anthropology," in *Anthropology and the German Enlightenment*, ed. Katherine M. Faull (Lewisburg, Pa.: Bucknell University Press, 1995), 196–237; and Charles W. Mills, "Dark Ontologies: Blacks, Jews, and White Supremacy," in *Autonomy and Community: Readings in Contemporary Kantian Social Philosophy*, ed. Jane Kneller and Sidney Axinn (Albany, NY: SUNY Press, 1998), 131–68. Communitarian critiques trace their lineage back to G. W. F. Hegel, *Phänomenologie des Geistes (Werke 3)* [1807] (Frankfurt: Suhrkamp Verlag, 1970) (English: *Phenomenology of Spirit*, trans. A. V. Miller [Oxford: Oxford University Press, 1979]) and *Grundlinien der Philosophie des Rechts* (English: *Elements of the Philosophy of Right*). Contemporary communitarian critiques may be found in Michael Sandel, *Liberalism and the Limits of Justice* (Cambridge and New York: Cambridge University Press, 1982); Alasdair MacIntyre, *After Virtue* (Notre Dame: University of Notre Dame Press, 1981); and Charles Taylor, *Human Agency and Language, Philosophy and the Human Sciences: Philosophical Papers, 2* (Cambridge and New York: Cambridge University Press, 1985), and *The Ethics of Authenticity* (Cambridge, Mass.: Harvard University Press, 1992).

These criticisms, I think, are both more interesting and more threatening to Kantian moral theory than charges of coldness and hyper-rationalism, because they cut closer to the bone. They take ideal Kantian free rational willing directly to task. I would be satisfied here if I made it easier to entertain these criticisms, and to think productively about whether Kant has resources with which to reply. I would be happy if I made it harder for Kantians to rehearse philosophical arguments that all rational agents, in some minimal universal senses of 'rational' and 'agent,' are compelled to adopt Kantian principles. It seems to me these arguments avoid the real fight. I would very much like to see Kantians and anti-Kantians take the debate to the floor, arguing the merits of their views in terms of the self-conceptions and lived lives they offer.

The account of Kant's project I offer here is thus at once more and less modest than some. It is more modest because it does not try to argue that all agents, or all rational agents, or all agents with free wills, really, despite what they may think they believe, do or must adopt Kantian morality. This was often Kant's rhetorical strategy, but it is not mine. My aim is to step back and try to shed light on Kant's underlying conception of free rational agency, and to show why he thought this kind of agency would command respect and commitment sufficient to ground moral theory and practice. The account here is also more modest than some in that it does not begin the hard work of figuring out what Kantian morality would have us do in many important cases. Arguing that the ultimate aim of Kantian moral theory is free rational willing just sets the stage for hashing out implications, and doesn't give deliberators procedural rules that are foolproof. Hence the less modest aspect of the account offered here: although it offers no algorithm, it does offer what is meant to be a reasonably substantive conception of the good, a sketch of a vision of human flourishing and excellence capable of inspiring allegiance and guiding struggle and real deliberation. In the end, it is, of course, for the reader to decide whether this approach is worthwhile.

CHAPTER 2

A sketch of Kantian will: desire and the human subject

Reason, in practice, has to do with a subject and especially with its faculty of desire. KpV 5:20

Like many people I know, I often try to draw abstract, non-spatio-temporal things on blackboards. When I start trying to draw the Kantian will, students become particularly hopeful. Having a clear picture of this unwieldy faculty would make life a lot easier – but after the first few circles and arrows, we all end up discouraged. There are too many different parts and pieces, interacting in too many different ways. These early chapters represent my efforts at offering, instead, a sketch in writing of the complicated Kantian will.

Kantian will is complicated because it is at once a faculty that desires, makes choices, and issues action-guiding rules. To say it desires is to say that it wants and wishes, that it has inclinations and interests. To say it makes choices is to say that it decides between possible ends or aims of action, picking which desires we act upon. To say it issues action-guiding rules is to say that it is a faculty that formulates maxims,¹ as well as rules for deciding among possible maxims; it is to say that will authors, and represents to itself, and determines itself according to, principles.

To further complicate matters, Kantian will – encompassing desire, choice, and rule-making – is also at once thoroughly rational and thoroughly free, and also often incompletely rational and incompletely free. Kantian will is thoroughly rational in that it always employs concepts and representations, and always does what it does in ways that invoke reasons and hence always make (some kind of) sense. Kantian will is thoroughly free in that it is never determined, in its choices or in the rules it issues, by anything other than itself, and hence always does whatever it does

¹ Or what Kant also calls ‘subjective rules for action’ (G 4:420n) – personal, local rules. Much more on maxims will be found below in [Chapter 3](#).

ultimately of its own accord, that is, freely. However, for Kant, there are further senses of 'rationality' and 'freedom.' In the first senses, just invoked, rationality and freedom are thoroughgoing – will is thoroughly (inevitably, pervasively, ineliminably) rational and free. But in their further senses, rationality and freedom are neither pervasive nor ineliminable, but are hardly inevitable achievements to which the Kantian will should (but does not always) aspire. The Kantian will *ought* to choose – to determine itself, to set itself on courses of action – on grounds that are at once *most* rational (that make the most sense, that have the strongest support of our best considered judgments) and *most* free (that are most deeply its own, that reflect the deepest interests of our true selves). While it cannot help being thoroughly rational in a basic sense, Kantian will is only rational in a perfect, complete, or full sense when it identifies, adopts, and is guided by *the best* reasons; similarly, while it cannot help being thoroughly free in a basic sense, Kantian will is free in a perfect, complete, or full sense only when it is determined by those parts of the self that are our best parts. This, at any rate, is what I will try to show in this chapter, in as clear and helpful a way as possible. Will is at the center of Kant's moral theory, and having a clear picture of it will, I hope, make it easier to see that it propels the shape and aims of Kant's moral theory overall.

In addition to beginning the project of sketching Kantian will, this chapter advances a particular argument. Kantian will must have an end or aim in order to act – this is not controversial, and is a point Kant himself makes (see, e.g., KpV 5:34). But the argument is that, according to Kant, will can (and ultimately should) make *itself* its own end, that is, will can (and ultimately should) make its own free rational activity the ultimate action-guiding end of will itself. On my view, Kant's moral theory as a whole hangs together only once we see this clearly. This is so because the activity of free rational will is the only end fully compatible with the will's own complete freedom: if it took up any other end, as its ultimate end, it would subject itself to determination by external forces. The activity of free rational will itself is also the only end fully compatible with the will's complete rationality: free rational willing is the only end reason itself must nominate. What does it look like when will makes itself – its own free rational activity – its ultimate end? Will can, on the one hand, aim at its own free rational activity by instantiating it and, on the other, by honoring and respecting it. This, at any rate, will be my argument.

This chapter begins by looking at what might, for Kant, be described as will's most basic, minimal, or primary capacities and characteristics. It builds on these toward will's more elaborate, 'higher,' capacities and

characteristics. There is something artificial in this, since for Kant the ‘lower’ and ‘higher’ capacities that human will comprises are in truth mutually informing and interdependent; they could not exist as they do without each other, they would not be themselves without each other. But it is a way of proceeding, and as good as any given that one must start somewhere.

DESIRE, CHOICE, WILL

Desire

We begin, as Kant does in many discussions, with *desire* [*Begehren*] (e.g., MS 6:211–13; KpV 5:20). Like all living creatures, human beings have, for Kant, a ‘capacity of desire’ [*Begehrungsvermögen*] (MS 6:211; KpV 5:9n). Kant describes the capacity thus:

The *capacity for desire* is the capacity to be by means of one’s representations the cause of the objects of these representations. (MS 6:211, Kant’s emphasis; see also KpV 5:9n)

This uninviting definition essentially says that to desire is to move toward the realization of something of which one has an idea. To put it in words closer to Kant’s definition, it says that the ability to desire is the ability to be, by means of, or through, one’s representations the cause of the coming into being or the happening of the objects (understood as aims or ends: ‘objectives’) of those representations, be those objects spatio-temporal things, events, states of affairs, the instantiation of moral value, or whatever.² A creature has a capacity for desire if a creature has a capacity to do this – if the creature can, to put it loosely, try to make an idea a reality. Of course, as

² The idea of causality that takes place according to representations may give some readers pause. It is clearly what Kant has in mind: at MS 6:357, for instance, Kant writes “the causality of a representation (whether the causality is external or internal) with regard to its object must unavoidably be thought in the concept of the capacity for desire.” The grounds for pause are that causality through representations looks a bit like causality that is teleological and not efficient or mechanistic: an object is realized in accord with a final cause. Kant rejects this worry. In the *Critique of Practical Reason*, Kant writes that even though “psychological” causality, or causality that “produce[s] actions by means of representations and not by bodily movements” (KpV 5:96), is not material and so not “mechanical,” we can assimilate it to the causal order of nature because it exhibits “the necessity of the connection of events in a time series as it develops in accordance with natural law” (KpV 5:97). Causality through representations, in other words, conforms to the rule that when the concept of causality (‘every effect has a cause’) is applied to nature, the cause precedes the effect in time (see, e.g., A202/B247). When desire for an object causes action, desire precedes action, conforming to this rule. The idea is still peculiar, however, and how exactly causality through representations or conceptions interacts with material mechanism remains a question.

Kant knows (see MS 6:356), we who desire do not always *successfully* make our ideas reality; we sometimes fail to successfully cause the objects of our representations. But a capacity for desire is precisely a capacity to organize our energies and organs to *strive* in a certain way, namely toward the realization of something that is (only, for now) in our minds. Desire, as Kant elsewhere puts it, is a capacity for *causality through representations* (MS 6:357).

This is probably not what most readers, off the top of their heads, would have offered as a definition of desire. But it captures all the wanting and longing and inclining and being interested that more ordinary definitions would mention. To want, or long, or incline, or be interested, is to organize oneself and one's energies toward an object in a certain way: this is what Kant's definition emphasizes. In doing so, Kant's definition helps us see desire as a capacity that *animates* in a certain way, namely toward something we want but don't yet have. And this, if reflected on briefly, surely is what most of us understand as desire.

Kant's definition also captures a crucial basic sense in which a capacity for desire is a capacity for self-directed action. If a creature's movements are structured by an internal representation, the creature's movements issue – at least in a crucial basic sense – from the creature itself, and not from external forces. Of course, the desire may itself be something the creature cannot help; the impetus for the action may not ultimately be under the creature's control. But wherever there is desire, there is movement that is organized by, however we ultimately understand this, an internal goal and not (just) by external forces.

Now, for Kant, all living creatures desire (MS 6:211). Kant doesn't say much about what the capacity for desire is like in animals (or in plants),³ but we can assume Kant imagines a capacity to process external stimuli, form representations, and respond, however automatically or instinctively. A gecko flicks its tongue to catch a bug, or darts away from a predator; it has moved in ways structured by representations and goals. This is already a lot (and it could bear a lot more exploration), but for Kant it is just the beginning. Beings that represent things by means of *concepts* – beings Kant

³ The capacity for desire is, for Kant, characteristic of life itself. Kant writes: "The capacity of a being to act in accordance with its representations is called life" (MS 6:211; see also KpV 5:9n). It is hard to tell from Kant's comments whether he means to extend the capacity for desire to plants as well as animals. Insofar as plants 'pursue ends' (sunlight, water) and 'sense' (hence, represent to themselves?) the environment, they may perhaps be said to 'be by means of their representations the cause of the objects of these representations' (e.g., plants achieve adequate sun and water). Unfortunately, Kant's comments on life elsewhere do not help settle the question (MAN 4:544). Nothing turns on this at present, however, so although the question is interesting, we leave it unanswered.

identifies as *having reason* – have a more complicated capacity of desire (MS 6:213). The capacity is the capacity for *choice*.

Choice

Kant calls the capacity for desire in beings with reason *Willkür*, which he glosses as the capacity of “doing or refraining from doing as one pleases” (MS 6:213). *Willkür* is best translated as the *capacity for choice*.⁴ In German, *willen* is ‘to want’ (or ‘to will’) and *küren* is ‘to elect or choose’; *Willkür*, accordingly, is the ability not just to desire, but to elect or choose what we will ‘at will,’ we might say, or as we please. Most radically, this will mean that *Willkür* is where the buck stops: *Willkür* determines, but is not in turn determined by anything else. *Willkür* will therefore be the seat, so to speak, of radical freedom. But before we get to that, we need to see why Kant might suggest that ‘adding’ reason to desire gets us a *Willkür* or capacity for choice at all.

How does being a creature with reason or, as Kant writes, ‘representing by means of concepts’ (MS 6:213), transform a capacity for desire into a capacity for *choice*, for doing or refraining from doing as one pleases? How does reason make this difference? Kant does not answer this question directly, but we can piece the start of an answer together from other things Kant writes.

First, we need to see how ‘representing by means of concepts’ makes one, for Kant, *rational*. A short sketch of a story will serve our purposes here. For Kant, a concept is an abstracted representation of something, a representation that ‘contains’ all and only the characteristics that would qualify something to be an instance of the concept.⁵ The concept ‘bird,’ for instance, contains or represents all the essential characteristics of birds. The concept ‘bird’ does not contain the characteristics peculiar to species, or to exemplars, which would be contained in ‘lower-level’ species concepts (‘parrot,’ ‘crow’) and eventually instance concepts (‘my parrot Bruno’). If I have concepts, I can compare other things – objects, mental representations, experiences, other concepts – against them, and make judgments: this is one of those; that is not; these are similar, but not identical; those are opposed,

⁴ We must all thank Mary Gregor for this translation of ‘*Willkür*.’ See her translation of Kant’s *The Metaphysics of Morals* (MS), 282n11.

⁵ For a wonderfully clear and informative account of Kant’s understanding of concepts, and of what it means for concepts to ‘contain’ information, see R. Lanier Anderson, “The Wolffian Paradigm and Its Discontents: Kant’s Containment Definition of Analyticity in Historical Context,” *Archiv für Geschichte der Philosophie* 87 (2005), 22–74.

etc. Concepts thus serve as mental standards or models.⁶ Now, making judgments about whether, for example, an instance falls under a concept, about whether this is one of those, always requires a primitive invocation of reasons, of justification ('this is one of those because it has feathers and a beak').⁷ And to invoke reasons is to *be rational*, at least in a minimal, basic sense. Employing concepts involves comparing and judging against standards, which involves primitive justification, and creatures that do this qualify as rational.⁸

Having seen, at least very roughly, how possessing and employing concepts makes one *rational*, we need, next, to see how concept employment, that is, basic rationality, might transform the capacity for desire so that it becomes a capacity for *choice*. It is useful here to turn to Kant's "Conjectural Beginnings of Human History" (which we also drew on above). Here Kant, speculating about the beginnings of human reason, describes not only the advent of the capacity to compare and measure against standards (concepts) but also the significance of the development of long-term memory, and of our (consequent) recognition and awareness of the trajectory of our human lives, from birth to eventual death. Memory plus awareness of birth and death gives rise to previously unknown cares, worries, and dreads, but also spurs us (with the help of concepts) to construct aims, projects, ambitions – ends, objects – that transcend immediate circumstances, going beyond the demands and desires of the immediate situation (MAM 8:III–13). The gap this generates between stimuli and action makes room for much of what we ordinarily think of as choice. Rationality of the sort described here enables a being to become the kind of being capable of regarding options, of weighing and assessing and therefore of *choosing* between, rather than being reflexively or instinctively impelled toward, possible courses of action. This ability to frame alternatives and weigh their merits is, it seems to me, a

⁶ Support for these suggestions can be found in Kant's 1786 essay, "Conjectural Beginnings of Human History" (MAM 8:III–12).

⁷ In addition to his comments in "Conjectural Beginnings," Kant's comments in the *Anthropology*, §§6–11 (ApH 7:137–46), support the view of judgment and basic rationality suggested here.

⁸ Human beings are rational creatures, but animals, for Kant, are not. As living creatures, geckos desire, for Kant, and must therefore 'represent' their objects (e.g., bugs) in some sense; and geckos do not, I think Kant thinks, have stimulus-independent abstract representations, or concepts; they do not have a concept 'bug,' and do not, therefore, make judgments about whether particular objects of experience are bugs. They just eat them. Kant's use of the expression "animal choice" or "*thierische Willkür*" at MS 6:213 (and also at A802/B830) does threaten to confuse the issue, but I take it that the qualification 'animal' should lead us to read choice/*Willkür* in these cases as if it were in scare quotes. Allen Wood has a useful discussion of the difference between animal desire and human choice (Allen W. Wood, *Kant's Ethical Thought* [Cambridge and New York: Cambridge University Press, 1999], 51). Wood emphasizes the fact that human choice is subject to (or subjects itself to) norms, where animal desire is 'mechanical.'

piece of what Kant has in mind when he tells us that desire, in a rational creature, becomes the capacity of doing or refraining from doing as one pleases, that is, becomes the capacity for choice. Where there is room for reasons (in setting priorities and ordering actions), there can be choice.

There is another piece. To choose, for Kant as for many of his (and our) contemporaries, is also to be *free*. That is to say, for Kant as for many others, a choice is only *really* a choice if, in some ultimate sense, *I* make it. For Kant, creatures that are mechanically moved to act toward the objects of their strongest inclinations have no choice. A (real, free) choice must, for Kant, be an endorsement that is *underdetermined* by anything, save by my will itself. This means, first, that it must be underdetermined by inclinations rooted in our ‘sensuous constitutions,’ that is, by inclinations that have ultimate roots in our physical needs and desires (G 4:446). It also means that my choice must be underdetermined by reason, despite the fact that, for Kant, as we shall see, rational creatures find their own best grounds for action in reason. The crucial point for us here, however, concerns underdetermination: the fact that I have available to me grounds for acting furnished by nature (by my physical needs and desires) *and* grounds that have an altogether different source (reason itself) means that I am not pulled inexorably toward either. Rationality, and hence the possibility of ‘determination by reason,’ plays a direct and immediate role in opening up a space in which *I can choose*. This is not the whole story about Kantian freedom, or its relation to rationality; we will need to be shown how reason itself can determine the will without threatening its essential freedom, as well as how reason can be more and other (to anticipate the broadly Humean, naturalist objection) than a tool for organizing what it might be argued are all, at base, sensuous drives and desires. But all we need to notice here is that it makes sense for Kant to depict the transition from mere desire to choice as also already a transition, made possible by reason, from external determination to freedom.

Having seen what it means to say, for Kant, that rational creatures have a special kind of desire, viz., the capacity for choice, or *Willkür*, and having claimed that, in transforming desire into the capacity for choice, reason brings with it the possibility that actions will be undertaken on rational grounds, and will be undertaken freely, we now turn away from a focus on desire as choice and turn to it as a form of *causality*. Recall: when something is desired – or, by rational beings like ourselves, chosen – action toward that object is initiated (MS 6:211; KpV 5:15; G 4:417). This is ‘analytically,’ Kant would say, what it is to choose something: to choose something *just is* to move toward its realization, by analysis of the concept ‘choice.’ This deserves emphasis. To understand the capacity for choice as a kind of

causality is to understand it as always and essentially *effecting* something. This something need not be the thing desired or chosen – sometimes we fail to attain the object we want, just as a desiring gecko sometimes fails to catch the bug (see MS 6:356). But the living being's faculty of desire – choice or *Willkür* in us – always causes movement, striving, toward whatever object it represents to itself as desirable (KpV 5:15; G 4:417).⁹ This point can be clearly seen by noticing Kant's distinction between wishing and choosing. Wishing, in contrast to choosing, involves, Kant tells us, the representation of some object as desirable without attendant movement toward the object (MS 6:213; G 4:394; ApH 7:251).¹⁰

There are two important reasons to emphasize the fact that the capacity for choice is a kind of causality, for Kant. The first is that it underscores choice's 'executive' role in the whole complex that is Kantian will. Contra a connotation that may be carried by 'capacity for choice,' *Willkür* as Kant conceives it is not particularly deliberative. It issues decisions, yes, but in doing so it is more the jury foreman who announces the jury's decision, setting the courtroom into a particular kind of motion, than it is the jury itself; it *makes choices*, it *effects* choice, simultaneously declaring one option the winner and initiating the relevant action. Deliberation, with all the capacities it calls on, belongs properly not to *Willkür* itself, but to will full-blown, to will as a whole, the sketch of which we are still just beginning.

The second important reason to emphasize that the capacity for choice is a kind of causality is that it forces us to notice that it must, for Kant, operate *according to laws* (see KpV 5:15 and A539/B567). This is because, for Kant, causality always and fundamentally takes place according to laws: the concept of causality, Kant writes, "always requires that something *A* be of such a kind that something else *B* follows from it **necessarily and in accordance with an absolutely universal rule**" (A91/B123–4, Kant's emphasis; see also KpV 5:89). Cause and effect are not cause and effect unless they proceed in a lawful manner, and to think about causality, for Kant, is

⁹ "Movement" may include 'inner' movements of the mind (as when I choose to concentrate on a task), efforts at refraining from ordinary movement (as when I choose to sit very still), as well as 'ordinary' movement. It may even include 'movement' in a metaphoric sense, as when we choose our moral characters, a choice that takes place, for Kant, outside of space and time (see REL 6:25). (If action outside space and time seems incomprehensible, recall that for Kant and many of his contemporaries all of God's activity is non-spatio-temporal. This may not make non-spatio-temporal action easy to grasp, but at least it gives a sense that the concept is not completely bizarre or unmotivated.) Thanks to Simon Evnine for asking me to clarify the meaning of 'movement.'

¹⁰ Kant, however, notes that even when we merely wish, our bodies are set in motion: our hearts pound harder and our blood rushes faster (KU 5:177–8n; see also MS 6:356–7). If anything, this just underscores the point that, for Kant, desire is always and everywhere a power with physical, empirical effects.

always, necessarily, to think about law; to try to think causality otherwise is to think an incoherence. Choice, or *Willkür*, it will turn out, is the seat of a special sort of causal activity because its causality is causality that can be governed by laws the will represents to itself.

These features clearly point up that we need more than *Willkür*, even for *Willkür* itself. We need deliberation, and we need the capacity to self-govern by means of laws we represent to ourselves. To make sense of our peculiar capacity for desire, *Willkür*, we need to turn to the capacity Kant calls *Wille*.

Will

The German term, *Wille* (translated 'will'), is often used by Kant as an umbrella term, denoting the entire faculty of desire as it is found in rational beings, *Willkür* and all. This is how we have used the term 'will' thus far. But in his late (1797) *Metaphysics of Morals*, Kant explicitly contrasts *Wille* and *Willkür*. (We drew attention to this discussion in the Introduction.) While Kant's use of these terms, before and after he explicitly contrasts them, is not always consistent, the distinction he draws in the *Metaphysics of Morals* is worth noting and can be useful for labeling conceptually (if not really) separable capacities.¹¹ *Willkür*, as we saw above, is the capacity for choice. In contrast, *Wille*, Kant writes, is,

the capacity for desire considered not so much in relation to action (as the capacity for choice [*Willkür*] is) but rather in relation to the ground determining choice [*Willkür*] to action. (MS 6:213)

A few pages later, Kant writes that *Wille* is

not directed to actions but immediately to giving laws for the maxims of actions (and is therefore practical reason itself). (MS 6:226)

Kant's distinction is between that part of will that *provides rational guidance* for action and the part that *chooses and initiates* action. The first quoted passage points to *Wille's* role in furnishing grounds (i.e., reasons) that can 'determine,' or shape, action; the second points to *Wille* as giving laws, in particular 'laws for the maxims of actions,' that is, general principles suitable to guide the adoption of local, personal plans.¹² Now, it will turn out, as we

¹¹ The *Wille/Willkür* distinction has received wide attention from commentators. Helpful and influential discussions include Lewis White Beck, *A Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960), 177–80, and Henry Allison, *Kant's Theory of Freedom* (Cambridge and New York: Cambridge University Press, 1990), 129–36.

¹² As noted above, Chapter 3 has much more to say about Kantian maxims – and defends the sort of gloss of 'maxim' just offered here.

shall also see, that furnishing grounds and furnishing laws are two sides of the same coin; *Wille* always furnishes both at once, since grounds for acting are always implicit in practical laws, and some law or other is always implied by a given ground for acting. But this is getting ahead of the story, encroaching too fast on territory that will be covered by Chapter 3. What we need to notice here is just that Kant, in articulating a distinction between *Willkür* and *Wille*, describes a complete will that has both the capacity to choose and the capacity to give itself grounds for choosing. Or, to put it another way, a complete *Wille* is composed of both *Willkür* and *Wille*.

One might wish that Kant had used more terms, so that 'will' (or '*Wille*') did not have to do double duty. But the fact that the term is the same can actually be helpful in reminding us that, ultimately, the whole will has to work as one. On this point it can be useful to notice an analogy with Kant's use of the term 'reason.' Though Kant does not make this suggestion himself, *Wille* in the restricted, partial sense may be thought to stand to will overall as 'pure' reason stands to reason as a whole. How so? Pure reason, for Kant (as for his contemporaries), is our ability to think things and discover true connections between thoughts 'a priori,' that is, independent of a given, empirical reality.¹³ For example, 'pure' reason, acting 'a priori,' establishes concepts and discovers truths of mathematics and logic, which don't depend for their contents or validity on sensory experience of external spatio-temporal objects, that is, on investigations of empirical reality. Because pure reason does not rely on empirical reality – which may change and which might in any case have been otherwise – pure reason can affirm *necessary, universal* truths: 'one is not zero,' 'a thing cannot be *and* not be.'¹⁴ These truths shape and inform the activity of reason as a whole, preventing it, for instance, from affirming that this particular empirical thing both *is* and *is not*. Pure reason is thus the 'head' of reason taken as a whole. Now the thought about *Wille* is this. *Wille*, as ground- and law-giver, that is *Wille* in contrast to *Willkür*, stands to *Wille* generally (will as a whole) as pure reason stands to reason generally. Just as pure reason informs the rest of reason, directing the shape of its operations, so *Wille* as ground- and law-giver directs *Wille* as will overall.

This also fits nicely with the fact that, in various places (as above), Kant glosses 'will' as "practical reason itself" (MS 6:213, 226). 'Practical' reason is

¹³ Thought that depends, in contrast, on empirical reality for its truth warrants, for Kant and for his contemporaries, is not 'a priori' but 'a posteriori.'

¹⁴ Readers inclined to challenge these claims are asked to suspend their Hegelian, or Buddhist, or other such objections, valid and interesting as these may be, and notice the significant and undeniable role claims such as these play in organizing ordinary thought and action.

opposed, for Kant, to ‘theoretical’ reason: practical reason is reason employed to guide practice (or action), where theoretical reason is reason employed to construct theories or knowledge (paradigmatically, for Kant, natural scientific theories or knowledge).¹⁵ It is useful here to think about what would be required for reason to guide practice: what would need to be the case in order for there to be truly practical (action-guiding) reason? Reason would need the capacity to furnish and represent principles, the capacity to choose among principles, as well as the capacity to judge cases in light of those principles and to figure out how to act accordingly. Given this list, we can indeed say that will, as Kant understands it, is reason in its practical employment. The entire apparatus that makes rationally guided action or practice possible is thus ‘*Wille* as a whole’ – while the ‘head’ of such an apparatus or faculty is ‘pure’ *Wille*.

Now, let us try to get a deeper understanding of the thought we left above, viz., the thought that *Willkür* can be governed by laws the will represents to itself. “Everything in nature,” Kant writes in the *Groundwork*, “works in accordance with laws” (G 4:412). But, Kant continues, “only a rational being has the capacity to act *in accordance with the representation of laws*, that is, in accordance with principles, or has a *will*” (G 4:412; Kant’s emphasis). Later in the *Groundwork*, Kant writes, “the will is thought as a capacity to determine itself to acting in conformity with the *representation of certain laws*” (G 4:427; Kant’s emphasis). What is the thought here? Causality always happens according to law; law-governed happening just is what causality is. Rational beings are special in that we can cause – can act in the world – according to laws we represent to ourselves. Being able to do this is the same as having, for Kant, a will. OK – but how does this all work?

The ability to act according to laws I represent to myself is the ability to order my actions, *to cause*, in conformity with an internal guideline. It seems to me helpful to point out that, in order to do this, I must be able to do three things (which correspond to the three capacities mentioned just above). First, I must of course be able to represent a law or guideline to myself: I must have something like an internal mental life. Second, I must be able to choose or endorse the law, that is, I must be able to decide to order my actions in accord with it, and to actually order them thus: I must have *Willkür*. Third, I must be able to judge what actions do in fact accord with

¹⁵ Practical reason and theoretical reason are not, it is worth noting, two divorced faculties, but are shorthand descriptions for what Kant also calls ‘reason in its practical use’ and ‘reason in its theoretical use’; ‘practical’ and ‘theoretical’ describe two uses of reason, separable just in light of divergent aims and accordingly divergent principles.

the law I've chosen; I must, that is, be able to interpret the law, applying it to cases. But then this is just it: a rational, free creature is a locus of causality via laws it employs itself. To say we cause according to laws we represent to ourselves is to point to all this.

There is, of course, a clear contrast here between rational will and 'animal' desire, which does none of these things, causing movement instead according to instinct or reflex. There is, for Kant, no question of animals representing laws to themselves, let alone of choosing them or judging that certain actions accord with them: non-rational creatures are limited to behaving according to laws that operate upon or through them. Of course, human beings also have reflexive behaviors, and we are certainly subject to laws of nature (such as the law of gravity) whether or not we 'represent' them to ourselves. But when we behave reflexively, or, when our movements can be characterized solely in terms of unwitting obedience to natural laws, human will, properly speaking, is not part of the causal picture. When will *is* part of the picture, when we are *acting*, as opposed to merely behaving, our efforts are both described and guided by our own representations of laws.

DESIRE AND THE HUMAN SUBJECT

Before consolidating our account of desire as it works in rational creatures and moving on to [Chapter 3](#) to look in more detail at the rules and principles involved in Kantian willing, we need to make note of two related issues. First, the descriptions of desire and choice and action according to principles one represents oneself just offered on Kant's behalf should not be taken as a phenomenological or experiential description of willing. We may or may not be *conscious* of the ends, grounds, laws, etc. in accord with which we act. We often seek such awareness, and we are of course often called on to produce it ('what were you trying to do?'; 'what did you think you were doing?'), but Kant is clear that failure to properly identify the laws governing our wills is always possible – or, to put it in more familiar terms, we may always fail to identify our true motives, we may never be sure just what we are doing or why (see G 4:407). And if this is true, it must also be true that whether we are conscious of a rule does not have anything to do with (or at least does not decide) whether it is truly the rule according to which we are acting. Kant did not have an explicit theory of the unconscious, but it is clear throughout his writing that much of the activity of reason, in all its forms, takes place outside awareness (see, e.g., A555–6/B583–4). Kant, then, rejects the Cartesian and Lockean views that all thought is conscious – that

consciousness is a qualifying mark of the mental. For Kant, even activities like deliberation and choice can and do take place outside consciousness. To say that the capacity for desire is the capacity to be by means of one's representations the cause of the objects of these representations, and to say that specifically human desire (or desire in any rational creature) is the capacity to cause not only in order to realize represented objects, but also according to laws the will represents to itself is not, for Kant, to make a set of phenomenological or experiential claims. It is rather to make claims about the kind of causal power desire is, and about the kind of underlying, rationally structured activity it is when it is found and as it functions in rational creatures, like human beings, whether we think about or are aware of it or not. For Kant, desire, choice, the structuring of human action takes place, so to speak, in a space of reasons and meanings that is vast and largely outside our awareness.

Second, as has probably already occurred to some readers, the account of will sketched here implies that human actions are best described, indeed are the actions they are, in virtue of the internally represented laws they follow. In other words, an agent's internal principle determines the proper description of the action. Whether I am running from a fire or going for a job (or both) depends on how I represent my own activity to myself, that is, depends on the law of action under which I move, on the principle I have chosen and in accord with which I seek to act. The nature of my action, the true description of it, at various levels, depends on the guiding law I represent to myself. As we have just seen, none of this need be conscious (indeed, in running from a fire, we can hope to be spared much conscious representing to self), and for this reason, discovering the law that actually governs action in any particular case will not always be straightforward; as we mentioned above, there are, for Kant, in-principle limits to how well we can know the principles of our own actions. But the point here remains: on Kant's view, as on many twentieth-century views of human action, actions are the actions they are in virtue of agents' internal representations of what they are doing; that is, accurate descriptions of actions make essential reference to agents' internal principles.

Notice that the second point marks a sharp distinction between human action and the movements of non-rational beings. Indeed, the first and second points together – that practical reasoning is often unconscious and that all action as such is rationally guided – give us a picture of human action as inevitably rational, that is, as inevitably the result of a process that employs and is responsive to reasons and meanings. A version of this thought has deservedly received significant attention in the literature, as

the thought that human action is characterized by ‘reasons all the way down.’¹⁶ To say this is to say that there is no point at which actions, properly speaking, are strictly dictated by ‘brute’ mechanism. Our actions may not always be undertaken *for the best* reasons, but they are always, insofar as they are actions at all, undertaken *for some* reason. A closely related point is that, on the picture here, action is also ‘chosen all the way down.’ Any time there is action, properly speaking, and not mere behavior, something in us has taken up and endorsed grounds or laws for action (see REL 6:24).¹⁷ This ‘taking up’ remains radically underdetermined by all the available reasons – we remain free in the face of reasons – but we nonetheless take up *for* reasons. Here we see a way to ground one of the claims I mentioned at the outset. To say action properly speaking is thoroughly rational and thoroughly free is to say that rationality and freedom, in the basic forms we have encountered thus far, are pervasive. Action, as such, is always the product of a will at once rational and free, at least in essential basic senses.

We began this chapter with the causal power, common to all living things, of desire – of being able to be, “by means of one’s representations the cause of the objects of those representations” (MS 6:211). We saw that in rational creatures, such as ourselves, this causal power, for Kant, takes the form of a capacity for choice, or *Willkür*. We saw how *Willkür* is part of the broader capacity Kant calls will, *Wille*, or practical reason as a whole. Will as a whole is the source both of human action and of the rational organization of that action – it is the site of reason made practical or action-guiding. We have also seen that the will thus far described is rational and free in ways that are pervasive and ineliminable. The [next chapter](#) says more about the principles that describe action’s rational organization by the will, that is, about the rules and laws rational wills represent to themselves, about the

¹⁶ The phrase is Barbara Herman’s (“Leaving Deontology Behind,” in *The Practice of Moral Judgment* [Cambridge, Mass.: Harvard University Press, 1993], 229). Herman there and elsewhere argues that any maxim contains, if only implicitly, a reason for acting – that having a reason to act according to one’s maxim is part of what it is to have that maxim (“Leaving Deontology Behind,” 220–4; “Moral Deliberation and the Derivation of Duties,” in *The Practice of Moral Judgment*, 143–4). The ‘reasons all the way down’ point has also been made, in another form, by Henry Allison. Allison argues that principles are adopted, made maxims, because they answer some interest of the agent, or, to put it another way, because the actions they propose are in some way justified in the agent’s eyes (Allison, *Kant’s Theory of Freedom*, 87–91). This picture is also suggested by Allison’s ‘Incorporation Thesis’ (*Kant’s Theory of Freedom*, 40–1).

¹⁷ See again Herman and Allison, *ibid.* Note that while reasons go all the way down, so does the possibility of acting on bad or immoral reasons. Kantian ‘radical evil’ names the deeply rooted propensity among human beings to allow self-love’s claims to rival those of the moral law. ‘Reasons’ are not always or only approved by pure practical reason, but can also serve self-interest. (Thanks to an anonymous reviewer for asking me to clarify this point.)

structure of the space of reasons and meanings just mentioned above. This will bring us to a discussion of maxims, imperatives, and the moral law, among other things.

Before closing here, it is worth noticing what an impressive faculty we have begun to sketch. The faculty of desire, or, in rational creatures, will – the faculty that desires, reasons, and chooses – accounts for some extraordinary things. To begin, the bare fact of desire in the world organizes energies in new ways – desires set organisms to work *toward* things. Desire therefore brings whole new modes of being and moving into the world. Yearning, longing, striving, hoping, needing, wanting, lacking, missing, loving, trying, urging, and so on, find places in the order of things where they were not before. Movement organized by desire, movement that is goal-directed, must be understood very differently from movement that results from external, mechanical ‘pushes.’ When the faculty of desire is also rational and free – when it is a will, properly speaking – this movement becomes even more significantly different from mechanically caused motion: intentional, free action makes its debut. Such action, as Kant well knew, threatens the supremacy of Newtonian physics, challenging any effort to explain all motion in terms of impersonal laws acting on inert matter. In an effort to cope with this threat, Kant sought, at once, to assimilate desire to a framework of mechanistic causality (as we noted earlier; see note 1) *and* to quarantine the most metaphysically troubling aspect of rational desire, namely its freedom, in an order of things not governed by Newtonian law, viz., the noumenal realm.¹⁸ I don’t think either effort was entirely successful, but, thankfully, deciding this question is not our task here. Here, we need just to begin to notice the way free rational desire, for Kant, is the source of something very different, something entirely distinct, from all that owes its being and nature to natural mechanism, to blind forces, to powers and causes that can in no sense be said to care, to hope, to want, to need, to lean toward, to yearn, to love, to hate – to forces that cannot be said to be frustrated or fulfilled, disappointed, surprised, or successful at last. We who have will, for Kant, bring into the world the possibility of forces that *do* care – that hope, want, need, lean toward, yearn, love, hate, and so on – and consequently of effects that represent fulfillments or disappointments, successes or failures, that embody our ideas, or fall short of them.

¹⁸ See Chapter One, p. 16, for a preliminary gloss on ‘noumenal’; also see [Chapter Four](#), p. 13 ff. [for final pages, search for first appearance of ‘noumena’ in each chapter]

CHAPTER 3

A sketch continued: the structure of practical reason

In this second stage of a three-stage effort to sketch the Kantian will, we turn to the structure of practical rationality, that is, to the structure of the will insofar as will is, for Kant, reason in its practical employment. This stage of the effort will take us through the architecture of Kantian practical reason. We will look at the kinds of representations and propositions, and at the kinds of possible relations between them that, taken together, constitute practical reason's basic structure and internal logic.

WILL AS PRACTICAL REASON: PRACTICAL RULES, LAWS, AND PRINCIPLES

A first thing to do is to clarify the many terms that figure in Kant's descriptions of practice, or action. Practice (or action), properly speaking, consists for Kant in movement according to rules we represent to ourselves – we saw this in [Chapter 2](#). But what does Kant mean by 'rule'? 'Rule' is, for Kant, a fairly all-purpose term. *Rules*, or, in German, *Regeln*, either describe or prescribe regularities.¹ Plants grow toward the sun; the first word of a sentence is capitalized; parking after 6 p.m. is by permit only. Each of these is a kind of rule. Now rules, as both we and Kant use the term, can admit of exceptions. Fungi, though plants, spurn sunlight; sentences in many non-Latin alphabets, in some other character sets, and in many people's emails, do not begin with capitalized letters; parking regulations may be suspended during special events. We might say that rules can apply *contingently* or *locally*. *Laws*, or *Gesetze*, in contrast, are rules that apply, for Kant, with universality and necessity (G 4:416, 420; KpV

¹ Some etymology: 'rule,' like the German *Regel* (rule), comes from the Latin *regula*, meaning a straight piece of wood, or a ruler – hence, a rule, a standard, and the cognates, 'regular,' 'regulate,' as well as the political 'rule' (*regis*). Etymology of course proves nothing, but it may help to keep ideas clear, and it is interesting besides.

5:33–4).² Because they apply with universality and necessity, laws, for Kant, are ‘a priori,’ or are the products pure reason. What does this mean and why is it so?

Rules, ideas, and so on are ‘a priori,’ as we began to see in [Chapter 2](#), if their truth warrants, their justifications, do not depend on sensory experience, that is, on experience of the spatio-temporal world of external things. ‘A thing cannot be and not be’ is true a priori. It is true, it holds, in virtue of ‘a relation of ideas,’ to use Hume’s expression.³ ‘A thing cannot be and not be’ is, to put it another way, a conceptual truth, and as such is true independent of empirical investigation.⁴ (Rules, ideas, and so on that *do* depend on sensory experience, that is, on empirical investigation, are, as noted above, known or discovered ‘a posteriori.’) Anything known a priori is thus, for Kant, a deliverance of pure reason. Now, laws are rules that apply with universality and necessity. Kant accepts Hume’s argument that rules that apply with universality and necessity cannot, *qua* universal and necessary, be discovered a posteriori, or in experience; no experience, Kant agrees, can ever be adequate, since experience is finite, to assure us of universality and necessity. Kant therefore insists that any law, any rule that holds with universality and necessity, must be discovered a priori.⁵ Laws, then, are rules that are universal and necessary and, as a priori, must derive from pure reason.

What about *principles* (*Prinzip*)? Kant often uses ‘law’ and ‘principle’ interchangeably, though principles, for Kant, are not always wholly a priori. Thus, Kant will write about laws as the ‘supreme’ principles of a system, where the system may also include subordinate principles that take their basic form from laws but incorporate significant empirical information.⁶ ‘Principle’ often serves as a general name for rules, principles, and laws of all sorts.

² More etymology: ‘law’ comes from Anglo-Saxon *lagu* or ‘laid,’ as in ‘laid down,’ ‘fixed,’ or ‘set.’ It is the same, *mutatis mutandis*, in German: *Gesetz* (law) comes from *setzen* (to set).

³ David Hume, *An Enquiry Concerning Human Understanding* [posthumous], ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1975), §4, p. 25.

⁴ Of course, the story gets more complicated, and for Kant, as for others who use the term, there is a distinction between pure a priori truths, the entire contents of which are a priori (e.g., ten times a hundred million is one billion), and a priori truths that are ‘about’ things known to us only empirically (e.g., a nephew is a male child of a sibling). Kant discusses this distinction at B3ff.; the examples are mine.

⁵ For Hume’s argument, see Hume, *Enquiry*, §4, parts I and II, pp. 25–39. For Kant on Hume’s argument, see Kant A95/B127.

⁶ So, for instance, the law of causality – that every event has a cause – which for Kant is a priori, lends universality and necessity to Newtonian laws of motion, which are not wholly a priori but incorporate empirical information (see B20n).

So much for basic definitions of ‘rule,’ ‘law,’ and ‘principle.’ A rule or law or principle is *practical* when it describes or prescribes the actions of a will – when, in other words, it describes or prescribes *practice*.⁷ The content of a practical rule, principle, or law will be a more or less generic specification of action and the conditions under which it is undertaken: bills are paid when they arrive; when someone is in need, help is offered; the right thing (whatever it is) is always done. These formulations are a bit awkward because we are used to encountering practical rules, principles, and laws in their prescriptive forms, that is, as *imperatives*: ‘pay the bills when they arrive,’ ‘offer help to those in need,’ ‘do the right thing.’ We will have more to say about imperatives below. The thing here is just to get a sense of the underlying form of practical rules, laws, and principles, and what their content might look like. This allows us to turn to practical rules as they are taken up by us, or, as *maxims*.

MAXIMS, OR SUBJECTIVE PRACTICAL PRINCIPLES

When a practical rule or principle or law is taken up by an agent as action-guiding, it becomes, for that agent, a *maxim*. A maxim, in other words, is just a principle that, from the subject’s point of view, in fact guides the subject’s own actions. Thus does Kant call maxims ‘subjective principles of action’ (G 4:420n; see also REL 6:21). To call something a maxim is to distinguish it from the many practical principles we can ‘represent’ to ourselves *without* their guiding action: I can have the ideas that ‘bills are paid when they arrive’ or ‘when someone is in need, help is offered’ without these in fact ever serving me as maxims – I can reject them as maxims, on whatever grounds. The thought here echoes an earlier claim, viz., that to act on a principle one represents to oneself is also always to have chosen or adopted that

⁷ It is perhaps worth noting here that in his introduction to the *Critique of Judgment*, Kant writes that not all practical principles belong to *practical philosophy*, properly speaking. Some practical principles, such as those describing how to construct a geometrical proof, or perform a chemical experiment, or indeed, how to maintain a good diet (KU 5:172–3), are really, Kant writes, “corollaries” of theoretical philosophy (KU 5:172). This is because, though practical in the sense we are interested in above, they “rest on sensible conditions” and produce effects that are “possible in accordance with natural concepts of causes and effects” (KU 5:173), rendering them part of the branch of philosophy that deals with such, namely theoretical philosophy. This distinction, however, need not trouble us at present. The will is involved wherever practical principles govern action, whether their content ultimately belongs properly to theoretical or to practical philosophy.

principle. We can now put it thus: to act on a principle is to have actually made it one's maxim.⁸

What are some examples of maxims? Kant's four famous *Groundwork* illustrations provide a few. (As it happens, each is a maxim that goes *against* duty – Kant's aim in the discussion is to show how each violates the moral law.) The maxims Kant identifies in the *Groundwork* illustrations are as follows:

From self love I make it my principle to shorten my life when its longer duration threatens more troubles than it promises agreeableness. (G 4:422)

When I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen. (G 4:422)

Let each be as happy as heaven wills or as he can make himself; I shall take nothing from him nor even envy him; only I do not care to contribute anything to his welfare or to his assistance in need! (G 4:423)

In the fourth illustration, Kant doesn't offer the subject's maxim as such, but we can easily reconstruct it, by paraphrasing, as:

Because I prefer to give myself up to pleasure, I will not trouble myself with enlarging and improving my fortunate natural predispositions. (G 4:422)

In the *Critique of Practical Reason*, Kant offers an example in which I have "made it my maxim to increase my wealth by every safe means," which circumstances prompt me to further specify as permission to "deny a deposit which no one can prove has been made" (KpV 5:27). Earlier in the *Critique*, Kant considered the maxim, "let no insult pass unavenged" (KpV 5:19). These, of course, are also morally bad maxims, but Kant does offer some that are morally sound. "That [my] freedom ... coexist with the freedom of every other in accordance with a universal law" is a maxim Kant identifies with a commitment to justice, and says we must adopt (MS 6:382). One can make 'pursuit of morality itself' one's maxim, Kant writes (MS 6:392). Other good maxims that Kant mentions are, "love of one's neighbor in general," "love of one's parents" (MS 6:390), and "the happiness of others" (MS 6:393). One could find more, but this should give an idea.

⁸ It is, I think, precisely because adopting a maxim involves *choice* that Kant, somewhat enigmatically, writes that maxims "proceed from" the capacity for choice where laws "proceed from" will (*Wille*) itself. The full German text is "*Von dem Willen geben die Gesetze aus; von der Willkür die Maximen*" (MS 6:226). It supports this thought that at REL 6:21 Kant glosses 'maxim' as "a rule that the power of choice itself produces for the exercise of its freedom."

We may notice that, for Kant, maxims take many syntactic forms and have content that ranges from the quite specific to the quite broad. Efforts to clarify Kant's conception of maxims by identifying them with specific syntactic forms or kinds of content seem to me not very helpful.⁹ We do better, I think, to simply bear in mind Kant's own description of maxims as 'subjective principles of action' (G 4:420n), or 'subjective principles of volition' (G 4:400n) – that is, as descriptions, from the subject's point of view, of what he or she is doing. Such descriptions will, of course, include information about intents or purposes, since actions *qua* actions must have intents and purposes. A maxim will often include information about strategy as well. I make it my maxim to increase my property *by* failing to repay a debt, or to develop my talents *by* going to college, or to pursue justice *by* committing myself to protecting the rights of others, or to try to help others be happy *by* doing whatever that means for them. While not uniformly the case, it makes sense that maxims (which represent intents and actions) would generally specify an end *and* a strategy or means for attaining

⁹ A lot of work has been done in these directions. It has, I think, been motivated primarily by the untoward results one can get in applying the categorical imperative's universalizability test. The test asks, roughly, whether a maxim could successfully be held universally, that is, by everyone. If it could, it's OK. However, 'I will kill my enemies when they will be missed by no one and want to die anyway,' seems to pass the test, but no one would deem it moral; 'I will play tennis Sundays at 10 a.m.' (to borrow an example Herman borrows from Scanlon) does *not* seem to pass the test, but is surely morally acceptable (Barbara Herman, "Moral Deliberation and the Derivation of Duties," in *The Practice of Moral Judgment* [Cambridge, Mass.: Harvard University Press, 1993], 138; Tim Scanlon, "Kant's *Groundwork*: From Freedom to Moral Community," cited in Herman, *The Practice of Moral Judgment*). Many commentators have hoped that clarification about what should count as a maxim will head off such untoward universalizability test results. Onora O'Neill proposes that we think of maxims as fairly general 'orchestrating' principles ("Kant after Virtue," *Inquiry* 26 [1983], 387–405 and "Consistency in Action," in *Universality and Morality: Essays on Ethical Universalizability*, ed. by N. Potter and M. Timmons [Dordrecht: D. Reidel Publishing, 1985], 159–86); Otfried Höffe ("Kants kategorischer Imperativ als Kriterium des Sittlichen," *Ethik und Politik*, ed. Otfried Höffe [Frankfurt: Suhrkamp Verlag, 1979], 84–119) and Rüdiger Bittner ("Maximen," *Akten des IV. Internationalen Kant-Kongresses, Mainz, 1974*, ed. G. Funke and J. Kopper [Berlin: Walter de Gruyter, 1974], 485–98) would have us think of them as general long-standing 'life rules.' Barbara Herman is willing to let levels of specificity vary, but proposes that only 'generic maxims,' which justify a *type* of action for a *type* of reason, be thought of as subject to moral review by the universalizability test ("Moral Deliberation and the Derivation of Duties," 147–8). In my view, the real answer lies not in finding a way to specify what a maxim is, such that only bad ones will fail and good ones pass the universalizability test, but in rethinking the meaning and use of the universalizability test itself. Whether a given maxim, or subjective practical principle, is universalizable and so morally acceptable, will never be a matter of mechanical application of a test – it will always be a matter that requires subtle investigation into underlying grounds for action, a general sense of what morality demands, of why universalizability matters to it, and of what therefore needs investigation, 'universalizability-wise.' The upshot? Maxims may take all shapes and sizes. Those who want to think more about maxims should turn to Herman's extremely useful discussion in "Leaving Deontology Behind," in *The Practice of Moral Judgment*, 217–24. For more on how I think the universalizability test should be understood and applied, see Chapter 4 below.

it. But it need not be so – see ‘love one’s parents,’ which contains an end and a general strategy at once. What all maxims certainly do, however we figure to express them, is represent some end and the action we undertake to realize it.

A few additional notes about maxims are in order. First, *qua* subjective principles of action, maxims reflect “the conditions of the subject” (G 4:421n). The content of my maxims will reflect my understandings and beliefs, be they true or false, partial or far-seeing. See the following comment, which appears in Kant’s *Lectures on Ethics*: “Any objective rule says what ought to occur, even if it never does. The subjective rule says what actually does occur, for even among the wicked there are rules by which they act” (VE 27:244).¹⁰ The result is that maxims are subjective not just in that they have been endorsed by subjects and are actually action-guiding for them, but also in that they will suffer whatever limitations and distortions beset the subjects whose maxims they are.

Second, we should note that, if we understand the nature and role of Kantian maxims as I have suggested we should here, we find that, for Kant, *all* action is maxim-guided.¹¹ Indeed, it is fair to say that for Kant actions are the actions they are in virtue [of the maxims that describe them. Whether I am spring-cleaning or ransacking the apartment (however it might look), whether I am offering an honest appraisal of a job candidate or trying to undermine a colleague (or both), whether I am window-shopping or loitering, depends on the subjective practical principle that describes my

¹⁰ See the following comment, also from Kant’s lectures on ethics: “Every man who is against morality has his maxims. A precept is an objective law, by which we ought to act; but a maxim is a subjective law by which we actually do act.” (VE 27:1427)

¹¹ This interpretation must therefore reject Allen Wood’s suggestion that some actions lack maxims (Allen Wood, *Kant’s Ethical Thought* [Cambridge and New York: Cambridge University Press, 1999], 52–3). Wood’s suggestion is motivated by his thought that weakness of the will occurs when a desire details action on a good maxim – without this necessarily involving the adoption of a ‘bad’ maxim. A full reply to Wood’s suggestion relies on a point to be made shortly, namely that adopting a maxim is the same as adopting a reason for action. I reject Wood’s suggestion because it implies that for Kant there can be action undertaken for no reason at all. Without a reason, endorsed at whatever level of consciousness and with whatever commitment, we have not action but behavior, or mere natural processes. And I suspect that ultimately, for Kant, Wood’s suggestion would have the consequence that ‘weak-willed’ actions are not free – but for Kant, this would absolve agents of responsibility for such actions, and Kant won’t make this move.

Does this mean that, for Kant, actions can never genuinely be undertaken ‘for no reason’? Such cases, if they occur, constitute for Kant cases of a diabolical will. But it is not clear they can occur, since even perversely deciding to act randomly is itself generally motivated. Whether such cases can genuinely occur or not, it is certainly the case that, for Kant, we cannot in any way understand them – any attempt at understanding action necessarily, for Kant invokes reasons for action, and so ‘unreasoned action’ is necessarily incomprehensible (see REL 6:35–7).

action.¹² What I am doing depends on my maxim; and conversely, in order for me to be doing anything at all, to be *acting*, there must be a maxim that describes that doing.¹³

Third, and to echo an earlier point, it is important to note that maxims need not be conscious, or consciously formulated (see, e.g., REL 6:20). Maxims play a role in a picture of action and will that is meant to be analogous to a picture of the natural world governed by natural laws (as just noted, see G 4:387–8; KpV 5:12). Maxims, *de facto*, describe the actions we undertake. Sometimes I am aware of the maxims guiding my actions; sometimes I am not. Sometimes I just choose clothes for school, talk to a friend, write a letter, whatever, without awareness of the maxim guiding my action. If someone asked me what I was doing and why, I could probably bring the maxim to awareness – though, as we have seen, Kant knows misidentification is possible. It should be clear that neither awareness of nor introspective infallibility about maxims is required for them to be *rational*, that is, for them to participate in the system of concepts, meanings, and reasons that is in play whenever there is human action (as opposed to mere behavior). Indeed, I am rarely consciously aware of all the rules and concepts I am employing as I go about daily activities. I enter a new building, having applied the concept ‘door’ to a particular kind of opening; I phone

¹² It is of course true, as was remarked earlier, that it is often hard to tell under what subjective principle an action is undertaken – but this is an epistemological problem, and does not, for Kant, threaten the picture of individual actions as structured and described by maxims. It is also true that actions often have unintended effects, and in this sense, I may be ‘doing things’ via action that I have not represented to myself at all. But these doings are not the action itself – they are its effects, and do not belong, at least not on this Kantian picture, to description of the action I undertook. A more extended discussion of these claims can be found in Jennifer Uleman, “External Freedom and Kant’s *Rechtslehre*. Political, Metaphysical,” *Philosophy and Phenomenological Research* 68:3 (May 2004), 578–601. The most damaging objection to the account I offer is that it renders individual wills authoritative (even if subject to epistemological limitations) about the proper meanings or complete conceptualizations of their actions. But as we know, for example from the case of remarks ‘not intended to offend,’ actions participate in a shared social world of practices and meanings, and actors can be wrong about what their actions consist in – even where they are not wrong about what they intended. It may be impossible to answer this objection on Kant’s behalf, since the objection ultimately demands recognition of a radically anti-atomistic social ontology foreign to Kant’s thinking.

¹³ It is often helpful to bear in mind that Kant conceives practical principles on analogy with the principles or laws that belong to scientific theories (see, e.g., G 4:387–8; KpV 5:12). In both the scientific and practical cases, an object or event is significantly *determined*, is the kind of object or event it is, in virtue of the principles that govern it. A thing is a solid or a gas depending on the laws governing its matter, just as an attraction is an attraction, a disintegration a disintegration, insofar as these events fall under relevant laws. For Kant, as the discussion above has already suggested, the same is true for ‘practical’ objects and events. A gift is a gift (and not a repayment), a lie is a lie (and not a mistake) insofar as it falls under relevant practical principles – insofar, that is, as it can be understood as an act of will governed by the relevant representations.

an office from which I need information, and assume and employ myriad concepts and rules about telephone etiquette. Just as these activities do not have to be conscious to qualify as rational, so Kant does not require that the formulation and adoption of a maxim be a conscious process.

These features of Kantian maxims should make clear that it is wrong to suggest, as some have suggested, that actions undertaken out of love, for example, do not require maxims, the interface between will and action being in such cases somehow more 'direct' or 'immediate.' This suggestion makes the mistake of seeing maxim formation and adoption as a step, and here an extra one. On my interpretation, there is, for Kant, no action properly so called without a subjective representation, or maxim, under which the action falls.¹⁴ One may of course adopt a completely different theory of the will and of action, but as long as one is trying to be charitable to Kant, it seems to me we must understand maxims as the ineliminable subjective representations that make our actions the actions they are.¹⁵

We have, so far, moved from a general overview of rules, principles and laws to the above discussion of maxims, or subjective principles of action. We have seen what role maxims play (that of subjective principles of action), what they look like (many things), and that there is no action, as such, without them (however hidden they may be from consciousness). It is now time to turn back to laws, which, as we said earlier, are Kantian principles that hold with necessity and universality, and to see how they are related to maxims. In particular, we need to see what it means for Kant to write that a maxim *contains* a law, or "contains the practical rule determined by reason" (G 4:420–1n). How does a maxim 'contain' a law? The answer here is key to understanding how reason's abstract rulings connect to actual human willing, and is crucial to understanding the inner architecture and logic of practical reason.

For Kant, maxims 'contain' the most general rules of reason – laws – whenever they express, interpret, or specify them – and maxims *always* express, interpret, or specify more general laws. Examples can help make this point. A maxim of supporting my friends is (likely) an expression or interpretation or specification of the general law that we should respect others as ends in themselves. (If undertaken in a calculating and self-serving way, supporting friends could also, less happily, be an expression or

¹⁴ Though his local target is utilitarian calculation, Bernard Williams' now famous line about an agent having "one thought too many" makes the relevant objection (Williams, "Persons, Character and Morality," in his *Moral Luck: Philosophical Papers 1973–1980* [Cambridge and New York: Cambridge University Press, 1981], 18).

¹⁵ See note 10 above.

interpretation or specification of the general law of ‘self-love’ or ‘self-interest.’) A maxim of cheating on my taxes is (most likely), for Kant, an expression or interpretation or specification of the general law of self-love or self-interest (though if it were done to protest war spending it might be an expression or interpretation or specification of the general law that we should respect others as ends in themselves). Now, the law demanding respect for others as ends in themselves is (one formulation of) Kant’s supreme moral law. As it turns out, it is not an interpretation or expression or specification of a more general law. A law that commands self-interest is likewise as general as general laws get. The moral law and the law of self-love are the two principles at the end of the Kantian road: beyond them there are no more general practical laws. All maxims ultimately express one or the other. We have not yet said much about the supreme moral law, or its competitor – for now, I just hope the examples given make intuitive sense as expressions of such laws. The thought we need here is just the thought that any given maxim is an expression or interpretation or specification of a more general law – which it therefore implicitly ‘contains.’¹⁶

This thought may be familiar from another area of Kant’s thought. Maxims contain more general laws in the same way that empirical principles of natural science – rules describing the physics of sound, for example – ‘contain’ the more general, a priori laws of nature. Indeed, empirical principles of natural science derive whatever necessity and universality they have, that is, they derive their law-like status, precisely from the more general a priori laws they express, interpret, or specify (B14–18). The advantages of this sort of structure run deep for Kant. For Kant, lower-level principles, whether empirical scientific principles or subjective principles of human action, must advert to – ‘contain’ – laws insofar as they purport to describe kinds of causality. It is thus that local rules can describe genuine causal relations (which must be law-like). This sort of ‘containment’ structure also makes it possible for us, as enquirers and deliberators, to assume that the world and its all its motions follow a small number of a priori laws, which play out in a variety of ways. Insofar as we make such an assumption, and insofar as our enquiries and deliberations prove the assumption warranted, we find ourselves in a world that makes sense, that is, in a world to which our rational faculties are adequate. One could also put the point this way: wherever our rational capacities are adequate to the theoretical and practical tasks at hand, wherever they have proved competent, we are in

¹⁶ To use Barbara Herman’s language, any given maxim is *informed* by a law (Herman, “Leaving Deontology Behind,” 222–3). My discussion here has benefited from Herman’s work.

territory, for Kant, populated by objects and concepts that relate to each other in law-governed ways. The 'realms' described by natural science and by moral theory are, for Kant, such territories. Outside such places, if there are outsides, our abilities to make sense of things are shakier.

But enough about Kant's general view of reason and the world, and back to maxims and practical laws. Note that maxims can contain laws whether or not the subjects who adopt them self-consciously understand themselves to be expressing or interpreting or specifying laws. Indeed, as we have already seen, subjects do not always even self-consciously understand themselves to be adopting maxims. To say maxims contain laws is not to say something about how I think of them, but is to say something about what maxims *are*, namely that they are the subjective specifications of the kind of more general laws that can govern *Willkür's* causality. What those laws are, and how ordinary maxims are in fact interpretations or specifications of them, is something we take up now.

GROUNDS FOR ACTION: THE REPRESENTATION IN A
PRINCIPLE OF SOMETHING AS GOOD

"Every practical law," Kant writes, "represents a possible action as good" (G 4:414). What does this mean? Practical laws and principles do not merely describe (or command) actions. Practical principles also contain or imply *reasons* to undertake the actions they describe – in representing an action they also always represent it *as something that would be good to do*. A practical principle, with its 'good' fully represented, might be something like, 'bills are paid when they come because it is good to avoid forgetting about them and incurring late fees,' or, 'others are respected because respecting others is good in itself.' As these examples suggest, a principle may, as Kant writes, represent an action as "good merely as a means *to something else*" or as "*in itself good*" (G 4:414; Kant's emphases). The good may, in other words, be a *conditional* or 'hypothetical' good: if you want to avoid late fees, then it is good to pay the bills when they arrive. Or it may be *unconditional* or 'categorical': it is good to respect others. Practical principles, in any event, recommend themselves by promising to realize some sort of goodness.

Representing a possible action as good gives an agent a reason or ground to adopt the principle in question. We can now start to see how the Kantian will is moved. Agents act on laws they represent to themselves, and they take these laws up, acting on maxims that express them, insofar as the laws manage to recommend themselves, to present persuasive reasons for action. Persuasive to what? To me, as a whole, given my complicated set of rational

interests and sensuous desires. Our choice of maxims is a choice – executed by *Willkür* – in favor of the promise offered by some course of action, a choice of the good a practical law represents, of the reason for acting that it offers. A choice is a ‘yes, OK’ to the case a practical law implicitly makes. When we make a principle our maxim, we adopt, as the ground of our own action, the reason for acting offered by the principle.

We can now say in another way something we have been saying all along: Kantian agents are moved by reasons, and in fact, *qua* agents (as opposed to passive matter), are moved by nothing else. “Will,” Kant writes, “is a kind of causality of living beings insofar as they are rational” (G 4:446). “In such a being,” he continues later, “we think of a reason that is practical, that is, has causality with respect to its objects” (G 4:448). In acting on maxims, our actions are caused by reason(s). If it feels less strange, one might instead say that actions are caused by us, who act on reasons – but the claim is the same. As soon as I adopt a reason or ground for acting, the reason or ground is part of the proper lawful description of my causality: reasons, here, cause. Kantian agents are always and only moved to action insofar as they see something as good, that is, insofar as they take something to be a reason for action. This can be seen in the fact that, as a rational creature, an agent’s choice of maxims is subject to the demand for justification. ‘Achieving this end will satisfy a desire’ is a possible justification; ‘performing this action is in accord with duty’ is another. Kantian agents always and only move as they do because *something* has been deemed, has been taken by them to be, good.

Now, principles represent things as good – that is, present reasons – by adverting to *ends*. Earlier, we said that maxims express intents and purposes; talking about ends returns us, from a different angle, to that same idea here. (Indeed, at this point in the sketch, we find ourselves addressing many ideas again, but from slightly different Kantian angles.) The thought is this: for a practical principle to give a reason for acting *just is* for it to be directed at an end, or, for a practical principle to represent something as good *just is* for it to represent that thing as an end. The end represented in an action-guiding principle may, as we saw above, be *good as a means to something else* (G 4:414) – listening to the traffic report is good in order to avoid traffic. Or the end may be good *in itself* (G 4:414) – helping a stranger or telling the truth or otherwise acting out of respect for another’s humanity might be represented as good in themselves, and hence as ends in themselves. And ends do the job: “an end,” Kant writes, “is an *object* of free choice [*Willkür*], the representation of which determines it to action” (MS 6:384; Kant’s emphases). “Every action,” Kant continues, “has its

end" (MS 6:385). The point here is that by representing something *as the end* to be attained by an action, a principle also represents the action as good and, which is the same thing, gives a reason for undertaking the action.

As we will see in more detail in [Chapter 5](#), Kant sometimes usefully distinguishes 'objects' of action from 'grounds' or reasons for acting, reserving the term 'objects' for the empirical states of affairs to be realized by action, and understanding 'grounds' as the deepest aim of, or reason for, action (see, e.g., KpV 5:34). This distinction is useful in sorting concrete, local ends from overarching goals and aims of action. But Kant is not consistent: 'object' is also often a general term for aim or end (think 'objective' of action). Setting concerns about these distinctions aside for the time being, and using 'end' loosely to cover aims of all sorts, the idea should be clear: principles represent actions as good by adverting to ends the actions serve. All action-guiding principles, be they rules, principles, or laws, not only describe actions but advert to ends and therein offer *reasons* for action – which reasons may interest, and so motivate, a will to take the principle up as a maxim.

Before moving on, an important aside. While it is true that choices are always made, for Kant, for some reason, it is also true that choice is, for Kant, also always radically free. That is, no reason or set of reasons can *compel* my choice; I can always turn away and choose differently for other, perhaps worse, reasons. It is not good for me to turn away from good reasons: the less rationally defensible my choices, the more I approach both radical incomprehensibility and diabolical will (REL 6:35–7). But I can do it; my hand, so to speak, is never forced. It is important for Kant to insist on this, as it makes real attribution of responsibility possible. If agents were compelled, either by reason or by nature, to act as we do, there would be an important sense in which we were not authors of our actions. (This is a point made by Kant in many ways in many places, often alongside discussion, not entered into here, of the difficult metaphysical issues freedom raises; see, e.g., MS 6:213, 226–7; KpV 5:95–7.)

To sum up: we have seen here that practical principles, to be practical, that is, to be capable of motivating, always 'contain' a representation of a possible action as good, and that principles do this by adverting to ends, which constitute reasons for action. We are almost done looking, from a variety of angles, at Kantian practical principles. We turn here to our final key topic, namely toward the form practical principles can take as *imperatives*.

IMPERATIVES

Where maxims are subjective principles of action, imperatives, as Kant tells us, are ‘objective.’ What does this mean? In general, for Kant to call a practical principle ‘objective’ is for him to loose it from attachment to any particular subject: objective principles are impersonal, are tethered not to subjects but to the world of objects, that is, to the world in which subjects act. For this reason, when Kant calls a practical principle ‘objective’ he also means that it accurately describes what does (or should) happen under relevant circumstances: an objective principle is one that does not suffer from subjective distortions or limitations on knowledge or information (see G 4:420–1n). Because objective principles are accurate, in the sense just described, objective practical principles also hold for everyone (see G 4:13–14; KpV 5:19).

Objective practical principles, then, accurately describe what should be done – in such-and-such circumstance, or to achieve such-and-such a good, or in general. Being the limited and imperfect creatures we are, we sometimes set out on courses of action that do not ‘objectively’ make sense: we can be ignorant and obstinate, we are not always moved by the truth, and so on (G 4:412–13, 414). Hence, objective practical principles get couched, for us, as imperatives, as ‘oughts’: not ‘the bills are paid’ but ‘pay the bills’:

[Imperatives] say that to do or omit something would be good, but they say it to a will that does not always do something just because it is represented to it that it would be good to do that thing. (G 4:413; see also KpV 5:20)

If we were *perfectly* rational, in a full and not just a minimal or basic sense, that is, if we always did the objectively right thing, objective practical principles would function for us just as laws of nature function for non-rational beings – they would simply describe what we actually did (G 4:413–14).¹⁷ But we are not perfectly rational, and so objective practical principles present themselves to our wills as urgings, as very strong suggestions, that is, as imperatives.

Now, we have so far used the term ‘objective’ in what we might call its ‘weak’ sense. In the weak sense, a principle is objective if it is accurate (conveys factual information that agrees with objective reality) and is impersonal (is not thought of as belonging to any particular subject). However, a principle’s being accurate and impersonal does not make it

¹⁷ Henry Allison puts the point nicely when he writes that objective practical principles describe how we “would act if reason were in full control” (Henry Allison, *Kant’s Theory of Freedom* [Cambridge and New York: Cambridge University Press, 1990], 86).

matter to everyone: ‘filters in wall-unit air conditioners should be replaced every two months to reduce airborne allergens’ is an objective principle (let’s assume), but it only matters to, it ought only govern the wills of, those in charge of wall-unit air conditioners that condition air around allergy sufferers. It is objective in a weak sense. A principle is objective in a *strong* sense if it is not only accurate and impersonal, but also makes a claim on, that is, necessarily ought to matter to, everyone (see, e.g., KpV 5:19). Another way to say this, which takes advantage of the discussion above, is to say that a principle is objective in the strong sense if it represents an action as good on grounds that everyone ought to find persuasive, or if it adverts to an end everyone must (i.e., unconditionally ought to) have.

The distinction between weak and strong objective practical principles is useful because it corresponds to Kant’s distinction between two types of imperatives, hypothetical and categorical (terms employed above, to refer to kinds of goods). The terms ‘hypothetical’ and ‘categorical’ have long philosophical histories as names for syntactic or logical forms of judgment: ‘hypothetical’ judgments affirm ‘if ... then’ claims and ‘categorical’ judgments affirm ‘is an *x*’ claims. A hypothetical *imperative* as Kant develops the concept commands an agent to act in a certain way *if* she has some particular end; it commands a certain action *under the hypothesis* that the agent has as her end the object that is to be realized by the action where she may not (G 4:414), where the end is optional: it commands hypothetically. ‘Pay the bills when they come if you want to avoid incurring late fees’ is such a hypothetical imperative. Its end is optional: I might rather incur a late fee than bounce a check; I might be planning to declare bankruptcy; I might not mind paying late fees in general, etc. The action recommended by the imperative is not one I *must* (again, i.e., unconditionally ought to) pursue – the reason given for the action’s being good is not a reason I *must* (unconditionally ought to) accept.

Now, another important way to put the point is this: a hypothetical imperative will always ultimately advert to, will always represent as good, something that, for Kant, is only *conditionally good*. The good represented in a hypothetical imperative is good *conditional on* some non-necessary, non-universal feature of the agent, such as his having some further aim or end, or having particular appetites and inclinations given by his particular sensuous constitution (G 4:416, 420, 425, 427–8). A hypothetical imperative is a principle that represents an action as conditionally good since it serves an end that is itself conditionally good, that is, is good only under certain (non-necessary) conditions.

Categorical imperatives, in contrast, always ultimately represent a certain type of action *as good in itself* (G 4:414): ‘always do the right thing because it is right/good *simpliciter* to do so’ would be an (admittedly uninformative) categorical imperative. A categorical imperative represents an action as good on grounds that must (unconditionally ought to) be endorsed, and that are therefore binding on all, universally and necessarily. That is, a categorical imperative represents something as *unconditionally* good, or good in itself (G 4:428). It is thus that a categorical imperative, in contrast to a hypothetical imperative, commands, as we say, categorically. Its applicability is only contingent on the agent’s being responsive to a reason for acting to which all rational agents, universally and necessarily, should be responsive – its applicability, that is to say, is not contingent at all (KpV 5:19, 20; G 4:420, 426, 428; MS 6:385). And thus, categorical imperatives express practical principles that are objective in the strong sense. They not only accurately describe the actions one must undertake in order to realize some end, but they describe actions all *must* undertake (even if we don’t), because they serve ends we all *must* adopt (even if we don’t), period; they are ‘valid’ for (applicable to) all rational beings.

Many parts of the Kantian picture come together here. For Kant, hypothetical and categorical imperatives systematically express distinct practical laws. To see why this is so, we need to notice several things. First, we need to notice something about conditional goods, the goods ‘represented by’ hypothetical imperatives. Truly conditional goods are conditional ‘all the way down,’ so to speak. A good may *appear* conditional without really being so. For instance, ‘not lying’ in the following looks, syntactically, like it is a conditional good: ‘if you want to respect others, then you should not lie to them.’ However, the ‘if’ in ‘if you want to respect others’ does not signal a real option. Though we may not, as a matter of empirical fact, always want to respect others, and though we may, as a matter of empirical fact, often fail to respect others, this does not change the fact that for Kant we are categorically commanded always to respect others. Respect for others is not really optional, for Kant, but is unconditionally required. By extension, so is telling the truth. ‘Not lying,’ insofar as it is a way of respecting others, is a way of realizing an unconditional good. So what does it mean to say that some goods are ‘conditional all the way down’? All goods, for Kant, whose goodness depends on our having sensuous inclinations for them, that is, inclinations that derive from our natural, physical constitutions, are conditional all the way down. This is so because our natural constitutions are themselves conditional, that is, they are the complicated results of the complicated set of contingent natural facts that condition them, that

make them what they are (facts about everything from human evolution and the nature of the earth to my particular genetic makeup, the language I speak, the place I grew up, what I ate today, etc.). As a result, any goods that depend for their goodness on our sensuous constitutions are inherently conditional. And this conditionality goes 'all the way down,' or better, extends infinitely (A417–19/B445–7). Nature, for Kant, is an infinite web of mutually conditioning conditional facts (this is what it is for nature to be an infinite causal nexus), and we, as physical, natural creatures, complete with sensuous inclinations, belong to this nexus. A good is 'at bottom' conditional if its being a good depends on an inclination, a desire, a reason for acting, that itself depends on our natural conditioning.

An idea just mentioned bears repeating: our natural constitutions, our sensuous inclinations, *furnish us with reasons for acting*. The fact that I desire something is, *prima facie*, a reason for pursuing it. Where a good is conditional, my reason for pursuing it ultimately derives from nature. For Kant, every subject wants to satisfy her sensuous inclinations, big and small, sophisticated and simple, long-term and immediate. It is indeed, for Kant, a *law of nature* that we want to satisfy such inclinations: their satisfaction contributes to our own preservation as well as to our happiness (understood by Kant precisely as the maximal satisfaction of inclinations), and, as natural creatures, we necessarily desire both our own preservation and happiness (MS 6:386; KpV 5:25; G 4:444, 453).¹⁸ Kant calls this law of nature the law of self-love (KpV 5:22). Now we come to a crucial and very interesting point: if I choose to follow a hypothetical imperative, my action is guided by the law of self-love and is, in this sense – though not in any deterministic sense – *action according to a law of nature* (G 4:444). Nature has here provided me with the operative reason for action. Hypothetical imperatives express nature's law – the law of self-love – for the causality of my will, and the end I pursue is one furnished by nature. My will, *qua* will, should not ever be conceived as behaving deterministically in accord with natural laws, but it certainly can and sometimes must be conceived as

¹⁸ If there are, as there are and as Kant knows there are, people who do not wish to go on living, this is not because the law of self-love fails to govern them, but rather because they feel that the pleasure to be had by continuing to live is so outweighed by pain that they would, from the point of view of their own happiness, be better off dead. (In this they may of course be wrong, and Kant in fact prohibits suicide, but for moral reasons; see G 4:421–2.) It would indeed be interesting to explore further, from a Kantian point of view, the cases (many of them less dramatic) when sensuous desires for self-preservation oppose sensuous desires for pleasure. They might be seen as particularly telling cases of the need, as Kant sees it, for rational principles to govern inclinations, which cannot be trusted, so to speak, to cooperate even amongst themselves (see VE 27:360).

choosing obedience to what is fundamentally a natural law. This is just what I do when I follow my inclinations.

Now a categorical imperative adverts to an unconditioned end, one I ought to have regardless of the contingencies that constitute my conditioned condition. Whence such an end, whence an unconditional good? The answer is reason: reason, not nature, is capable of furnishing an unconditional end, and recommending it to us. How? If reason discovers something that can be shown good for its own sake, it will have discovered a good that is unconditioned. And if it can be shown good a priori, reason can claim it as an unconditional good that is good universally and necessarily. Of course, for Kant, reason makes just such a discovery – indeed, made this discovery long ago, though philosophers have often been slow to recognize it. Reason finds “humanity” (G 4:429) in one of Kant’s formulations, or in another “rational nature” (G 4:428), or “the good will” (G 4:393; see also KpV 5:62), or, in the formulation I prefer, reason finds free rational activity itself unconditionally good a priori. And the law that represents such an unconditional good as its end is reason’s own law. Though I just named them, or it (by various names), I have carefully avoided trying to show how or why reason recognizes what it does as an unconditional good – I want to leave those arguments and that discussion for later. I have also carefully avoided saying much about the contents or demands of what I just called ‘reason’s own law,’ which is, of course, Kant’s moral law, though these contents and demands will be well known to many readers already. It is difficult, and artificial, to want to look at parts of Kantian practical reason without tracing the ways each part leads to the whole, but there are reasons to try to do it. I promise much more on the value of free rational activity and on the contents of Kant’s moral law anon. Here, the aim is to see how imperatives work and what kinds there are, and to see, as we have just been seeing, how imperatives, as lower-level principles, ‘contain’ more general laws. The moral of the story we have just been telling is this: just like inclination-satisfying maxims, hypothetical imperatives ‘contain’ or express a law of nature, the law of self-love, and just like pure-reason-responsive maxims, categorical imperatives ‘contain’ or express a law of reason, the moral law. And, in seeing this, we also start to see a key finding: at the end of the day, all practical principles, for Kant, express either a law of nature or a law of reason.

The account of imperatives just offered is part of an account of the Kantian will as a faculty of action-guiding reason, or (which is the same thing) of reason-guided action. My hope is that readers see at this point how thoroughly Kantian will is shot through with reason, how completely it is a

capacity for desire that operates not mechanically but through conceptualizations, representations of possible principles, and attentiveness to aims and goals, how it can only be comprehended as doing what it does *for reasons* (even if some of these reasons are furnished by nature). Actions, we have seen, are for Kant the actions they are in virtue of the subjective conceptualizations and representations that guide them; actions as such are always undertaken for some reason, are always purposive, even where subjects are not consciously aware of the reasons or purposes for which they are acting. Kantian will is always reason made practical: will, as a causal power, initiates a chain of effects under the guidance of a subject's own reason. All of this indeed makes Kantian will an interesting (and metaphysically complicated) causal power. Recognizing Kantian will as a causal power moved by reasons may help us begin to grasp why will is so central to Kant's conception of our special nature and special place in the world, and why, for Kant, the human will is of such inestimable value. But before we get to that, we need to look more carefully at a final and more radical sense in which Kantian will is rational, and which will have profound ramifications for Kant's sense of our value: will, for Kant, is not only determined by reasons in a broad sense, but is, as we began to see above, determinable by pure reason itself.

PURE PRACTICAL REASON, OR THE POSSIBILITY OF A
CATEGORICAL IMPERATIVE

If the above has tried to show anything, it has tried to show that action, for Kant, is always rational in a minimal or basic sense. A way into the next thought here is to see that, while *all* action is rational in a minimal sense, some actions are nonetheless 'more' rational than others, now in a normative sense of 'rational,' that is, a sense according to which one can succeed, and in greater and lesser degrees, to live up to a standard of full rationality. Actions are *more* rational when they are supported by *more* or *better* reasons – or, to put the same point slightly differently, actions are more rational when reason, as a faculty that identifies and applies standards, that seeks justification, finds that they better meet standards or are more justified. Reason's judgment that some actions are better or more justified than others takes two important, and importantly different, forms. The first form assumes given ends. Here, reason can judge that some means to attaining ends are more efficacious or judicious or otherwise prudent than others. I am uptown and I want to get downtown; reason judges that an express train is better than a local as an efficient means to attaining this end. I have plenty of food at home and little money at the moment, and I am hungry and want

dinner: reason judges that cooking is a better means to attaining dinner than going out. Or I have plenty of food at home and little money, and I am hungry and want dinner, but a colleague with whom I would love to talk has just proposed supper at the corner restaurant: reason here may judge that going out is the best course. These cases all illustrate actions judged *more* rational than viable alternatives because they more effectively accomplish given ends, all things considered; ‘more rational’ here means ‘more prudent’ just in that sense.

The second form of judgment that some actions are more rational than others – indeed, are completely or purely rational – is judgment that some actions are aimed at ends that are themselves rational. Insofar as reason judges some ends themselves inherently rational, that is, judges that reason itself requires their adoption, actions aimed at such ends will be *rational* in the strongest possible sense, will be completely rational. In contrast, actions prudently organized to attain ends given by circumstance, by the condition I’m in, no matter how efficacious, judicious, thrifty, fun, etc., are less than completely rational because they are in themselves indifferent to the rationality of the ends they serve.

For an end to be perfectly (or completely, or purely) rational is for it to answer to interests that belong to rationality itself. That is, its appeal cannot be to the subject’s sensuously (naturally, contingently, a-rationally) constituted array of needs and desires, but must be to an interest that is internal to rationality itself. What could it mean for an end to answer *an interest of reason itself*? The idea will seem strange to readers who take for granted a view like Hume’s, according to which all ends ultimately are a-rationally given; reason, for Hume, can hope just to help us attain our ends effectively. On a view like Hume’s, judgments that something is good, justifications for actions, always ultimately bottom out in a brute given of human need or desire. This instrumentalist view of reason is of course expressed in the famous Humean dictum that, “reason is, and ought only to be the slave to the passions,”¹⁹ and dominates contemporary ‘rational choice’ accounts of practical reasoning (influential in economics as well as in philosophy). On this picture, reason is a tool for satisfying desires, not a source of desire itself. Kant’s understanding of reason is profoundly at odds with this picture. For Kant, reason is a faculty, a capacity, to which interests are integral, and which indeed has interests of its very own (MS 6:212–13; also KpV 5:79; G 4:413n, 449, 460, 460n). We could easily call these interests reason’s *desires*

¹⁹ David Hume, *A Treatise of Human Nature* [1739], ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1978), Book II, Part III, §3, p. 415.

but for the fact that Kant tries to reserve 'desire' for interests that have sensuous origins.²⁰ But whatever we call them, reason for Kant certainly has ends, aims, or goals that are its own. If the idea seems strange, consider the aims that philosopher of science Thomas Kuhn argues inform choice of scientific theory: accuracy, consistency, broad scope, simplicity, and fruitfulness.²¹ We can understand his claim as the claim that accuracy, consistency, etc., are interests that are *built into* the project and processes of scientific inquiry, that is, into scientific rationality. The suggestion here is that interests are similarly built into Kantian rationality, both theoretical and practical. Kantian reason in its theoretical employment, for instance, is guided by an interest in discovering mechanical causes of events. (This interest, unless kept in its proper domain, famously conflicts with another of reason's interests, namely the interest in completing series [see, e.g., Kant's discussion of his 'antinomies' at A405/B432ff.].)²² For Kant, practical reason likewise has interests that are internal to its project and processes. Now, interests are (or generate) ends, and ends in which reason itself is interested are *really* rational, all the way down; they are valued by reason itself, as such.

If practical reason has such an end, and only if it does, can there be, for Kant, a true categorical imperative (G 4:428). What Kant asks here is that pure reason identify an end the goodness or desirability of which does not depend on empirical facts, but which instead is taken to be good or desirable in itself, by reason, on rational grounds. Such an end will be able to ground a

²⁰ Kant's German term normally translated 'desire' is *Begehren*; 'interest' is *Interesse*. I am mostly, following Kant, in the habit of reserving 'desire' for sensuous inclination, using 'interest' to designate reason's own inclinations. Some recent commentators have admirably gone ahead, defying strict Kantian usage in favor of ordinary English usage, and attributed 'desires' to Kantian reason. Allen Wood, for instance, making a point with which I am obviously sympathetic here, writes that

Kant is sometimes misrepresented as holding that to act from duty is to act without (or even contrary to) *desire*, for Kant denies that we could ever act without desire ... Kant's view, however, is that pure reason can determine the will because it is also *a source of desire*. (Wood, *Kant's Ethical Thought*, 53 and 348n6; emphases in original)

Alenka Zupančič likewise does not hesitate to call the interests of Kantian reason 'desires' (Alenka Zupančič, *Ethics of the Real: Kant, Lacan* [London and New York: Verso, 2000]).

²¹ See Thomas Kuhn, "Objectivity, Value Judgment, and Theory Choice," in his *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), 320–39.

²² It may also be worth recalling that Kant's first words in the Preface to the first edition of the *Critique of Pure Reason* point to (among very many other things) the idea that reason has concerns (here, questions and problems) of its own:

Human reason has the peculiar fate in one species of its cognitions that it is burdened with questions which it cannot dismiss, since they are given to it as problems by the nature of reason itself, but which it also cannot answer, since they transcend every capacity of human reason. (*Avii*)

categorical imperative, and indeed the moral law itself, since it will be an *unconditioned* end. Recall: nothing in nature is unconditioned. But reason has access to unconditioneds. An end internal to reason is an end I must have, *just in virtue of being rational* – and if it is an end reason has identified as intrinsically and not just (even for its own purposes) instrumentally valuable, then it is an *unconditioned* end and its claims on me are unconditional. Such an end will not ground a hypothetical imperative – ‘if you have this end, then you ought to act in this way’ – but a categorical one – ‘you have this end, so you must act in this way.’ Reason alone is, if anything is, capable of furnishing an unconditioned end for action, and of authoring a law aimed at serving this end, and so of grounding a categorical imperative (see G 4:408, 426; KpV 5:19–20; MS 6:216–17).

Now, I will here assert again that for Kant, the ultimate, unconditioned, unconditional end of reason is the free rational activity of will itself. Kant, as I have noted, does not use such a formulation, but my argument in the remainder of this book will be that the many names Kant *does* give practical reason’s unconditionally good end (including, as we saw, “humanity” at G 4:429, “rational nature” at G 4:428, and “a good will” at G 4:393, as well as, at KpV 5:62, “a will whose maxim always conforms with [the moral] law”) as well as the contents of his categorical imperative (most famously formulated roughly as ‘act only on maxims that you can at the same time will as universal laws’ [see G 4:421 and KpV 5:30]) effectively endorse the activity of free rational willing itself as the ultimate aim we should set ourselves. Moral thought and action, as I claimed in the Introduction ([Chapter 1](#)) and as I again assert here, take place for Kant when the free rational will wills itself. When I will in accordance with the end of free rational activity, or when the end of my willing is free rational activity itself, that is, when the aim of my action is to respect, preserve, and promote free rational willing, then and only then is my action perfectly or completely free and rational.²³ Hegel was right: Kant’s categorical imperative commands the free will to will itself, with philosophically spectacular results.

Before concluding this chapter, and moving to the third part of our sketch of Kantian will, I want to clear away a concern that the promised argument cannot, in principle, be made on Kant’s behalf. Kant sometimes writes as if moral action, or action on the categorical imperative he identifies, *can have no ends*. For instance, in the *Critique of Practical Reason*, Kant

²³ Whether this actually ever happens, or whether we can be sure it has happened when it does, is something about which Kant himself has doubts (see note at A551/B579). The point here is that such willing represents the moral ideal at which we must, and Kant thinks do in fact, aim.

writes that there can only be a practical law, that is, a categorical imperative, if “all material of a law, that is, every object of the will considered as a ground of its determination, is abstracted from it” (KpV 5:27). This passage is often read as claiming that the moral law can have no object or end. There are many such passages.

Such passages militate in favor of reading Kant as claiming that doing one's duty consists in submitting to rules for their own sake, rules that serve no ideal or promise no good, save perhaps correct rule-following itself. Many have read Kant as making this claim, and accordingly Kant is often criticized for holding an unmotivated (or only very opaquely motivated), rule-obsessed view. Were this Kant's view, Kant could indeed be properly criticized for an implausible moral psychology, according to which the will can be moved by rational principles alone, but not in virtue of anything that might be identified as the good promised by those principles. Such a reading, however, elides two important features of Kant's view. First, it overlooks the many passages in which Kant discusses moral ends (e.g., G 4:428; MS 6:380–5). Second, and perhaps more importantly, it fails to notice that whenever Kant writes as if a categorical imperative can have no end, Kant means just that a categorical imperative cannot have an *empirical* end, that is, cannot have as its ultimate end some empirical state of affairs (see REL 6:6–7). He does not mean that action on a categorical imperative can have no end in the sense that there can be no *reason* to undertake it at all. Indeed, the very possibility of a categorical imperative depends precisely *on there being* a purely rational reason for acting as it commands. [Chapter 5](#) makes this case in more detail, but the grounds of my argument for this claim are already deeply embedded in the reconstruction of Kantian will offered here. As I have (I think accurately) presented it, Kantian will could never get off the ground, as it were, without objects at which to aim. Will, even at its most rational, is still for Kant a kind of desiring faculty, seeking the realization of ends, striving to make ideas realities, and doing so because it takes doing this to be good.

The moral, then, of this second stage of my sketch of Kantian will is that Kantian will, with all its desiring and striving and wanting, is also rational. It is a faculty both responsive to and productive of reasons for acting, and in several senses. It is, as we saw in [Chapter 2](#), both inevitably and thoroughly rational in that it always employs concepts and representations, and always does what it does in ways that invoke reasons and hence always make (some kind of) sense: it initiates movement that is always run through and shaped by reasons. Beyond this, it can be better and worse at responding to better and worse reasons – it can be more and less rational, in a less minimal,

normative sense of ‘rational.’ It is also, at least in principle, capable of achieving complete, pure rationality. Will, as pure practical reason, can identify a purely rational good. When it does this, will gives itself not only *some* reason for acting, but a reason for acting that owes its being not to nature, not to anything contingent, but to practical reason itself. This brings us right to our next discussion. Where the first stage of this three-stage sketch of will focused on will as a kind of activity (as causality organized by desire), and where this second stage focused on will’s rationality, the third stage, to which we turn now, looks at its freedom, which significantly includes the will’s ability to act on grounds not given by anything external to itself.

CHAPTER 4

A sketch completed: freedom

AN OVERVIEW OF THE FREE KANTIAN WILL

The third and final stage in this initial sketch of Kantian will is focused on the will's freedom. For Kant, freedom in the most central sense consists in being self-determining, that is, in being a kind of causality whose determining ground is internal and not external (see, e.g., A444–6/B472–4). Given the ground we have already covered, we can easily see that, for Kant, a will determinable by reason itself, that is, a will that can give itself an end and a corresponding law of action, is also for this reason and at the same time a free will. The aim of this short [final chapter](#) on the will is to say this again, in greater detail, and to show how it is related to some of Kant's more famous discussions of freedom and the good will.

Despite the brief gloss just given, freedom is a complicated business for Kant. Just as there is for Kant a sense of rationality according to which rationality pervades all action *and* a sense of rationality according to which rationality is something to achieve, and can be achieved more or less, so there is a sense of freedom according to which freedom pervades all choice and all action, *and* a sense of freedom according to which freedom is something to achieve, and can be achieved more or less. Just as determination by pure reason is the pinnacle of 'achievable' (non-inevitable) rationality, so complete freedom, or determination only on internal grounds, would be the pinnacle of 'achievable' (non-inevitable) freedom. Freedom in the inevitable sense, according to which *all* choice and *all* action is free, follows from the fact that the will is always underdetermined by the grounds or reasons present to *Willkür* (see KpV 5:29). We are, in contrast, free in the 'achievable' sense – choice and action are more fully, or truly, or completely free – to the extent that will is in fact determined by (pure) reason. Here, we are more fully free because the grounds on which we act are more truly our own – we are more self-determining (see KpV 5:29).

Both senses of freedom are crucial. Freedom in the first sense – freedom that is inevitable for us – underwrites our belief that action is not predetermined and beyond our control. *Willkür*, the capacity for choice, determines, but until it determines, I could act in any of several ways. Freedom in this sense is required, Kant thinks, if we are to hold people responsible for their actions; if actions were predetermined, or followed an inevitable trajectory, if we were just particular locations (no matter how complicated) on infinite causal chains, people would be no more responsible for their actions than clouds are responsible for raining (KpV 5:96–7).

The second sense of freedom – as something that can be achieved more or less fully – has less to do with freedom in the moment of choice and more to do with choosing in a way that sustains freedom. In the second sense of freedom, I am *fully* free when I choose in accord with grounds that are mine ‘all the way down,’ grounds that come from my reason itself, as opposed to grounds that ultimately have an external source: when I choose on grounds that are mine ‘all the way down,’ I self-determine in the fullest sense (see KpV 5:33). The significance of this is far-reaching: when, for Kant, I choose to act on *reason’s* reasons, so on reasons that are most truly mine, the moment of (free) choice passes, but rather than slumping back, so to speak, into the order of nature and acting in ways that can be chalked up (ultimately) to natural mechanisms, freedom – now as self-determination – is sustained. In choosing self-determination, I act freely (in the first sense) in a way that sustains freedom (in the second sense).

The two senses of freedom are like the two senses of reason not just in being two, and not just because one describes something ubiquitous and ineliminable and the other something to strive for. The likeness extends to the way the two senses are related. Kant could have chosen different words for the first and second senses (of each), but instead ran the risk of confusion and conflation by sticking with one term (for each). This is because the senses *are* intimately related. In each case, the first sense gets the whole thing off the ground, makes it possible, gets the project – of rational action, or of free action – going. The second honors and perfects the same capacity by taking it as far as it can go, giving it a chance to ‘amount to something.’ It would not be quite right to say that the first sense describes a potential and the second its realization, though there is something right there too. The first describes a sheer capacity, a capacity that already distinguishes us sharply from creatures that belong wholly to nature; the second describes what can happen when that capacity is itself further elaborated and treated as something to be valued and pursued for its own sake.

THE FREE KANTIAN WILL IN MORE DETAIL

Inevitable freedom

Let us turn now to look more slowly at each of these claims, as well their relations to Kant's well-known discussions of negative freedom, positive freedom, and autonomy. We turn back first to freedom in the first sense. Recall the idea that any determination of will ultimately rests on a choice, on *Willkür*. Sensible impulses and inclinations, which are what they are thanks to nature (to our bodily constitutions and to other natural facts), offer us reasons to pursue certain objects. But these impulses do not themselves determine the will. "Human choice," Kant writes, "is a capacity for choice that can indeed be *affected* but not *determined* by impulses" (MS 6:213).¹ Likewise, the proposals of reason fail to ineluctably determine our wills – people can, and routinely do, turn away from the most rational grounds for action and choose on what even they deem to be less rational grounds. So what determines a will in the end for Kant? *I* do, in some way that can never, for Kant, be fully explained, but that also must be granted. We may sometimes *feel* compelled, as if the objects of our wills themselves, by some natural law, make us act (see Kant's description of 'concupiscence' or 'lusting after something' at MS 6:213). But, for Kant, until we decide to pursue a given object, until we endorse an inclination on the one hand or an end set by reason on the other, until we adopt a maxim that represents some object as choice-worthy, our wills are not, for Kant, determined.² This radical freedom of the individual will must be granted, because to deny it, or even to remain agnostic about it, is to deprive oneself the intellectual resources needed to hold people, including oneself, responsible for actions.

Kant's 'transcendental' idea of freedom is in play here. For Kant, a transcendental idea is one reason is pressed to posit, and is justified in positing, because rational reflection and argumentation show it to be a necessary condition of something else that is given and undeniable. Kant's

¹ The following passage from one of Kant's precritical texts expresses the idea more evocatively:

[M]ortals commit sins voluntarily and as a result of an inmost state of mind, for the chain of antecedent grounds does not hurry them along or sweep them away against their will; it attracts them. (PPC I:404)

I owe awareness of this passage to Martin Schönfeld, *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford University Press, 2000), 157.

² These comments owe much to accounts of end adoption offered by Henry Allison (*Kant's Theory of Freedom* [Cambridge and New York: Cambridge University Press, 1990]) and Barbara Herman (*The Practice of Moral Judgment* [Cambridge, Mass.: Harvard University Press, 1993]).

transcendental idea of freedom is an idea reason is pressed to posit, and is justified in positing, because rational reflection and argumentation show it to be a necessary condition of attributions of responsibility, or 'imputation' (see A448/B476 and A533/B561; see also REL 6:21).³ The transcendental idea of freedom is precisely the idea of an inexplicable gap between possible determining grounds and determination itself, a gap I fill, a gap my *Willkür* fills, though we know not quite how. The transcendently free will, or *freie Willkür*, is a sort of uncaused cause, a "faculty of beginning a state from itself" (A533/B561; see also A541/B569). If the 'just how' of this determination were known, if the event were fully explicable, it would, for Kant, be locatable within a deterministic causal order – but this location is just what free will resists. Free will initiates actions and effects, without this initiation being itself the effect of any cause.

I said earlier that this sort of freedom is ubiquitous – all choice and all action are free in the sense that all agents choose among potential grounds for action, none of which inevitably determines the will. Is this true even in extreme cases? What if I am threatened – are my actions free then too? In the sense at issue here, absolutely yes: I remain free in the face of coercion. In his legal theory, Kant will introduce a notion of freedom – 'external freedom' – with which external compulsion is not compatible (MS 6:214). One will therefore ultimately have resources, within Kant's overall view, to argue not only that coercion is wrong, but even that it infringes on freedom.⁴ But one cannot claim to lose the freedom at issue *here* in the face of coercion. In the *Critique of Practical Reason*, Kant offers the example of a man who is threatened by a ruling prince with "immediate execution" unless he gives "false testimony against an honorable man whom the prince would like to destroy under plausible pretext" (KpV 5:30). "He would perhaps not venture to assert whether he would do it or not," Kant writes, but the man, and anyone putting himself in the man's place, "must admit without hesitation," Kant argues, that he *could* choose to tell the truth (KpV 5:30). For Kant, neither princes nor nature nor anything else can absolutely compel us. All human action is free in the radical sense that it always issues

³ For Kant, the transcendental idea of freedom is also necessary to 'ground' appearances, that is, to make possible that the 'world of sense,' the order of nature, is as it is through some free intelligible cause (see A537/B565). The idea here is that reason needs accounts not only of how (causally) things come to be but also of why (rationally) they come to be as they do.

⁴ For more on Kant's legal notion of 'external freedom,' and its relation to freedom of the will, see Jennifer Uleman, "External Freedom in Kant's *Rechtslehre*: Political, Metaphysical," *Philosophy and Phenomenological Research* 68:3 (May 2004), 578–601.

from a choice between grounds of action, heavy and great as the push or pull of some grounds may be.⁵

'Achievable' freedom

So much for the first sense of freedom. In the second sense, freedom is far from inevitable and ubiquitous, but is instead something to be achieved. A will that can give itself its own principle, while certainly radically free in the first sense, is also capable of being free in the 'fuller' second sense that it acts on home-grown grounds of action, that its self-determination consists not in mere choosing 'for some reason,' but in choice of and then adherence to reasons that are most deeply *its own*. When I choose to act on a principle I have myself authored, in the service of a good I myself commend to myself, I remain free 'all the way down.' My choice itself is self-determined, as choice always is, but it is also a choice on reasons that are deeply my own, rather than a choice to order my actions in accord with an aim or good fundamentally external to me. In the second sense, to say a will is free is not just to say it is *able to choose* a rational law, but to say that it *has chosen* determination by reason. When I act freely in the second sense, I am living out the deepest possibility that freedom in the first sense offers: I have achieved action that issues purely from myself.

Kant's 'negative' and 'positive' characterizations of freedom

What I am describing as first and second senses of freedom is largely captured by Kant's famous 'negative' and 'positive' characterizations (Kant calls them 'definitions') of freedom. Kant's negative characterization invites us to focus on what the free will *is not*, namely, determined by 'alien causes' or by something external to itself (G 4:446). Kant's positive characterization draws attention to ways we can think about what the free will *is*, namely, Kant tells us, *determined by reason*. These characterizations are meant by Kant

⁵ Of course, as we have said above but as bears repeating here, human bodies, *qua* bodies, are also subject to natural laws about which we have no choice. Many of the movements of my body fall outside the jurisdiction of my will: my stomach and my heart, for example, move in ways I do not control, and I follow the law of gravity just as a stone does. But when I follow such laws, I am not *acting*; the arena of the will's jurisdiction is precisely the arena of action as such. (It is of course extremely interesting to look at the boundaries of this jurisdiction, including at limit cases, that is, at cases where will seems to extend further than common sense would guess, or where it seems more restricted: abilities to willfully regulate and heal the body on the one hand, addictions and compulsions that defeat will on the other. But interesting limit cases notwithstanding, we all understand the idea that some bodily movements and behaviors are under will's control and some are not.)

to capture two sides of a single coin – I don't think Kant sees them as distinct species of freedom, but rather as characterizations that emphasize different aspects of one complex thing, the freedom of the will. Nonetheless, playing out their different emphases can generate senses which correspond well to freedom in our first and second senses (which are also ultimately intimate).

In the *Groundwork*, Kant offers the following “definition of freedom” as “negative” (G 4:446):

Will [*Wille*] is a kind of causality belonging to living beings insofar as they are rational, and *freedom* would be that property of such causality that it can be efficient independently of alien causes *determining* it, just as *natural necessity* is the property of the causality of all nonrational beings to be determined to activity by the influence of alien causes. (G 4:446; Kant's emphases)

In the *Critique of Practical Reason*, Kant describes “freedom in the negative sense” as “independence from all material of the law (i.e., a desired object)” in the determination of the will (KpV 5:33), and in the *Metaphysics of Morals* he writes that “*freedom of choice is this independence from being determined by sensible impulses; this is the negative concept of freedom*” (MS 6:213). Each of these negative formulations says that a will is free if it is *not* compelled, determined, or necessitated by something other than itself, most often nature or natural desire. The connection to freedom in our first sense should be evident. We are (‘negatively’) free if we are free *not* to pursue objects that are attractive on grounds of self-love – that is, if we are not compelled to pursue, among other things, the urgings of inclination, of ‘sensible impulse,’ of ‘desire’ in the narrow, sensuous sense (see A534/B562).

The positive characterization, in contrast, takes us toward being able to say what free will *is*. This is not easy, for Kant, since he is unwilling to describe free wills as he is used to describing other things, viz., by elaborating their behavior under natural laws. If we are going to construct a positive characterization of free will, a characterization of what it is and not just of what it is not, we need to think in other terms, and specifically in terms of laws *other than* natural law. Kant's positive characterization accordingly describes free will as “causality in accordance with immutable laws but of a special kind” (G 4:446):

Since the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with natural laws, is not for that reason lawless but must instead be a causality in accordance with immutable laws but of a special kind; for otherwise a free will would be an absurdity. (G 4:446)

As a kind of causality, will needs to follow *some* law or other (see A539/B567). If free will does not deterministically fall under nature's laws, and if we are going to think about it without thinking an absurdity, we have to begin to think about it as following some other 'special kind' of law. But the alternative to determination by natural law ultimately can only be determination by the law of reason, or the moral law; laws of nature and laws of reason *are the* Kantian alternatives (see A532/B560). Here is where the 'positive' sense of freedom emerges. To be free, in this sense, is to give oneself, that is, to have one's pure reason give oneself, a law. It is to have reason itself 'legislate,' or 'be practical'. In the *Critique of Practical Reason*, Kant writes that the, "intrinsic legislation of pure and thus practical reason is freedom in the positive sense" (KpV 5:33); in the *Metaphysics of Morals*, "the positive concept of freedom is that of the capacity of pure reason to be of itself practical" (MS 6:213–14).

But didn't we just say that a radically, or transcendently, or inevitably free will is not compelled by *any* law? We did – and it is not. But this does not mean, for Kant, that it can be lawless, at least not insofar as it is *thinkable*, as we can think about it in a meaningful way. Any time a thing causes, it follows a law – it is just that, if the thing is free will, *it* chooses the law. Freedom in Kant's positive characterization is free will exercised in favor of a practical law of reason, and explicable, causally speaking, in terms of determination by reason.

The difference between Kant's negative and positive characterizations of freedom is, then, close to the difference between our first and second senses of freedom. The negative emphasizes the fact that *I*, and not something else, determine my will; the positive looks at what determining grounds I can really call my own: both concern self-determination. But where the negative characterization emphasizes openness and choice, and describes an inevitable, ineliminable, pervasive feature of our wills, Kant's positive characterization emphasizes a possibility that is neither inevitable nor pervasive, but at which we can succeed or fail, and by degrees: it describes determination by pure reason.

Kant writes that the positive conception of freedom is "richer and more fruitful" than the negative (G 4:446). What does Kant mean? First, the positive characterization makes freedom of the will *make sense*. We might, of course, think of a free will 'merely negatively,' in terms only of what does not necessarily determine it. But for Kant, this would be to conceive it without law or principle – and will conceived 'lawlessly' is, for Kant, deeply incomprehensible. Will is, after all and as we have said, a kind of causality for Kant. Even as it remains impossible to observe the operation of 'free

causes' – since such would violate the laws governing the observable world of spatio-temporal objects in a Newtonian order – and hence even as it remains impossible, in Kant's terms, to theoretically 'know' or 'cognize' free causes as part of that world, the possibility of coherently *thinking* free causes (through the category of causality) requires that we think them in terms of law (G 4:446).⁶ Genuinely lawless freedom 'operates' only in cases of a diabolical will – but these are precisely cases where the causes or grounds of action remain *unthinkable* for Kant (REL 6:21; KpV 5:29). The positive characterization is fruitful, then, first, in making free causality make sense, in making it thinkable.

Second, Kant's positive characterization of freedom is fruitful insofar as it provides a way into the derivation of the moral law – and a way into this makes the characterization rich and fruitful if anything does. The fact that reason alone, without any assistance from nature, gives a law is the starting point from which Kant derives his first formulation of the categorical imperative. The hallmarks of this formulation – 'act only on maxims that you can at the same time will as universal laws' – are that it denies natural incentives pride of place in determining the will and demands law-like-ness itself of maxims. Further discussion of the derivation and contents of the moral law will be found in [Chapter 6](#) below. What we need to note here is just that the positive characterization of the free will as rationally self-legislating provides Kant with materials to derive the substance of the moral law. This makes it, as we said above, fruitful indeed.

Autonomy

It now makes sense, not only in terms of Kant's own texts, but also in terms of our exposition, to come to Kant's famous concept of autonomy.⁷ Kant glosses 'autonomy' as the "will's property of being a law to itself" (G 4:447; see also G 4:433). The term is from *auto*, 'self,' and *nomos*, 'law': to be autonomous is to give a law to oneself, to rule or legislate for oneself. To be *able* to do this is to be both free and rational in our minimal, inevitable senses. To actually do it – to achieve autonomy – is to approach full determination by reason, and hence full self-determination, and hence freedom in the fullest or most perfect sense. As we will see in more detail in subsequent chapters, a will that is autonomous is also, for Kant, a morally

⁶ On the contrast between 'knowing' or 'cognizing' and 'thinking,' see Bxxvi.

⁷ Autonomy is introduced in the *Groundwork* alongside Kant's negative and positive characterizations of freedom – see G 4:446–7.

good will. ('Kantian autonomy' accordingly receives wide attention in the literature, and moral theories that stress the value of autonomy nearly always claim a Kantian heritage.) The term opposed to 'autonomy' is 'heteronomy,' or rule by another. A will that chooses on natural (or other external) grounds makes grounds *not* deeply its own determinants of the will: such a will chooses 'heteronomously.'⁸ Heteronomous choices, choices grounded in *others'* grounds, are still choices; the heteronomous will is still free will in our first sense. But autonomous choices, choices ultimately grounded in our own reason, exhibit freedom in the second sense, as full-blown self-legislation. For this reason, autonomy is something to which we should always, for Kant, aspire.

Phenomena and noumena

It is becoming urgent that we start saying something more about this aspiration, and about why Kant thinks we have it. Before we do so, and before we leave this preliminary sketch behind, it is only fair to allow that profound puzzles and questions about Kantian freedom remain. They begin with this question: how can Kant maintain that will is radically free without disrupting the deterministic Newtonian universe to which he was also committed? Kant's famous distinction between noumena and phenomena is indispensable here, though many have wished they could dispense with it. For Kant, as we have mentioned in earlier chapters, the radically or transcendently free will is conceived as part of a 'noumenal' realm or order. A 'noumenal realm' is a law-governed order of concepts and objects that are not licensed or verified by evidence accessible to the senses or to spatio-temporal intuition generally.⁹ God and the immortal soul, along with the free will, are among the objects in the noumenal realm. This realm or order and its objects are known by the intellect, mind, or νοῦς (*nous*), acting independent of sense. Kant accordingly calls noumenal concepts and

⁸ I also choose heteronomously, for Kant, if I choose to submit my will to authority, whether human or divine. Considered cooperation can be autonomous – if I have decided, for reasons of my own, that some authority is correct in its demands, action in accord with those demands can be autonomous. But autonomy is forfeited when I choose to abdicate my role as ultimate jury and judge, when I say 'your will, not mine,' in whatever context (see generally WiA).

⁹ 'Intuition' for Kant is (the faculty of) immediate apprehension; to know or grasp 'intuitively' is to know or grasp as immediately evident, without need of discursive or step-wise demonstration. Kant famously holds that human beings have only 'sensible' intuition, that is, intuition of things as they can be pictured, felt, or otherwise tracked in a spatio-temporal theater. We lack 'intellectual intuition,' which immediately grasps the inner essences of and relations between things. (God, presumably, has intellectual intuition.)

objects 'intelligible,' distinguishing them from everything that is 'sensible,' or accessible to the senses. Things accessible to the senses Kant calls 'phenomena,' or 'things that appear,' from the Greek root verb φαίνω (*pheno*), 'to reveal' or 'to make appear.' Now phenomena – ordinary spatio-temporal physical objects, paradigmatically – behave according to the Newtonian laws of nature, the laws of mechanistic causation. Phenomena are the basic subject matter of natural science, the data with which science works and which it explains. But noumena are not like this, and noumena do not behave according to Newtonian laws – noumena have their own laws (sometimes called, by Kant, the 'laws of freedom') (e.g., G 4:387; or KpV 5:65; or MS 6:214). Kant solves the apparent conflict between freedom and determinism by arguing that will, insofar as it is free, belongs to an order of concepts and objects and laws other than the natural.

Many have worried that this makes freedom of the will a kind of illusion, an idea we want to posit, since it enables attributions of responsibility, but which *really* does not describe things as they are. But it is important to keep in mind that, for Kant, the phenomenal order of appearances, governed by rigidly deterministic Newtonian law, is no more or less real than the noumenal order of intelligible objects, governed by 'laws of freedom,' and accessible only to the intellect (see A537/B565). Each system has its own laws of evidence and its own set of warranted claims, and in some particularly interesting cases – the case of the will, for example – an object can belong to both systems. The ways each system determines or characterizes the object might conflict if they were meant to belong to the same explanatory and justificatory schema, but they are not (A536/B564–A541/B569). Many have been and are committed to the exclusive 'really realness' of the objects described by natural science, but Kant himself was unwilling to privilege the natural scientific and phenomenal over the moral and noumenal, leaving transcendental freedom at least as real as universal deterministic causation.¹⁰

RATIONAL FREEDOM

It is interesting and eminently worthwhile to think further about Kant's solution to the dilemma posed above, but doing so is beyond the scope both of the current chapter, and of the investigations to which this book is primarily committed.¹¹ Our purpose right here – coming to a basic

¹⁰ I would indeed argue that Kant privileges the moral over the scientific. See A797/B825–A800/B828.

¹¹ Readers interested in pursuing them can find a wealth of literature elsewhere. See [Chapter 1](#), [note 11](#), for some places to start.

understanding of will as free, for Kant – is served adequately without solving these problems. The most important thing, for our purpose, is to understand the close relationships Kant posits between rationality and freedom, that is, between freedom of the will and the possibility of pure practical rationality. As Kant claims, if pure reason is to be practical, then we must have free wills, that is, wills capable of being determined on other-than-natural grounds, and conversely, if we have free wills, then pure reason must be able to be practical, that is, able to furnish grounds for action that provide a genuine alternative to natural grounds and that make deep self-determination possible (KpV 5:29). For Kant, the rationality and the freedom of our wills mutually imply each other.¹²

We close this description of freedom and the Kantian will by returning to the thought that freedom of the will makes morality possible for Kant. Freedom is of course crucial to the possibility of assigning responsibility. But this is not the whole story. For Kant, freedom from determination by nature makes action on rational grounds itself possible, and the possibility of acting on rational grounds is the possibility of aspiring to a way of life and to ideals of action that transcend nature. This is a radical possibility indeed. It says we can do things differently, that we can change courses, that we are not required to repeat, season after season, generation after generation, the same moves. And it says that we can change in the service of a way or life or ideals we think are *better*. Which way of life is that? What ideals are those? What I assert here (and defend in the following chapters) is that Kantian morality not only presupposes the free rational activity of the will, but also aims at it – that it presupposes freedom and rationality in the first, bare, senses described here, and then aims at them in their second, fuller, senses. The way of life Kantian morality demands of us is one that honors the promise of free rational activity; the ideal reason presents us with is an ideal of free rational activity; the ultimate ground of the moral law is the good of the ongoing free rational activity of a human will.

This idea, as promised, will be developed further in the next chapters. I end here my attempt to sketch a picture of the Kantian will, with all its faculties and capacities and reasons and ends and laws and freedoms. My sketch has emphasized the fact that Kantian will needs reasons for action and that action is always undertaken for the sake of something the subject understands to be good. I have sought to make clear that reason's

¹² Kant concludes his KpV 5:29 discussion with this: "Thus freedom and unconditional practical law reciprocally imply each other." Henry Allison's explication of this thought, which he calls Kant's 'reciprocity thesis,' is very useful (*Kant's Theory of Freedom*, 214–18).

own capacity to *furnish* reasons or grounds of action is also an essential component of Kantian freedom, and to set us up to start asking what reason's own grounds might be. How will pure reason come up with reasons for acting? How will these reasons motivate? What will be the content of the principle of action reason urges? The centerpiece of Kantian moral philosophy has often been taken to be the 'purely formal' supreme moral principle (the categorical imperative), understood as a principle that has no aim of its own. I hope my presentation of the Kantian will makes clear how untenable such a view is. It just does not make sense to imagine that Kant thought the categorical imperative would apply to us and command our allegiance without expressing any ideal or aim whatsoever, without, that is, containing a significant and compelling ground. I hope here to have shown this by having shown, at least in outline, how the Kantian will – the real centerpiece of Kant's view – does work.

Against nature: Kant's argumentative strategy

THE PROBLEM

The problem is a problem of long standing. In the Presidential Address to the December 2001 Eastern Division American Philosophical Association Meeting in Atlanta, Georgia, for instance, Virginia Held raised doubts about research programs in ethics that insist on *naturalism*.¹ Held targeted recent efforts in sociobiology and evolutionary psychology to ground ethics in natural facts about human beings, and argued against philosophical projects – like those of Daniel Dennett, Paul Churchland, and Allan Gibbard – that seek to give ethics psychobiological foundations. She cautioned fellow feminist thinkers against embracing naturalism, congenial as embracing nature might seem, especially to thinkers – Held names Annette Baier – eager to revalue the ‘natural’ moral practices of care, and to generally reclaim the denigrated natural sphere – home to bodies, emotions, and the mundane, messy, as well as ‘sinful’ facts of reproduction – to which women have often been consigned. Held’s argument, roughly, was that nature, by its very nature, cannot be turned to for answers about what is morally good, what evil, what called for, what forbidden.² We rebel, as we should, against systems of gender hierarchy, no matter how rooted in ‘nature.’ We reject, as we should, callous selfishness, again no matter how natural. The normative, Held argued, cannot come from the natural; we should not try to ground moral oughts, her thought goes, in what ‘by nature’ is.³

In making this sturdy and in many ways Kantian set of moves, Held was moving against a tide. A commitment to naturalism – that is, to

¹ Virginia Held, “Moral Subjects: The Natural and the Normative,” *Proceedings and Addresses of the American Philosophical Association* 76:2 (Nov. 2002), 7–24.

² Held, “Moral Subjects.”

³ See Kant, A547/B575: “indeed the **ought**, if one has merely the course of nature before one’s eyes, has no significance whatever. We cannot ask at all what ought to happen in nature, any more than we can ask what properties a circle ought to have; but we must rather ask what happens in nature, or what properties the circle has.”

explanations that emphasize impersonal causality and that eschew reliance on super- or extra-natural forces, objects, or beings – has become *de rigueur* for many twentieth- and twenty-first-century philosophers and intellectuals. Naturalism in this form emerged from early modern scientific materialism, with its commitment to seeing the natural world as constituted by essentially inert matter ‘animated’ by impersonal physical laws.⁴ The spectacular eighteenth- and nineteenth-century successes of technologies based in mechanistic sciences, as well as theoretical advances in the sciences themselves, inspired the rise of ‘scientific’ approaches to social, economic, political, cultural, and psychological phenomena. Sociology, economics, political science, anthropology, and psychology emerged, and sought laws analogous to the laws of physics to describe and predict the behavior of their objects. Even science got more scientific: Darwinian theory, via mechanisms of adaptive evolution, promised to explain away what had looked like the intelligent design of living organisms and ecosystems, that is, of the objects of biology. The world we live in, the world our theories were expected to explain, became, as Max Weber put it, ‘disenchanted,’ a place not of willful spirits, intelligent forces, and signs and meanings, but instead of machines and of mechanistic systems.⁵ By the early twentieth century, the most respectable objects of intellectual inquiry were objects that natural (or naturalistic) science could track, and the most legitimate questions about them were questions that could be addressed in terms of just those concepts and laws that natural (or naturalistic) scientists warranted.

Kant himself may have contributed to the rise of naturalism, though not intentionally. Kant’s famous – or infamous – solution to the problem of free will in a deterministic Newtonian universe, a solution that relies on an ‘extravagant’ two-world metaphysics, makes just the kind of moves naturalists love to hate. Kant, as we saw in [Chapter 4](#), segregates free will (and any kindred ‘uncaused causes’) from deterministic nature, placing them in a ‘noumenal’ realm, apart from the ‘phenomenal’ realm of nature, thereby insulating them from the reach of deterministic law. But such an apparently dualistic metaphysics is anathema to naturalism: Kant’s metaphysical

⁴ See, for example, Peter Dear, *Revolutionizing the Sciences: European Knowledge and Its Ambitions, 1500–1700* (Princeton: Princeton University Press, 2001), 86–90. For a terrific discussion of the key seventeenth-century debate over ‘living matter’ – a debate won by those who argued for matter’s ‘deadness’ – see Martin Schönfeld, *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford University Press, 2000), chs. 1 and 2.

⁵ See Max Weber, “Science as Vocation” [1918–19], in *From Max Weber: Essays in Sociology*, trans. H. H. Girth and C. Wright Mills (Oxford: Oxford University Press, 1946), 155. Nietzsche’s famous claim that ‘God is dead’ makes a similar point (at, e.g., Friedrich Nietzsche, *The Gay Science* [1887], trans. Walter Kaufman [New York: Random House, 1974], §125, p. 181).

response to natural science could easily serve naturalists as a *reductio ad absurdum*, an object lesson in the perversities to which one commits oneself if one tries to reserve room for the non-natural. My aim in this chapter, however, is to revive Kant's insistence that nature is not all there is. It is important to do this, if not for its own sake, at least for the sake of understanding Kant aright.

Kant rejected the idea that the proper aim of morality could ever be something in or of nature, and he rejected it with a vehemence perhaps not seen before or since. The natural world, for Kant, is deeply inadequate to anything we might recognize as a moral system. It cannot furnish a properly moral end or ground of action, and it is therefore inadequate to produce moral imperatives, principles, or even feelings. It does not exemplify moral concepts, at least not unambiguously, and so cannot be the source of such concepts. The infinitely interconnected and mutually determining causal nexus that is nature makes meaningful assignments of responsibility impossible. In an important sense, nature is not even the site in which moral value is realized, for Kant. And yet, for Kant, the moral as we know it is most certainly a fact – so nature cannot be all there is. These are strong claims, and, as suggested above, cut against the grain of much contemporary thought generally as well as of many sympathetic reconstructions of Kant's moral philosophy. In my view, however, Kant's reasons for rejecting what we would call 'naturalism' are absolutely central to his understanding of morality. They inform the content of the moral law at which he eventually arrives, and which Kant regards as the supreme achievement of his moral theory. Understanding Kant's anti-naturalism is therefore, I think, key to understanding the core commitments of his moral theory.

Kant's rejection of a natural basis for morality is at its heart a deep rejection of the idea that the essential facts about humanity, about our lives and options and practices and possibilities, are given to us by senseless, impersonal forces. For Kant, freedom and rationality, not determinism or brute fact, must lie at the heart of our lives. To reject the possibility of a morality grounded in nature is to reject the thought that the ultimate aims, and the ultimate structure, of human practice – and, therefore, of much of human life – are or ought to be given to us by anything other than ourselves, that is, by human beings, individually and collectively, acting both freely and rationally. It is to reject, in other words, the idea that our lives ought to be organized by things that are given. Kant also, not accidentally, rejects the idea that God furnishes our lives with their ultimate aim and structure: if they are to have moral worth, our lives must be exercises in our own *active*

making, not in passive obedience or acceptance of external givens. For human practice or action to have moral value, that is, for it to have the highest value it can have, its structure and aim must be arrived at rationally and freely by human beings (see WiA). Where its structure and aims are not arrived at in this way, our practical lives are characterized by acquiescence to external coercion or compulsion, by passivity in the face of givens, by brutish instinct, and by deep senselessness. This, I will at any rate argue, is Kant's view.

This chapter proceeds by looking at Kantian arguments based mainly in what Kant takes to be ordinary intuitions, commitments, and experiences of morality. These, as we shall see, are key to Kant's overall argumentative strategy. All of them tell, for Kant, against natural grounding. Ordinary intuitions, commitments, and experiences of morality reinforce the view that freedom and rationality are prerequisites of morality: without them, we cannot attribute responsibility or legible intentions to people. Ordinary intuitions, commitments, and experiences of morality also all recognize the special superiority of freedom and reason. This point is crucial. Nature is not only *inadequate to the possibility* of morality, that is, it not only fails to contain the freedom and rationality that are morality's prerequisites. Nature contains nothing in itself *worthy of moral choice*. We saw in earlier chapters that we are, in one set of senses, free and rational whether we want to be or not, but that there are further senses of freedom and rationality to which we can aspire, 'full flowerings' of freedom and rationality that we can strive to attain, or not. These are worthy of moral choice. The point here is that, for Kant, our ordinary thinking about morality demands the non-natural *both* for morality's possibility, and for morality's end. In arguing against a natural ground for morality, Kant is not just pointing out a sad but true inadequacy, but is mounting an attack against settling for the given, for passivity, obedience, and senselessness, given that we can, and in fact already expect ourselves to, do better.

In making the depths of Kant's anti-naturalism clear, the chapter thus also seeks to advance the more general argumentative arc of the book. For one, it should make clear that Kant's rejection of natural (or empirical) goods as a suitable grounds for morality is not a rejection of the idea that there is *some* good which grounds morality – though Kant sometimes asks to be read this way. In addition, by rejecting the natural, it should begin to lay bare the contours of Kant's commitment to the intrinsic value of 'unnatural' free rational agency, or the activity of free rational will. In doing so, it sets the stage for [Chapter 6](#)'s discussion of the moral law's grounding in free rational willing itself.

KANT'S UNDERSTANDING OF NATURE

The first order of business is to get clear about Kant's understanding of nature. "Now, nature in the most general sense," Kant writes, "is the existence of things under laws" (KpV 5:43). In the full passage here, Kant mentions both 'sensible' and 'supersensible' nature (KpV 5:43), or the systems of both phenomena and noumena, which fall under 'laws of nature' and 'laws of freedom' respectively. In passages in the *Groundwork* (e.g., G 4:395), Kant also writes about 'a nature' that is a system of things falling under *teleological* laws, and teleologically organized nature is, of course, a central subject of Kant's later *Critique of the Power of Judgment* (see KU 5:167–98 and 357ff.). But in most contexts, and in all those relevant for our purposes, Kant understands by 'nature' just what he above calls 'sensible' nature: the world of spatio-temporal material or physical things available to our senses and described by mechanistic Newtonian laws. The 'naturalism' Kant opposes is a naturalism that understands nature thus.

Kant often refers to this nature – to the world of spatio-temporal material or physical things available to our senses and described by mechanistic Newtonian laws – as the 'empirical.' The term 'empirical' itself derives from the ancient Greek 'Empirics,' a group of physicians who believed that experience (*empeirikos*), as opposed to theory, was the surest route to sound knowledge. Kant, like many of his contemporaries, uses the term 'empirical' to pick out those things that can be known only through experience, that is, for Kant, that can be known only through investigation that relies on evidence from the senses. Things known this way may also be said, in the terminology of the day, to be known a posteriori. Hence, for Kant, the 'sensible,' the 'empirical,' 'objects of experience,' 'things known a posteriori,' and the 'natural' are often equivalent.

Now there *are* things about nature that we know a priori – things we know, that is, *independent* of experience, on the basis of evidence that does not ultimately rest in sensory experience. For Kant, we know a priori that objects of sensory experience will have temporal duration and (if they are external objects) spatial extension,⁶ though we do not know a priori what particular objects there will be, or what durations or extensions they will have. We know a priori that objects of experience will stand in causal relations with each other, though we don't know a priori what those relations will be. Kantian nature, in other words, is a joint product of a

⁶ Kant allows that some 'inner objects,' such as the train of my own thoughts, are not spatially extended (A22–3/B37).

priori 'forms' and a posteriori 'matter.' Things – such as facts, objects, concepts, and regularities – that we know empirically – that is, a posteriori, experientially, sensibly – are the things that, in all their particularity, make up the 'matter' of a system of nature the most basic structure of which is given by forms (of space and time; see A19/B33–A49/B73) and concepts or principles (the 'pure' categories of the understanding; see A66/B91–A83/B113) that we know a priori.⁷ Much more might be said about all this, which we will not say here. It is enough, for our purposes, to get the basic idea. One last note: despite its a priori structure, Kant often uses 'the empirical' to refer to nature as a whole. This makes sense because *as a whole* nature is accessible to (if not exhaustively known through) sensory experience.

Now as human beings, *we* are natural insofar as we are spatio-temporal physical things. Our bodies are in nature, and so are our sense organs and our brains and our nervous systems. Desires that arise from the natural constitutions of our bodies are also in and of nature, for Kant. Hungers, sexual drives, all desires aimed at things because they please the senses or contribute to the overall feeling of well-being of the organism are, for Kant, natural. Kant's rejection of an empirical or natural basis for morality is a rejection of the idea that morality could in any way emerge from (could depend for its central concepts or motivations on), or could aim ultimately at, anything in spatio-temporal Newtonian nature, including our natural desires or natural aims or our natural human constitutions. We turn now to look more carefully at the bases of this rejection.

KANT'S COMMON-SENSE CASE AGAINST A NATURAL FOUNDATION FOR MORALITY

The moral commitments of 'common rational cognition'

As noted above, the route I take into Kant's anti-naturalism is one cut by Kant himself. In his 1785 *Groundwork of the Metaphysics of Morals*, Kant begins with what he calls the ordinary understanding's 'common rational cognitions' (G 4:392–3), or what we might call the demands of common sense.⁸ Kant accepts these demands as sound, and indeed stakes much of his

⁷ We come to know these a priori in coming to see that they are conditions of the possibility of the experience we in fact have. Helping us see this is Kant's task in the "Deduction of the Pure Concepts of the Understanding" in his *Critique of Pure Reason* (A84/B116–A130/B169).

⁸ For an excellent discussion of the role of 'common rational cognition' in Kant's *Groundwork* strategy, see Paul Guyer, "The Strategy of Kant's *Groundwork*," in his *Kant on Freedom, Law, and Happiness* (Cambridge and New York: Cambridge University Press, 2000), 207–31.

case on fit with them. Common rational cognition, Kant thinks, already has key foundational moral facts right; the main task of the *Groundwork* is to properly discern, explain, and theoretically ground those facts, not to revise them (see, e.g., G 4:397). In his 1788 *Critique of Practical Reason*, his second major work on moral theory, Kant insists that he offers no new moral principle, but simply a perspicuous formulation of the principle that ordinary moral people have respected all along (KpV 5:8n). Kant wants his moral theory, in general, to be understood as an attempt to work out the philosophical details that will allow pre-existing commitments about the nature and demands of morality to be integrated into a systematic account of the rational, human subject and its place in the world: he wants, in other words, to clean up morality's philosophical house. As an argumentative strategy, Kant is inviting readers to travel with him along a path he argues is demanded by their very own starting points. So, what starting points, what commitments and demands of 'common rational cognition', does Kant find, and how do they militate against natural grounding?

The unconditional goodness of the good will

Kant opens the [first chapter](#) of the *Groundwork* with this claim: "It is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation, except a *good will*" (G 4:393; Kant's emphasis). Kant explains: it always, on reflection, turns out that the other things we might have been tempted to regard as "good without limitation," or unconditionally good, are dependent for their goodness on the goodness of the relevant will. Kant gives examples: think, he says, of "talents of mind" such as "understanding, wit, judgment"; "qualities of temperament" such as "courage, resolution, and perseverance in one's plans"; "gifts of fortune" such as "power, riches, honor, even health"; and "moderation in affects and passions, self-control, and calm reflection" (G 4:393–4).⁹ We all, Kant thinks, agree that in the hands, as it were, of a bad will, none of these is good: talents of mind, qualities of temperament, gifts of fortune, and even self-control can all be used for evil ends, and so are good only conditionally. Even happiness – or "complete well-being and satisfaction with one's condition" – cannot be regarded, Kant tells us, as unconditionally good (G 4:393). Kant appeals to our disapproval of the "uninterrupted prosperity of a being graced with no feature of a pure and good will" (G 4:393), that is, to the fact that it irks us

⁹ It is noteworthy that Kant parenthetically identifies the last group as virtues wrongly accorded unconditional value by "the ancients" (G 4:394).

when bad people seem happy. The idea is that even happiness, if it is to win our approval, must belong to one who has a good will – which shows that we do not regard happiness itself as unconditionally good. While good will is not, Kant acknowledges, “the sole and complete good” – other things are good, if conditionally so, and the *complete* good would include a divine distribution of happiness proportionate to moral goodness – the good will is still the only thing we can think of that, on reflection, can qualify as “the highest good and the condition of every other” (G 4:396). Thus the opening: “It is impossible,” to repeat, “to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a *good will*” (G 4:393; Kant’s emphasis).¹⁰

Having secured agreement that good will is the only unconditional good, Kant’s next move in the *Groundwork* is to unpack our understanding of good will itself. Kant starts off ‘negatively,’ with a claim about what good will is *not*. By ‘good will,’ we (who share the common-sense understandings Kant identifies) do *not* understand a will that produces certain kinds of effects. A will, that is, is good in virtue of something other than what it manages to produce. Kant is making appeal here to the intuition that particular effects – flowers taken to a sick room, help with a stroller on a staircase – good as these may be, are never *good in themselves*, but derive their goodness from the underlying *intention* of the agent. Taking flowers to a sick friend loses its goodness if I do it only because I owe her money and want to stall repayment demands; helping someone with a stroller up the stairs loses goodness if I do it because I want a better look at the peculiar child it holds. Effects, for Kant, are good just insofar as the intention that produced them was good. But then, a will is not good *in virtue of* its effects – the case is rather vice versa.

To press the point, Kant draws us into agreeing that a good will loses none of its goodness even if “with its greatest efforts it should yet achieve nothing” (G 4:394). The intuition, again, is that goodness is measured not by effects but by quality of intention. Indeed, even where well-intentioned actions have disastrous, unforeseeable effects, the goodness of the will that

¹⁰ It is worth noting what a bold claim it is that nothing in *or even beyond* the world can be thought of as unconditionally good except a good will. Kant’s contemporaries would have thought immediately of God as a prime candidate for something unconditionally good. Kant was not ignorant of this – his “indeed even beyond [the world]” signals his awareness of God’s candidacy for unconditional goodness clearly enough (G 4:393). In the *Groundwork*’s second chapter, Kant writes that even the “concept of God as the highest good” comes “solely from the *idea* of moral perfection that reason frames a priori and connects inseparably with the concept of a free will” (G 4:409; see also G 4:443): Kant thinks that even God’s goodness can be conceived only in relation to the more fundamental conception of the goodness of a good will.

willed them is not impugned. I brought flowers for the right reasons, but they aggravated another patient's asthma, so a nurse threw them away. Finding them in the trash, a disgruntled floral worker rehatched an arson plan. Some bad things happened. But my will, Kant insists, was still good, as was the act of bringing flowers as I did. The goodness of any action derives not from its consequences, but from the intention that animated it. This, Kant thinks, all agrees with common sense. Notice, however, that it already delivers a significant blow to any empirical foundation for morality. Effects are spatio-temporal events in the empirical world. If they can never be good in themselves, it looks like the empirical as such can never host, so to speak, unconditional value. If we want, and Kant thinks we do want, an unconditional good at the heart of our morality, we should understand ourselves already here, at the very start of the *Groundwork*, to be on notice that nature may come up short.

Perhaps aware that, in rejecting effects as bearers of goodness, he is asking us to grant a lot, and fast, Kant slows things down. He remarks that, "despite all the agreement even of common understanding," the idea that a good will, the only unconditionally good thing, is good independent of its effects is "so strange" [*etwas so Befremdliches*] that, having stated it explicitly, we might begin to suspect it of being, "a mere high-flown fantasy" (G 4:394). And the idea, though it takes advantage of familiar intuitions, is indeed strange. It says that the thing we value most highly, the thing to which we accord intrinsic and ultimate value, is some quality of an agent's will – and at the same time it refuses to make that quality out in terms of its production of the everyday things (effects, states of affairs) we normally think of as good.

The analysis of duty

Kant turns to assure us that the strange idea is not misguided, not a confusion or fantasy, by moving to "explicate the concept of a will that is to be esteemed in itself and that is good apart from any further purpose" (G 4:397). He does so by examining the concept of *duty*, which, he writes, "contains that of a good will" (G 4:397),¹¹ and describes action that is good for its own sake. The idea here is to take something that captures our common-sense sense of what makes a good will good, or that characterizes

¹¹ "Under," Kant adds, "certain subjective limitations and hindrances" (G 4:397), namely the limitations and hindrances to which all human wills are subject. The contrast here is with purely good wills (such as those belonging to angels), which can do no wrong, and which the concept of *duty* (or moral obligation) therefore does not (need to) 'contain' (see G 4:412–13).

good willing, and further explicate *it*. We are (at least tentatively) persuaded that the goodness of a will does *not* depend on the effects it produces, and we are willing to say that the goodness of a good will consists in its having the right kind of intention. This amounts, Kant thinks, to saying a good will is one that wills from *duty* (even if we are not yet sure what exactly *that* consists in). So we turn to the concept of willing from duty itself, and see what we can discover about it.

Kant makes three claims (or ‘propositions’) about willing from duty. His first is that to will from duty is to will some action or effect *because* it is what duty requires, that is to say, because we take willing the action or effect to be morally good or right (G 4:397–9). This kind of willing stands in contrast to pursuing some action or effect because it will satisfy a sensuous inclination or desire, that is, because we anticipate that it will satisfy (some part of our) self-interest. Now, we do not yet know exactly what it means to will something because it is morally good or right because we do not yet, given the order of Kant’s exposition, know what it *is* for something to be morally good or right. At this point, Kant is just appealing to what he takes to be the common-sense distinction between doing something because one ought to (because it is duty, because it is morally good or right) and doing something because it answers some inclination (because it promises to satisfy some sensuous desire, because it serves short- or long-term self-interest). He takes it that we all feel this distinction, even if we can’t (yet) give a full account of it. But we still get a result: by calling on our intuitions that dutiful willing is special, Kant secures the thought that there is a gap between (morally) good willing and inclination-driven willing.

Kant’s second proposition grows out of the distinction drawn above between action undertaken out of duty and action undertaken to satisfy sensuous desire. Kant writes:

The second proposition is this: an action from duty has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon, and therefore does not depend upon the realization of the object of the action but merely upon the *principle of volition* in accordance with which the action is done without regard for any object of the faculty of desire. (G 4:399–400)

We saw a version of this point above, when we saw Kant insisting that the goodness of a good will does not depend on its effects. The new thought expressed here is this: whether my will is dutiful or not, hence good or not, is a matter of the *kind of reason* I allow to guide my will, the *ground* of action I accept. Maxims, or subjective principles of volition, always advert to, or ‘contain,’ reasons or grounds for acting – so maxims, not outcomes, are

what matter.¹² The second proposition thus makes more explicit our intuition that moral worth lies in the *reasons* someone acted, and, hence, in the kind of guiding principle that is operative for the will.

Now, as it will happen, the difference between good or dutiful willing and other sorts of willing is also, for Kant, the difference between willing from an interest in duty itself (about which, it is true, we so far know little) and willing from inclination for some empirical object. Before finishing with Kant's 'second proposition,' it will be useful to formally introduce a distinction between *grounds* and *empirical ends or objects* of human action. Kant makes such a distinction in several places (KpV 5:21, 27–9, 34; see also G 4:400), though his use of the terms is not always consistent with the distinction. It is nonetheless a helpful distinction. Empirical ends or objects of action are those *concrete, empirical, spatio-temporal things* or states of affairs at which action aims. (Kant also calls such objects or ends 'the material' or 'matter' of will [e.g., KpV 5:21, 34].) Grounds, in contrast, are the *reasons* actions are undertaken. Now, any act of willing requires an empirical object, that is, a spatio-temporal thing or state of affairs that the will aims to bring about. Any action, including dutiful action, must – in order to be action and not mere ratiocination – seek to realize *some* empirical end (KpV 5:34; see also REL 6:4). "No free action," Kant writes, "is possible unless the agent also intends an end (which is the matter of choice)" (MS 6:389). Actions must of course *always also*, as we have already seen, be undertaken for some reason, or on some *ground* or other: any act of willing requires not only an empirical end or object but also a ground. There must be a reason for wanting to realize this or that object, even if it is the simple reason that I have a desire for the object.

Imagine I fill my car's gas tank with gas. The empirical object of my action – the concrete state of affairs I seek to realize – may be simply a full gas tank, or perhaps, more generally, 'gas for the week,' or 'enough gas to get myself through the Everglades without stopping.' The *ground* of my action may be the same as the object – the reason I am putting gas in the tank is that I want there to be gas in the tank, or I want to be able to drive to work without needing to get gas again all week, or I want to get through the Everglades without stopping. But perhaps I put gas in the tank because I am lending my car to a friend, and last time I borrowed his car I used up all his gas. I believe it to be my moral duty to repay him by offering my car with a full tank. In this case, the ground of my action is not identical to whatever

¹² See Chapter 3 above for full discussion and defense of the idea that maxims always contain reasons for acting.

objects I seek to realize (here, a full tank), but is the fact that I believe I have a moral duty that I can fulfill by realizing those objects.¹³ Distinguishing empirical objects from grounds, as Kant saw (KpV 5:34), allows us to make sense of the possibility of action undertaken for a reason (on a *ground*) *other than* that of realizing some desired empirical thing or state of affairs. This possibility is tantamount, for Kant, to the possibility of dutiful willing itself. This is so because, as we have just suggested, dutiful or morally good willing is distinctive in being grounded in something other than desire for some empirical object.

We return to our story, the distinction between empirical object and ground in place. Kant concludes discussion of his second proposition with this dense passage:

In what, then, can [moral] worth lie, if it is not to be in the will in relation to the hoped for effect of the action? It can lie nowhere else *than in the principle of the will* without regard for the ends that can be brought about by such an action. For, the will stands between its a priori principle, which is formal, and its a posteriori incentive, which is material, as at a crossroads; and since it must still be determined by something, it must be determined by the formal principle of volition as such when an action is done from duty, where every material principle has been withdrawn from it. (G 4:400)

The principle of duty, however it is grounded, is not, Kant is insisting, grounded in inclination for empirical ends (“the hoped for effect of the action”). When action is done from duty, the principle that guides the will must, Kant tells us, be “a priori” and a principle from which “every material principle has been withdrawn” – it must, that is to say, be guided by “the formal principle of volition as such” (G 4:400). But what is a “formal principle of volition as such”? A formal principle of volition as such can only be a principle of will that is grounded in something apart from the sensuous desirability of some matter or material end. Kant’s second claim, then, not only reminds us that we assess goodness on the basis of an agent’s reasons, but combines this thought with the thought that action undertaken out of inclination does not qualify as morally good. Kant accomplishes a lot here: our intuitions about duty imply, according to Kant, that for there to

¹³ Taking advantage of Chapter 3’s discussion of maxims, principles, and imperatives, we can say the following. Whenever the maxim of an agent’s action expresses a hypothetical imperative, the *empirical end* of the agent’s will (the state of affairs the agent seeks to realize) is at the same time the *ground* of the action (the reason the agent undertakes the action); here, empirical object and ground are *one*. However, when the maxim of an agent’s action is informed by the categorical imperative, the empirical object aimed at in action is not identical with the ground (with the reason the agent undertakes the action) (KpV 5:34; also KpV 5:27, 28, 29, 57–8); here, empirical object and ground are *two*.

be unconditionally good willing there must be grounds of action distinct from empirical ends.

It is in fact worth pausing a bit longer here, before turning to Kant's 'third proposition,' to notice more fully what we have just seen Kant do – and to notice what Kant has not yet done. Kant opens the *Groundwork of the Metaphysics of Morals* with a set of appeals to intuitions or common-sense understandings about what makes a will morally good. These intuitions imply that unconditional goodness, moral goodness, resides, first, in something other than effects and, second, in something other than effect-motivated actions. For Kant, this means it must reside in intentions governed by an a priori, non-empirical, principle. We have here the strong beginnings of an argument that morality cannot be grounded in nature. But Kant's mode of argumentation here does not tell us what is immoral (or even simply amoral) about satisfying inclinations, and so about valuing empirical objects – rather, it simply appeals to our sense that duty calls for something else. Kant, at least here, relies on his readers' background antipathies toward inclination-driven action to do a good deal of work. This is fine, but can leave one wondering just what Kant, and Kant's readers, and maybe even oneself, have against action aimed at satisfying sensuous inclination. Kant's third proposition doesn't answer this question, but gives us a way into asking it. Let us turn to that proposition now.

Recall: we are unpacking our ordinary ideas about duty to understand the nature of the good will. Kant's third proposition is this: "duty is the necessity of an action from respect for law" (G 4:400). It says that acting out of duty consists in recognizing some action as necessary (as required, as demanded) because, and simply because, the principle (or law) demanding the action is itself *worthy of respect*. This is, as Kant says, "a consequence" of the two previous claims: to adopt a principle of action irrespective of empirical consequences can only be to act from respect for the principle, or law, itself. To act on a principle out of respect is precisely to appreciate and adopt it on grounds *other* than its helpfulness in satisfying inclinations. Respect [*Achtung*] is what we feel when we know we are in the presence of something that deserves our allegiance independent of what it can do for us. The good will – itself the only unconditionally good thing – is a will that acts to realize whatever the law of duty demands, out of nothing more than respect for that law.¹⁴

¹⁴ Does this start to mean that the moral principle is itself unconditionally good? In a sense, yes, and Kant writes about it as something that occasions respect, awe, etc. But in another important sense, no. The moral principle is the *standard* of goodness, and so does not compete with the good will for status as the only unqualifiedly good thing, that is, the only thing that lives up to the standard.

What occasions respect?

So, what law occasions respect, and why? This is the question Kant must answer now. It is all well and good to unpack the ordinary intuition that a good will is a will that does the morally right thing, and that does it simply because it is morally right. It is fine to gloss this by saying that good will acts out of duty, and that to act out of duty is to act out of respect for a principle aimed at something other than inclination-satisfaction. But none of this tells us anything directly about what the morally right thing *is*, or about what other than inclination can interest our will, or about what we are respecting when we respect the demands of duty. What principle deserves our respect? In what is it grounded? And why, again, *can't* the answer be 'empirical ends'?

The first thing to do here is understand what Kant means by 'respect.' The German is *Achtung*, which, besides 'respect,' can mean 'attention,' 'esteem,' and 'regard.' None of these English terms quite captures the distinctive character or power of the experience Kant is trying to name. At one point in the *Critique of Practical Reason*, Kant glosses respect as "boundless esteem" [*grenzenlosen Hochschätzung*] (KpV 5:79), and H. J. Paton, in an effort to capture Kant's meaning, translates *Achtung* as "reverence."¹⁵ 'Reverential respect' might be a good compromise – though I will stay with 'respect' to agree with more recent translations. In any event, the feeling of *Achtung*, or respect, is, for Kant, in a class apart from other feelings. Most feelings, Kant thinks, originate with sensuous appetites and aversions. Love and hate index things we want, and things we don't. To feel mad, sad, or glad is to feel frustrated, despairing, or satisfied in our pursuit of empirical objects. Feelings like longing, bliss, dismay, boredom, giddiness, and desperation, for example, all might be said to express senses of how, roughly speaking, the world is treating us, *qua* sensuous beings. Add complex and long-term self-interest (or 'self-love'), rooted in desire for overall sensuous well-being, and feelings like chagrin, pride, relief, and resolve can (at least unless and until they get involved with moral motives) be understood as having sensuous origins. Respect is fundamentally different (see generally KpV 5:71–81).

The key difference consists in the different relation these different feelings have to judgment. Feelings with sensuous origins, feelings that result from the interaction of our inclinations with the empirical world, *are the*

¹⁵ H. J. Paton, *Groundwork of the Metaphysic of Morals* (New York: Harper & Row, 1964), at, for example, 68 (G 4:400). Paton's English translation of the *Groundwork* (first published in 1948) was the gold standard for almost fifty years, until the 1997 publication of Mary Gregor's translation by Cambridge University Press.

basis for judgments about whether things are good (for us, in any relevant sense). A feeling of respect, in contrast, *follows on* a judgment that something is good. Other feelings *ground* judgments; respect is a feeling that is *produced by* a judgment. And the judgments themselves differ, in a way that is crucial. Judgments based on feelings are judgments that things are going well or badly for me, as far as inclination-satisfaction goes; they are, that is, judgments about *conditional* goodness. In contrast, judgments that give rise to feelings of respect are judgments that something is good *unconditionally*, irrespective of my particular inclinations. The difference is crucial because making the latter sort of judgment does not require me to 'check in' with my inclinations – the judgment itself is not dependent on feelings. But it does, and this is the key point here, *give rise to* a feeling. Kant writes that "respect for the moral law is a feeling that is produced by an intellectual ground" (KpV 5:73). Respect is a *consequence* of recognizing the unconditional good, represented in the moral law, not a foundation for it, or, as Kant writes, respect is "the *effect* of the law on the subject, and not ... the *cause* of the law" (G 4:401n).

It can be helpful, I think, to notice that Kant's distinction between respect and other feelings tracks an earlier philosophical distinction, one important especially in discussions of voluntarism.¹⁶ The distinction is between deeming a thing good because one loves it, or loving a thing because it is good (where 'love' is a catch-all that covers all kinds of desiring and approving). To respect something (in Kant's sense) would, in this earlier idiom, be to love something because it is good, that is, to love it as a consequence of a rational judgment that it is good in itself, unconditionally, apart from one's desires and inclinations. To have other sorts of positive feelings toward something – to be fascinated by it, or filled with lust for it, or swept off one's feet by it, or be mildly inclined toward it – would be to deem something good because one loves it. Respect is a feeling that follows judgment, rather than driving it.

On Kant's analysis, the actual feeling of respect is a complex physiological result (for Kant, all feelings are physiological) of a rational recognition. Kant's quite interesting idea is this: when I make a judgment about goodness that sets aside inclinations, inclinations lose standing, and they, my normal physiological guides, are thus painfully "subordinated" to (G 4:401n),

¹⁶ 'Voluntarism' is the position that God's will (power of volition) *decides* standards of goodness, rather than being responsive to them. Voluntarism thus sustains maximal freedom for God, since it frees him of constraint by external standards, but it does so at the cost of rendering God's decisions arbitrary, rather than rationally motivated. Debates about voluntarism were central to scholastic and early modern philosophy.

even ‘struck down’ and ‘humiliated’ by (KpV 5:77), something (rationally) recognized as higher and more sublime. At the same time, this painful feeling is *combined* with a feeling, consequent on the rational recognition, that I have access to and can participate in something that *is* more sublime than satisfying my contingent needs and desires (G 4:401n; KpV 5:73–7), which makes the subordination of inclinations not altogether unpleasant. To feel respect is thus, for Kant, to feel simultaneously humiliated and ennobled, to feel that one’s ‘higher’ self is being called on, even at the expense of one’s lower self.

It should, at this point, make sense that Kant describes morally good willing as willing out of respect. Willing out of respect is willing responsive to a judgment that something is good in itself, independent of inclination and of its empirical effects – and this is just what we demand of morally good willing. After introducing his third proposition (that, recall, “duty is the necessity of an action from respect for law” [G 4:400]), Kant’s moves can seem to be just more reiteration. We never, ‘common rational cognition’ affirms, feel *respect* for the pull of inclination, nor do we respect principles or demands grounded in inclination. We’ve essentially heard this already. But clues about what *might* ground a principle worthy of respect lurk in Kant’s formulations here:

Only what is connected with my will merely as ground and never as effect, what does not serve my inclination but outweighs it or at least excludes it altogether from calculations in making a choice – hence the mere law for itself – can be an object of respect and so a command. (G 4:400)¹⁷

Only something that interests me – is “connected with my will” – as a ground or reason for acting in itself, “merely as ground,” excluding inclination, can “be an object of respect,” Kant tells us. This can only be “mere law for itself” [*für sich*]. But what can this mean? Am I to respect law for law’s sake? This seems odd and unmotivated – and would require me to respect all sorts of laws, even laws of nature, which is clearly not what Kant is after. Clues to a more satisfying interpretation are found in nearby passages. Kant had just made the following point: “For an object as the effect of my proposed action I can indeed have *inclination* but *never respect*, just because it is merely an effect and not an activity of a will” (G 4:400; Kant’s emphases). Aha! Respect is occasioned only by something that is an “activity of a will.” Kant says virtually the same thing in another nearby passage. I can never have respect for empirical objects or states of affairs, Kant suggests:

¹⁷ This passage is, of course, an example of a place where Kant does *not* use ‘object’ to mean ‘empirical object’ or ‘matter of the will,’ but uses it rather to mean the ‘target’ of a kind of thought or attention (as in ‘the object of my affections’).

For, all these effects (agreeableness of one's condition, indeed even promotion of others' happiness) could have been also brought about by other causes, so that there would have been no need, for this, of the will of a rational being, in which, however, the highest and unconditional good alone can be found. (G 4:401)

Empirical effects can never occasion respect because they might have been brought about "by other causes," by something other than the activity of a rational will (G 4:401). Kant's thought, it seems to me, is this. The natural world, the world of empirical objects governed by mechanistic causal laws, chugs along, and though it may be awe inspiring as a whole (see KpV 5:161), nothing in it, no possible empirical object, is as such worthy of respect. Why not? Here, Kant's deep commitments come into play. Products of nature are what they are due to blind mechanism: they are local events in an infinite causal nexus driven by impersonal, iron law. As such, effects in nature are essentially senseless, coerced, and passive. What occasions respect, in contrast, is the reason-guided activity of a self-determining or free will. A thing does not impress me in the same way if it is "merely an effect and not an activity of a will" (G 4:400); and empirical effects can, one and all, be understood as having been, "brought about by other causes" (G 4:401), that is, as having, for their existence, "no need ... of the will of a rational being" (G 4:401).¹⁸ Any respect that empirical facts or states of affairs *do* happen to occasion they occasion just insofar as they intimate origins in the activity of a free rational will (see KpV 5:77).

Now, the hallmark activity of such a will is, of course, the authoring and following of moral law itself. "Our own will," Kant writes, "insofar as it would act only under the condition of a possible giving of universal law through its maxims – this will possible for us in idea – is the proper object of respect" (G 4:440). Or, as in *Religion within the Boundaries of Mere Reason*:

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

It is *this* law, the moral law we author and according to which we determine ourselves, that Kant thinks we respect – *and* it is ourselves, insofar as we write this law, insofar as its authority 'lies in us.' But this law and the will that authors it are not natural; nature gives no essential cues, natural processes play

¹⁸ This claim is, of course, an expression of Kant's commitment to the full mechanistic determinism of the phenomenal (or empirical) world. Any event in space and time must have an explanation in terms in mechanistic causation.

no role. The pieces start to fall into place: authoring and self-determining according to a purely rational law is, if anything is, what a good will, worthy of respect, does. But a purely rational law can only be a law grounded in the rationally recognized unconditional goodness of free rational will itself. Because the proper object of respect is thus ourselves, *qua* free rational wills and not *qua* natural beings, Kant describes respect as a feeling “*self-wrought* by means of a rational concept,” rather than one passively “*received* by means of influence” (G 4:401n; Kant’s emphases). Respect is a kind of awe at the unnatural best in ourselves. But then this is what we have been aiming to say: the unconditional good capable of grounding a moral law, the thing that occasions respect, looks like the activity of free rational will itself – which activity cannot, for Kant, be conceived as part of nature.

We are here: respect is only occasioned, according to Kant, by something non-empirical, something that is outside nature. This thing, we have also begun to see, is the activity of free rational willing itself. Chapter 6 will have much more to say about how the activity of free rational willing self-reflexively grounds the moral law Kant proposes. Here, thus far, we have followed a trajectory that began with Kant’s appeal to the good will as the only unconditionally good thing. We saw that, for Kant and, Kant thinks, for all of us, the goodness of the good will must be made out in a way that eschews foundational reference to the value of empirical states of affairs. We have arrived at the conclusion that the goodness of the good will must, therefore, consist in some non-empirical state of the will itself. Analysis of Kant’s take on respect shows that this state is the state of authoring and following the moral law, that is, the state of being free and rational.

We might feel done. We have seen that, for Kant, ordinary commitments about what deserves our moral allegiance militate strongly against anything – any end, any principle, any motive – that is grounded in nature. The intuitions on which Kant draws favor ends, principles, and motives that are fundamentally independent of our sensuous constitutions and sensuous well-being, insisting that we go beyond concern for these in our actions. We might stop here; the Kantian case against natural grounding might seem made. But it is made only on very Kantian terms, despite its starting points in common sense. In order to try to strengthen the case, we trace another route Kant takes to the same conclusion. We look this time at the data of ordinary moral experience.

The data of ordinary moral experience

By ‘ordinary moral experience,’ I mean the whole range of familiar thoughts, perceptions, judgments, decisions, actions, and shared practices

in which moral categories are implicated. Nature, as Kant and most 'naturalists' understand it, cannot (Kant thinks) be the source, or even properly speaking the site, of this experience.

But isn't this strategy already in trouble? Isn't nature precisely the realm of 'experience'? Won't the data of 'ordinary moral experience' therefore, whatever they are, *belong to* (rather than somehow rebuke) nature? Kant after all insists that we do not 'experience' the free will or its morally charged rational activity at all (see, e.g., G 4:448ff.; KpV 5:95ff.; Bxxviii). And if we don't experience free will or its morally charged rational activity, what is the status in experience of moral thoughts, perceptions, judgments, decisions, actions, and shared practices?

We will not really answer the question here (though it is an interesting one), because in fact what Kant means by 'experience' (*Erfahrung*) and what we want here to mean by 'experience' are two different things. When Kant denies that we experience free will, he is denying that the free will can, as such, *appear to our senses*: Kant denies that we have empirical access to – that we see or smell or hear or otherwise feel – the free will and its activity. Kant also denies that we can fit the free will and its activity into our theories of the physical world, that is, that we can 'experience' free will as part of the spatio-temporal Newtonian world of objects available to our senses. But then, as we saw in [Chapter 1](#), 'experience,' *Erfahrung*, for Kant, is narrower than the ordinary English term 'experience.' The ordinary English term refers to *any* sort of awareness, regardless of whether its objects can be verified by the senses or fit into our physical theories. In ordinary English, we 'experience' the contents of dreams and, if we have them, of delusions, religious revelations, hallucinations, and other things that precisely *do not* fit with what we know about empirical reality;¹⁹ we also routinely 'experience' things like intellectual 'aha!' moments or the constraints of logic, which experiences would be hard to assimilate to a system of spatio-temporal Newtonian objects. We also, of course, 'experience,' in the ordinary English sense, all sorts of things that *are* empirical: the English term is broader than Kant's technical German one. There is another German term, *Erlebnis*, that can also be translated 'experience.' Its meaning is more capacious and closer to the English 'experience.' I ask readers here to understand 'moral experience' not as the (for Kant) potentially oxymoronic *moralische Erfahrung* but instead as something like *moralisches Erlebnis*, or moral awareness, which, as we will see, Kant certainly thinks we have. So

¹⁹ Kant indeed denies that any of these are part of experience, properly speaking (see his "Refutation of Idealism," B274–9).

much for terminological clarification. What follows here is a look at some of the moral experience, in the ordinary English sense of 'experience,' to which Kant points in order to secure his rejection of a natural grounding for morality.²⁰

The inadequacy of the empirical world to moral concepts

A first piece of Kantian evidence that tells against any natural grounding for morality is that no empirical state of affairs seems adequate to furnish familiar moral concepts. The *Groundwork's* second chapter opens with a discussion of the shortcomings of all attempts to ground moral theories on concepts drawn from the empirical world. Kant works here with a concept we saw analyzed in the *Groundwork's* first chapter, namely duty. The claim is that we cannot nor should we expect to be able to glean the concept of duty from anything we encounter empirically.

Why not? "It is absolutely impossible," Kant argues, "for experience to establish with complete certainty a single case in which the maxim of an action in other respects right has rested solely on moral grounds and on the thought of one's duty" (G 4:407). The claim is this: the empirical provides no unambiguous examples from which the concept of acting out of duty might be drawn (G 4:407). Even in cases where a person *seems* to have acted out of duty, there is always the possibility that she has acted, perhaps even unbeknownst to herself, on non-moral grounds. But – and this is the crucial point – this fact does not lead us to call the correctness of our concept of duty into question. Indeed, we (try to) measure empirical reality against its standard, rather than gleaning the standard from empirical reality. The concept of duty therefore cannot be an empirical one.

Nor, Kant argues, can other distinctly moral concepts. We may never, Kant writes, encounter true virtue (G 4:407), but this does not impugn the

²⁰ Kant's technical concept of 'experience' (*Erfahrung*) is closely related to his technical concept of 'cognition' or 'knowledge' (*Erkenntnis*). In Kant's technical sense, to 'cognize' or 'know' something is to assimilate it to a systematically interconnected Newtonian world of sense experience. (*Erkenntnis* is translated as 'cognition' in the new Cambridge Edition of the Works of Immanuel Kant, and in other newer translations, which helpfully gives the term a more specialized, technical ring than 'knowledge.')

Both technical restrictions are consistent with Kant's insistence on the difference between (empirically cognized) phenomena and (intelligibly grasped) noumena: we 'experience' and 'cognize' (or 'know') the former, in Kant's technical senses, in ways we cannot 'experience' or 'cognize' (or 'know') the latter. Kant's point is that we do not have sense data that help us fit free will or its activity into a theory of spatio-temporal objects governed by physical laws; his point is *not* that we cannot make sense of or think true things about the free will. As we have already seen, we certainly have all sorts of thoughts and commitments, and make all sorts of truth claims (like those made by 'common rational cognition') about morality – we 'know,' in an ordinary sense, a lot about the free will, and as we will see here, we 'experience,' in an ordinary sense, the moral law's demands routinely, even if we do not 'know' or 'experience' them in Kant's technical senses.

concept or alter our sense of its validity. Kant gives another nice example: “although up to now there may have existed no loyal friend, pure loyalty in friendship can be no less required from every man” (G 4:408). Hence, it must be that the “true original” of moral concepts “resides” not in experience, but “in reason” (G 4:409). The empirical world is inadequate as a model of the moral concepts we in fact possess and routinely employ.

The inadequacy of nature as the site of moral value

We turn now to the thought that goodness does not inhere in the empirical world as such. We saw earlier that particular effects – realized empirical objects and states of affairs – are good just insofar as they are the effects of a good will. Goodness, for Kant, does not inhere in them, *qua brute empirical objects*, at all. They are good just insofar as they manifest or signify or express or are the product of the good intentions of a free will. Recall Kant's objection, cited above, to finding value in effects per se since effects could have been “brought about by other causes,” that is, by causes other than “the will of a rational being” (G 4:401). Moral value is not, on this reasoning, to be found *in* the natural world, *qua* natural world, at all.

But this should give us pause. Kantian practical reason, after all, “expect[s] its ideas to have effects in experience” (A548/B576). But then isn't this to say that moral value *is* (or at least can be) realized in the empirical world? That nature is (or at least can be) the site of moral value? Kant's position, I think, must be this: only insofar as an empirical event or state of empirical affairs is regarded from a point of view that is *not* Newtonian, from a point of view that describes it as a product of free rational noumenal – and hence profoundly unnatural – will, can it be understood as an ‘effect’ of a good will, and hence as something of moral value.

To describe an empirical event or state of affairs as the effect of a free rational noumenal will is, of course, to describe such a thing in a peculiar way, as an effect of an unconditioned cause, but Kant is committed to this peculiarity. Kant writes, for instance, that if we investigate actions from the point of view of practical reason (or, “in a practical respect”), we “find a rule and order that is entirely other than the natural order” (A550/B578). Kant continues his thought:

For perhaps everything that **has happened** in the course of nature, and on empirical grounds inevitably had to happen, nevertheless **ought not to have happened**. At times, however, we find, or at least believe we have found, that the ideas of reason have actually proved their causality in regard to the actions of human beings as appearances, and that therefore these actions have occurred not through empirical

causes, no, but because they were determined by grounds of reason. (A550/B578; Kant's emphases)

Much could be said about this passage. Kant's central point here is that we must regard events in a light other than that furnished by our theories about nature, since unless we do so, we have no way to make out a difference between what *did* happen and what *ought to* have happened. But the follow-up comment is important too. "At times," Kant writes, "we find, or at least believe we have found, that the ideas of reason have actually proved their causality in regard to the actions of human beings as appearances" (A550/B578; see A547/B575). Finding, or believing we have found, that ideas of reason have caused things is what makes moral judgment and deliberation possible. Things that happen, if they are to be subject to a moral 'ought' at all, must be understood according to "a rule and order that is entirely other than the natural order" (A550/B578). Describing things according to such other rules and orders is just what we do – all the time, everyday – when we describe them as results of freely rationally intended willing.

We can see this at work in a now familiar context. Here, again, is Kant arguing that effects, as such, do not have worth. "Skill and diligence in work," Kant writes,

have a market price; wit, lively imagination and humor have a fancy price [*Affektionspreis*];²¹ on the other hand, fidelity in promises and benevolence from basic principles (not from instinct) have an inner worth. (G 4:435)

Fidelity in promises and benevolence from basic principles are morally special; they have "not merely a relative worth, that is, a price, but an inner worth, that is, *dignity*" (G 4:435; Kant's emphasis). But "their worth," Kant writes,

does not consist in the effects arising from them, in the advantage and use they provide, but in dispositions, that is, in maxims of the will that in this way are ready to manifest themselves through actions, even if success does not favor them. (G 4:435)

This is the familiar thought: dispositions, not effects, are what matter when it comes to specifically moral worth. "Nature," Kant writes, "as well as art, contains nothing that, lacking these, it could put in their place" (G 4:435). But here is where we see this now familiar thought generating a new result.

²¹ Something with a 'market price,' Kant tells us, is "related to general human inclinations and needs," whereas something with a 'fancy price,' "conforms with a certain taste, that is, with a delight in the mere purposeless play of our mental powers" (G 4:434–5); something with a 'fancy price' is, in other words, valued for its aesthetic qualities.

It must be granted that faithfully kept promises or benevolent gestures *are* things we experience – they do ‘appear,’ at least in an ordinary if not in a technical Kantian sense. There are specific events, particular configurations of sound, light, and matter, that ‘count’ as keeping a promise or being benevolent. But counting them thus, we can now see, must involve understanding them as having been caused by the special dispositions of a noumenal free rational will, in an ‘unnatural’ way which somehow runs parallel to natural causation. “Of the faculty of such a subject we would,” Kant writes, “form an empirical and at the same time an intellectual concept of its causality, both of which apply to one and the same effect” (A538/B566). Or, again, as we saw above:

At times, however, we find, or at least believe we have found, that the ideas of reason have actually proved their causality in regard to the actions of human beings as appearances, and that therefore these actions have occurred not through empirical causes, no, but because they were determined by grounds of reason. (A550/B578)

In these cases, where effects seem caused (non-mechanically, ‘unnaturally’) by reason, effects have value as intimations that a good (noumenal) will lurks behind. They are not therefore worthless. They are indeed the kinds of things to which we ought to point children, as part of moral training (KpV 5:154–6); they make for good conversation, as they awaken the deep interest we all have in morality (KpV 5:153);²² they are the kinds of things encounters with which strengthen our own commitment to morality, and are in any event the only available basis on which to judge ourselves and each other morally.

In one sense, then, it is too strong to say that nature is not the site of moral value – on Kant’s view, a free rational will shows itself in empirical effects. But in another sense, it is correct, for Kant, to say that moral value is never in nature *as such*, nature regarded or cognized *as nature*. In nature, there is just matter that follows natural laws. Insofar as we experience the world as a world of moral value – which we do, a lot – we are experiencing the world as a set of objects and events described by concepts and laws that do not belong to the systematic interconnected causal nexus that is Kantian nature. The fact that we can, and do, experience (in the ordinary English, non-technical sense) the world in terms other than those appropriate to

²² “Now, of all arguments there are none that more excite the participation of persons who are otherwise soon bored with subtle reasoning and that bring a certain liveliness into the company than arguments about the *moral worth* of this or that action by which the character of some person is to be made out.” (KpV 5:153)

describe nature as such is, I take it, what it is for there to be a 'noumenal world' – a world grasped by the intellect, rather than by the senses. A 'noumenal world' is not someplace else, but is *this* world (or parts of this world), understood in a particular set of rationally warranted, non-natural terms. One such set of terms is the set of terms that belong to moral discourse generally, which comprises all 'juridical' or legal concepts as well as concepts of ethical virtue.²³ The key point for us here is that Kant argues, and has good reason for arguing, that moral value is not realized in the natural world *as such*.

The inadequacy of empirical motivation to morality

We turn now to our ordinary experiences of moral motivation, and to Kant's thought that nature cannot furnish anything adequate to account for these experiences. We routinely experience ourselves as pulled to do the right thing but, Kant thinks he can show, no natural motive can constitute that pull. To bring the sort of experience in question to the fore, Kant, in the *Critique of Practical Reason*, offers a thought experiment. Kant imagines a man threatened by his ruler with death unless he bears false witness against an innocent person. (We saw this example earlier, in [Chapter 4](#).) We are asked to put ourselves in his position. We know, as the man knows, that he ought to tell the truth (whether or not he will have the moral strength to do so) (KpV 5:30). This 'ought' tugs at him. But what can account for this tug? Nature has provided the man, as it has provided all living things, with the desire to go on living – but this natural motive pulls *away* (in this case) from moral action. Other natural motives likewise fail to tug, here, toward truth-telling: telling the truth (and not bearing false witness) will cost the man his life, and since being alive is a precondition of satisfying natural sensuous desires, nature can find no toehold to tug in this direction. Maybe someone in this situation would resist giving false testimony out of concern for his posthumous reputation, and perhaps this concern could be understood as a 'natural' motivation (what Kant sometimes calls 'glory'). But this suggestion fails on two counts. First, it leaves unaddressed the question of why someone who sacrifices his life rather than bear false witness earns a good reputation – that is, it simply postpones the question of why one should side with truth-telling here, and so of what, in the end, tugs any of us against giving false testimony. Second, if our moral experience is anything

²³ The contours of the noumenal are, of course, a large and controversial topic, and I cannot defend my interpretation fully here. See Jennifer Uleman, "Everyday Noumena: The Fact and Significance of Ordinary Intelligible Objects," unpublished manuscript.

like what Kant thinks it is, the 'ought' in question here would be felt even under circumstances where there is no chance for posthumous glory (if one is being held, say, in an isolated detention center). No 'natural' motivation, no empirical object can, in these circumstances, be the source of someone's experience that he ought, no matter what else, to resist the corrupt ruler, even if it will cost him his life. Kant generalizes from cases like these, where the demands of duty (as commonly experienced) offer no plausible gain to the agent as a sensuous organism. He concludes that there must, therefore, be an alternate, non-empirical source of 'tug,' an alternate source of motivation or of reasons for action.

Punishment

In the *Critique of Practical Reason*, Kant makes an interesting argument against natural grounding that is based in reflections on punishment. His argument appeals to those (and Kant thinks this is most, if not all, of us) who have retributivist intuitions about what justifies punishment. The core retributivist intuition is that harm deserves harm. Punishment hurts, and so, for the consistent Kantian retributivist, must be justifiable as "mere harm in itself," independent of any "benevolent intention" on the part of those inflicting punishment, such as character reform or deterrence (KpV 5:37). The retributivist intuition is that it is right, that it is justified, to inflict harm, as such, on those who have harmed others. But this intuition only makes sense, Kant argues, on a criterion of rightness, of justification, independent of any criteria nature furnishes. Nature tells us that things are good if they *feel* good – if they cause, that is, sensuous pleasure. But on such a criterion, no infliction of pain or harm could ever be justified as such – retributivist punishment would just represent a second wrong, which could never make a right. Worse still for the naturalist, Kant argues, given a standard of right or justice based on pleasure and pain, punishment itself becomes a kind of absurdity: on a pleasure-based criterion an act might (at least in some cases) only be judged a *crime* (i.e., a bad act) once the punishment had been inflicted, because (at least in some cases) only then would the harm, all told, outweigh whatever pleasure the initial act provided the criminal (KpV 5:37).

Now, it is true that sophisticated utilitarians and others who might side with a pleasure-based criterion are not as helpless to justify punishment as Kant suggests – they need only advert to some overall increase in pleasure that punishing accomplishes, by deterring criminals, for example. But Kant thinks our practices do and should reflect widespread acceptance of the retributivist intuition that criminals deserve to be harmed, *regardless* of the

overall consequences this harm has for human happiness.²⁴ Insofar as he is right about this, Kant is right to argue that punishment, as we commonly understand and practice it, requires that we be able to advert to a standard of rightness *independent of and overriding* pleasure or pain or any other empirical goods. Moral practices like punishment demonstrate that we avail ourselves of non-natural standards.

The qualitative difference between morality and prudence

Further evidence of nature's inadequacy when it comes to grounding morality is found in our, for Kant, sure sense of a distinction between morality and prudence. Prudence, for Kant, is a kind of practical wisdom aimed at attaining happiness, that is, at satisfying some maximal set of sensuous desires. Prudential principles help us do this by helping us pursue our own empirical ends in ways that are efficient and successful; they are practical principles that serve the cause of self-interest. As such, they are grounded in nature. But Kant thinks we have a sure sense that there is another sort of practical principle, and that this other sort is specifically moral.

Kant offers examples. In the *Critique of Practical Reason*, Kant asks us to imagine our response to an acquaintance who tries to justify, "having given false testimony by first pleading what he asserts to be the sacred duty of his own happiness" (KpV 5:35). The acquaintance goes on to recount the advantages that accrued to him as a result, and "affirm[s], in all seriousness, that he has fulfilled a true human duty" (KpV 5:35). "You would," Kant writes, "either laugh in his face or shrink back from him with disgust" (KpV 5:35). Prudentially savvy as his actions may have been, the acquaintance's claims to having fulfilled a duty seem to Kant, and Kant thinks will seem to us, absurd and repulsive. Prudence and duty, self-interest and

²⁴ Kant sees retributivists as taking the moral high-road compared to those who would punish in order to benefit the offender (via reform) and/or society (via deterrence) (see MS 6:331–7). Retributivism, thinks Kant, respects and upholds the criminal's dignity more adequately than do penal schemes that fiddle psychologically or sociologically with the terms of her agency. In its "he asked for it" stance, retributivism assumes and seeks to uphold a strong conception of free agency, which is of course congenial to Kant. Kant's full understanding of punishment, moreover, invokes the idea that one *should* be happy just in proportion to one's 'well-doing.' In punishing a person, the community holds up this rational ideal and attempts to enact it in practice. See Kant's comment at KpV 5:61: "When ... someone who delights in annoying and vexing peace-loving folk receives at last a right good beating, it is certainly an ill, but everyone approves of it and considers it as good in itself even if nothing further results from it; nay, even he who gets the beating must acknowledge, in his reason, that justice has been done to him, because he sees the proportion between welfare and well-doing, which reason inevitably holds before him, here put into practice." Here, the gap between what is pleasant and what is right is evident once again.

morality, come apart. Another example: if someone recommended a steward, praising him, “as a prudent human being with masterly understanding of his own advantage and also as an indefatigably active one, who lets pass no opportunity to advance it” (KpV 5:35), you would, Kant concludes, “believe either that the recommender was making a fool of you or that he had lost his mind” (KpV 5:36). The point is not just that we would feel our own interests threatened by someone single-mindedly dedicated to his own, though most of us surely would. The problem, even more significantly, is that the person recommending the prudential steward has misidentified the virtues we would hope for in a steward, or in anyone we consider admitting into close company. The appalling thing – “you would believe either that the recommender was making a fool of you or that he had lost his mind” – is that we expect to have held up (for general endorsement, and for admission to trust) precisely people who both feel and are routinely moved by demands *other than* those of furthering their own personal ends. We expect characters to be recommended to us because of their sensitivity and susceptibility to demands *other than* those of inclination. A final example: Kant imagines a man who has won at cards by cheating. The man compares his own actions against his principles and finds himself wanting. The fact that he can find himself wanting is the thing. In order to be able, Kant writes,

to say to himself, “I am a *worthless* man although I have filled my purse,” he must have a different criterion of judgment from that by which he commends himself and says “I am a *prudent* man, for I have enriched my cash box.” (KpV 5:37)

Since the card player is able to distinguish moral worthiness from prudential success, there “must, therefore, be something other than the principle of one’s own happiness” at work (KpV 5:37) – there must be moral principles separate in kind from prudential principles.

All these examples are intended by Kant to show that we routinely recognize and feel the pull of a standard of conduct qualitatively distinct from any which essentially refers or appeals to empirical needs, desires, pleasures, outcomes, or advantages – which is what all prudential principles do. Prudence is not duty; self-love and self-interest, in all their natural inevitability, often conflict with moral demands. Kant writes: “So distinctly and sharply drawn are the boundaries of morality and self-love that even the most common eye cannot fail to distinguish whether something belongs to the one or the other” (KpV 5:36). The law of nature that provides for self-interest, that leads us to the prudential principles conducive to happiness, is not all there is.

The argument so far, and Kant's endorsement of freedom and rationality

So, for Kant, the data of moral experience militate against a natural grounding for morality. Kant argues, as we have just seen, that we experience real distinctions between moral and empirical concepts; that we routinely contrast (never fully realized) moral ideals with (observable) realities; that we routinely distinguish between the (potentially) morally worthy activity of free rational wills and the value of material, spatio-temporal, causally conditioned empirical consequences, which (I argue) is tantamount to denying moral value to anything in nature as such; that we experience a gap between moral and empirical motivation, routinely contrasting the tugs of duty with those of inclination; that our practices of punishment make sense only if we grant morality a non-natural foundation; and finally that we routinely distinguish between the demands of morality (duty) and of empirical well-being (prudence).

Kant stakes these claims on what he takes to be ordinary human intuitions about morality. The “clear conviction” that, “reason by itself and independently of all appearances commands what ought to happen,” and moreover, commands “inflexibly” (G 4:407–8) is of a piece with our confidence that, e.g., pure loyalty in friendship is demanded, whether or not it has ever occurred. Kant’s strategy for showing that morality can receive no natural grounding lies in showing that the possibility of the will’s determination according to something other than empirical concepts or empirically based laws is the *sine qua non* of our moral intuitions, experiences, and practices.

Of course not everyone shares these intuitions, claims these experiences, or endorses these practices. Kant does not have a knock-down, drag-out argument suitable to force acceptance of his account. People can choose, and some have chosen, to believe that we are creatures whose fates are not our own, or that we are all just acting on instinct, or that feelings of duty are just misplaced guilt. Moral philosophy is in fact full of attempts to articulate such views. Kant’s case, in the end, rests on a claim that such articulations fail to explain or to answer moral longings many of us in fact have. Most of us, if we are honest with ourselves, will, Kant thinks, find views grounded in nature confusing, unsatisfying, and unattractive (G 4:409–11; see also G 4:426). Most of us want to think we can take a view of things that sees beyond the ‘dear self.’ Most of us want to think that human life and action can answer meaningful, and not just instinctual, demands. Kant’s rejection of an empirical basis for morality is staked on this. If people don’t, or haven’t, or won’t, see what the rest of us see, feel the tugs the rest of us feel,

distinguish between prudence and morality, attend to intention over effects, etc., like the rest of us, there is not much we can do – not much, that is, beyond rehearsing examples meant to awaken them to distinctions we see and feel and act on.

The closest thing Kant has to a knock-down argument is his argument from moral responsibility.²⁵ Without a distinction between free reason on the one hand and nature on the other, we would not be able to distinguish the (free rational) self from the (natural) other in the radical way required to distinguish self-determination from external determination. Without the ability to distinguish self-determination from external determination, we lose the possibility of metaphysically radical responsibility. But we need metaphysically radical responsibility, since it makes blame and praise possible and also insulates us against ‘moral luck,’ that is, against our moral fates being subject to the (for us) unpredictable and brute facts of nature and its vicissitudes. Given the possibility of radical self-determination, I can be held accountable on the one hand, and on the other I can achieve moral goodness whether or not nature is poised to cooperate with me, because authoring and freely subjecting myself to the moral law is always within my power.

Knock-down arguments aside, in drawing our attention to the intuitions, experiences, and practices that confirm a strong distinction between nature and free rational will, Kant is doing more than confirming a metaphysics. Kant is also *siding* with reason and freedom, and trying to show us that we do the same. Kant's rejection of a natural grounding for morality is an *endorsement* of the attachments we have to actions that make sense (that are rational), and to our own free power (that are freely willed). We find the will that rationally, freely, determines itself *good*. We feel respect for the moral law, and for ourselves as its authors. We think there are things more important than pleasure and survival. For Kant, unnaturalness elevates us. Unnaturalness holds out the possibility of an existence that is not, at bottom, senseless and coerced, but that may approach, if asymptotically, full rationality and full freedom. How much better a world, and how much better the lives, Kant thinks, made hereby possible! Unnaturalness helps explain the high points and impressive actions of human history, grounding our admiration: we are able, sometimes, to say ‘look, they did that, freely, and because they saw it was best!’ For Kant, insisting on a split between

²⁵ It is a knock-down argument *given* Kant's understandings of nature and of reason, and given the moral intuitions and practices he seeks to accommodate. It is also the argument on which he relies most heavily in the *Critique of Practical Reason* to prove freedom: see KpV 5:28–9.

nature and the activity of free rational will keeps this possibility alive. It rejects a worldview drained of resources to explain or encourage the human activity of world-making, of believing in the possibility of the *new*. Kant's anti-naturalism represents his own deep choice to side, to the extent philosophy will allow, with responsibility, potency, agency, sense-making, and rational shaping, and to reject fatalism, impotence, passivity, chaos, and hopelessness.

We have, for the second time, reached a kind of conclusion. Our first followed the analysis of duty, and emphasized Kant's claim that only the activity and products of a free rational will occasion true respect; the second conclusion, here, again emphasizes the ways we routinely recognize something 'unnatural' at the root of moral value. We might, again, leave the case here. But there is one final piece of Kantian terrain that bears examination.

The special case of happiness

Happiness, mentioned above as the good served by prudence, warrants its own discussion. It warrants this despite the fact that everything we have seen Kant argue so far already tells against any moral theory based on happiness, at least as Kant understands it, since for Kant, happiness is fundamentally an empirical good: by Kant's lights, arguments against the very possibility of natural moral ends or naturally grounded principles should be sufficient to sink happiness-based views. But Kant rejected happiness often and on many other grounds; happiness has, historically, been a popular foundation for moral theory, and Kant wanted to defeat it specifically, definitively, and over and over. What is interesting for us, and what justifies a further look, is that many of the specific arguments Kant mounts against happiness do not really work by themselves. What we see instead is that happiness's real failing, for Kant, always lies in its relationship to nature.

How, to begin, does Kant understand happiness? The end of happiness represents "the unification of all ends that are given to us by our inclinations into [a] single end" (A800/B828); in the idea of happiness, "all inclinations unite in one sum," or systematic unity (G 4:399). And to attain happiness is to satisfy this sum of inclinations. Happiness is "the state of a rational being in the world in the whole of whose existence *everything goes according to his wish and will*" (KpV 5:124); happiness is getting what you want, as much of the time as is possible. Kant characterizes happiness, not surprisingly and in a phrase we saw earlier, as a feeling of "complete well-being and satisfaction with one's condition" (G 4:393).

In both the *Groundwork* and in the *Critique of Practical Reason*, Kant offers a host of arguments, all of which aim to show that happiness fails to meet common-sense criteria for being a supreme moral good. The meaning of 'happiness' is difficult, if not impossible to determine (G 4:399, 417–18); it is variable both within and across persons (G 4:417–19; see also G 4:444; KpV 5:25, 28); the contents of anyone's conception of happiness are always contingent (KpV 5:25–6); happiness has therefore nothing necessary or universal about it, and cannot be grasped a priori (KpV 5:26); happiness cannot, for any and all of these reasons, ground a categorical imperative (G 4:418–19; see also KpV 5:28). And not being able to ground a categorical imperative is tantamount, recall, to not being able to ground a moral principle, which in turn is tantamount to not being the supreme moral good. Now happiness, as Kant understands it, is not in itself bad or immoral. We in fact have duties to promote our own as well as others' happiness (G 4:399; MS 6:387–8; KpV 5:34). But the value of happiness per se is not the ground of those duties.²⁶ Kant's arguments are aimed at showing that happiness is not *itself* the moral good, that happiness is no candidate for supreme moral value. Happiness is no candidate because, to recapitulate, it is indeterminate, variable, contingent, not universal and necessary, not a priori, and is (therefore) capable only of grounding hypothetical imperatives, principles of prudence, but nothing truly moral.

Let us see how these charges fare. Look first at Kant's claim that happiness is too indeterminate and variable to ground a moral principle. It is true that different things will make different people happy, and that different things will make the same person happy at different times. But it is also true that different things will constitute doing one's duty for different people, and for the same person at different times. At a sufficient level of generality, happiness seems determinate enough: happiness is getting as much of what you want as much of the time as is possible. We cannot say ahead of time what this general description will dictate in real cases, but the same will be true for any moral principle, which must take all sorts of contingent facts into account, both subjective and objective, before it can take any determinate ('do this specific thing') form. The indeterminacy and variability charges seem to tell against Kantian moral principles as well as against principles grounded in happiness.

²⁶ We are obliged to promote happiness, for Kant, for two reasons. First and most importantly, helping others attain their empirical ends is a way of respecting their wills; bolstering others in their efforts to pursue their projects represents a deep nod to their free rational agency (which we have a hard time promoting as such) (MS 6:386, 387–8). Second, as a matter of rough empirical fact, happy people have an easier time doing their moral duty (MS 6:388).

Consider now Kant's claim that happiness is too contingent to be the moral good, that it fails for not being universal and necessary. The problem with this argument is that as an end, happiness is *not* contingent; the end of happiness is, Kant often acknowledges, necessary for "all rational beings," provided just that they are "dependent beings," that is, not God (G 4:415). Happiness as a purpose or end is universal too: it can be "presupposed surely and a priori in the case of every human being" (G 4:415). Its specific contents for me or for you are no doubt contingent on contingent facts about each of us – but again, this will be equally true for the demands of duty. Happiness in fact looks like a universal and necessary end. Indeed, Kant has given us the *law* aimed at pursuing it in the law of self-love, and told us that this law is a law of nature – and we know that all Kantian laws carry universality and necessity. Kant's rejection of happiness as a contingent and not universal and necessary end does not stand up to scrutiny.

Kant tells us that happiness cannot ground a categorical imperative, even if it can be presupposed to be an end at least for all human beings, because whatever action an imperative directed at happiness commands "is not commanded absolutely but only as a means to another purpose" (G 4:416). But this argument is also weak, for if one is willing to count an abstract aim as general as 'happiness' as 'another purpose' (as in, 'such-and-such an action is not commanded absolutely but only as a means to happiness') one can, by the same token, argue that imperatives of duty are hypothetical: they do not command their actions absolutely but only as a means to another purpose, viz., that of fulfilling my duty. The moral principle does not, for example, command me 'absolutely' to stop at red lights or to buy flowers for my friend, but commands these things, insofar as it does, just as means of realizing my most general moral duty. Now Kant could rightly object that to put things this way is to fail to recognize the deep difference between categorical and hypothetical imperatives, namely, that categorical imperatives are grounded in an end we *must* have. But then here, again, an imperative grounded in the end of happiness looks categorical enough.

The real problem is this. The good of happiness, no matter how abstractly conceived, no matter how necessary and universal a desire, no matter (therefore) how suitable to ground an a priori principle, a law, a categorical imperative, is itself always for Kant a thing of nature. Happiness, as Kant understands it, is dependent on our sensuous sides – and this means, for Kant, that happiness always depends on sides of us that are, in key respects, coerced and arbitrary or senseless. How so? Strivings the satisfaction of which would constitute happiness are, for any one of us, *given* – they are not under our control, but are given by our sensuous

constitutions. We are free to pursue them or not, but we are passive and unfree vis-à-vis the set of inclinations nature hands us: we are coerced by nature in seeking happiness. The pursuit of happiness is, moreover, arbitrary or senseless for Kant. Happiness answers no grand rationale – it is a fact that some things give us pleasure, satisfy our inclinations, but their satisfaction satisfies no inherently rational design (see KpV 5:26). Even if the things that make us happy also tend to promote our survival (Kant rightly notes that the evidence is ambiguous: KU 5:430–1), our survival itself is not demanded by reason (though our survival as a means to our continued free rational agency may be) (see G 4:395, 397–8). To add insult to injury, as it were, the success of any striving after happiness is in significant part a matter of luck, of the contingent cooperation of a nature that is fundamentally indifferent to us (see KU 5:450). It is not so rational for me to put too much stock in happiness.

But then the real failings of happiness are other than its indeterminacy, contingency, and failure to ground categorical imperatives (since, looked at in the right light, it has none of these failings). For Kant, happiness as an end really fails as a candidate for supreme moral good because it leaves too much to nature, which is to say, to the brute, mechanical, indifferent, the determined and a-rational order of things that gives us our senses, our sensuous desires, and our sensory impressions. Happiness, as an end, fails to call on or exhibit us *qua* free, rational, active agents. We thus also see that Kant is not necessarily an enemy of what many of us would understand by ‘happiness.’ He *is* an enemy of making mere animal satisfaction, or ‘feeling good,’ one’s highest aim. But this is because they represent determination by nature. Happiness, as Kant understands it, fails as an ultimate end, as something suitable to ground moral practice, because it fails to honor free, rational, active agents as we ought to be honored.

KANT AGAINST NATURE

By way of conclusion, this long and justly famous passage from the *Critique of Practical Reason* bears quoting in full:

Duty! Sublime and mighty name that embraces nothing charming or insinuating but requires submission, and yet does not seek to move the will by threatening anything that would arouse natural aversion or terror in the mind but only holds forth a law that of itself finds entry into the mind and yet gains reluctant reverence (though not always obedience), a law before which all inclinations are dumb, even though they secretly work against it; what origin is there worthy of you, and where is to be found the root of your noble descent which proudly rejects all kinship with

the inclinations, descent from which is the indispensable condition of that worth which human beings alone can give themselves?

It can be nothing less than what elevates a human being above himself (as a part of the sensible world), what connects him with an order of things that only the understanding can think and that at the same time has under it the whole sensible world and with it the empirically determinable existence of human beings in time and the whole of all ends (which is alone suitable to such unconditional practical laws as the moral). It is nothing other than *personality*, that is, freedom and independence from the mechanism of the whole of nature, regarded nevertheless as also a capacity of a being subject to special laws – namely pure practical laws given by his own reason, so that a person as belonging to the sensible world is subject to his own personality insofar as he also belongs to the intelligible world; for, it is then not to be wondered at that a human being, as belonging to both worlds, must regard his own nature in reference to his second and highest vocation only with reverence, and its laws with the highest respect. (KpV 5:86 7)

The “root” of duty’s “noble descent,” the source of our “reluctant reverence” for the moral law, a reverence “before which all inclinations are dumb,” can, Kant writes, “be nothing less than what elevates a human being above himself (as a part of the sensible world),” that is, above himself as empirical (all KpV 5:86). The root of duty, and hence the root of morality, lies in our being supremely unnatural.

We have seen that, for Kant, the only unconditional good lies in a good will; that a Kantian good will is a will moved by duty; that willing out of duty consists not in willing particular empirical objects, but in willing on a special a priori law; that this law is grounded in a good that occasions the peculiar feeling of respect; that nothing in nature, but only (a law describing) the characteristically free rational activity of a free rational will can in fact occasion respect. We have seen that moral concepts cannot, for Kant, be gleaned from nature; that nature strictly speaking is not where value is housed; that moral motivation properly speaking is not natural; that our intuitions about punishment bear out the claim that there are non-natural goods (which justify retribution); that morality and prudence are experientially distinct. We have, finally, seen that Kant’s arguments against happiness must be grounded less in happiness’s inability to fulfill the role of ultimate moral good and more in Kant’s determination to reject natural grounding for morality, which ‘happiness’ purports to offer.

The arguments Kant makes for these claims – arguments based in moral intuitions, practices, and experiences – echo long-standing philosophical calls against naturalist, empiricist views. Seventeenth- and eighteenth-century rationalists routinely mounted similar charges against their empiricist contemporaries. The claim that we cannot get from observables to ideals

echoes the Cartesian argument that concepts like 'infinity' cannot be derived merely from empirical data (which is by its nature finite), but require a 'non-natural' contribution (an innate idea, in Descartes' case).²⁷ The idea that intentions are more important than consequences in evaluating an action echoes a rationalist concern about inner grounds, ultimate rationales, and deep origins, rather than the superficial 'phenomena' that manifest empirically.²⁸ Insisting on differences between the demands of duty and properly moral motives, on the one hand, and the demands of prudence and empirical motives, on the other, is tantamount to insisting that we can be and are responsive to an order of things where 'best' is not the same as most efficient, or functional, vis-à-vis natural ends. Instead, here, 'best' represents 'supernatural' ideals, associated for rationalists with God: benevolence, creative agency, intelligence, justice. Kant's argument from retributivism reflects this last point: retributivism answers an ideal of justice, even if it violates all 'natural' demands. For Kant here, as for seventeenth- and eighteenth-century rationalists, the deliverances of our senses, the observable facts about the natural world and the interests that flow from these are, in fundamental ways, inadequate.

The thread that pulls this all together is the fact that nature is, for Kant, inherently senseless and coercive. It is senseless insofar as it is what it is and does what it does without rationale, intelligent design, final cause, or ultimate purpose. Nature is indeed marvelous, but what astonishes and impresses us, Kant thinks, is precisely that such a blind machine, such a collection of inert matter moving according to mechanistic causation, can be also so rich and elegant (see KpV 5:161–2). (In the *Critique of the Power of Judgment*, Kant explores and assesses the temptation, arising from just this marvel, to attribute purposes to nature [see KU 5:180–1].) Nature is coercive insofar as there can be no real freedom in it, no uncaused causes, no creation, nothing moved from itself alone, but always only the cumulative effects of an infinite network of causes that necessitates in all directions. In nature, there is therefore no responsibility, no sui generis agency, no possibility for a self that is ultimately more or other than a particular sort of locus in a deterministic causal nexus. Kant's anti-naturalism is, at base, a deep rejection of a ground for action at once senseless – arbitrary, blind – and coercive – deterministic, forced from outside, unfree.

²⁷ See, for example, René Descartes, *Meditations on First Philosophy with Selections from the Objections and Replies* [1641], trans. John Cottingham (Cambridge and New York: Cambridge University Press, 1986), 31 (AT 45). Cartesian innate ideas are given to us by God.

²⁸ See, for example, G. W. Leibniz, "From the Letter to Des Bosses (1712–16)," in *Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett Publishing, 1989), 197–206.

We can put the point – as Held in the address mentioned at the outset did – in another way. That which is distinctly moral – seeking to do better by oneself and by others, taking responsibility for oneself and one's actions, granting reason and reflection a role in figuring out what is right – is not, for Kant, given its due by efforts to ground morality in nature. These efforts inevitably tempt us to see human action (in and across individuals) as repetitive and predictable, and as something that we can't intentionally 'improve'; to shrug off strong invitations to responsibility; and to be skeptical about the efficacy of moral reason or reflection, that is, of effort committed to ideals of goodness and focused on discerning and honoring instantiations of these ideals. Moral theories grounded in nature tempt us also to 'settle' for the desires we have, resisting calls to examine or rethink them. As Kant sees it, unless we insist on a 'realm' distinct from nature, and on the real possibility of things (like a free will, or God) that operate according to another set of laws, we will be forced to abandon any ideas of ultimate responsibility or rationale, or of goods beyond survival and sensuous well-being.

In articulating the roots of Kant's anti-naturalism, this chapter has, I hope, made clear that Kant's rejection of natural (or empirical) grounds for morality is not a rejection of the idea that there is *some* good which grounds morality – it is just a strong rejection of the idea that this good could be empirical. [Chapter 6](#), to which we turn now, takes up Kant's commitment to the intrinsic value of 'unnatural' free rational agency, or the activity of free rational will, by looking of the moral law's grounding in free rational willing itself.

CHAPTER 6

The categorical imperative: free will willing itself

This chapter aims to defend an alternative to a widespread formalist interpretation of Kant's moral theory. It does so via a close reading of Kant's canonical arguments for his moral law in the *Groundwork of the Metaphysics of Morals* and the *Critique of Practical Reason*. On the interpretation I defend here, Kant's famous categorical imperative urges the 'free will to will itself.'¹ This is of course how Hegel put it. To put Kant's moral imperative this way is not, in my view, nor was it in itself for Hegel, a criticism of Kant, though Hegel of course did criticize Kant. Rather, it is to argue that Kant's moral law, expressed in the categorical imperative, has as its specific end the free rational activity of the will itself.

The first part of this chapter traces the roots of the formalist reading, and shows what Kant's formalism (really) demands, and what it doesn't. My aim is to loosen the grip of a formalist ban on asking what Kant cares about, or what his moral theory aims at, by showing that Kant didn't want to silence those inquiries. Having carved out permission to ask, directly, what Kant cares about, what he values, what his moral theory aims at, I turn to Kant's *Groundwork of the Metaphysics of Morals* and *Critique of Practical Reason* discussions of the moral law (the categorical imperative) and read them with these questions in mind. Reading in this way reveals, among other things, grounds for endorsing Kant's famous claim that all his *Groundwork* formulations of the categorical imperative are "at bottom" the same (G 4:436). They are the same, we discover, precisely because all urge us to respect and promote free rational willing, in ourselves and in others.

So to begin: against what reading of Kant am I arguing here? What, in other words, *is* the interpretation I am calling 'formalist'? And why and how should we loose ourselves from it?

¹ G. W. F. Hegel, *Grundlinien der Philosophie des Rechts* (*Werke* 7) [1821] (Frankfurt: Suhrkamp Verlag, 1970), §27. Translation: *Elements of the Philosophy of Right*, trans. H. B. Nisbet, ed. Allen Wood (Cambridge: Cambridge University Press, 1991), 57.

KANT'S FORMALISM

Kant as wholesale formalist

In both the *Groundwork* and the *Critique of Practical Reason*, Kant derives the *content* of the supreme moral law, the categorical imperative, from the *form* that moral principles must take (see KpV 5:26–7 and G 4:420–1). This move, on any (including Kant's own) account, constitutes one of the most distinctive achievements of Kant's practical theory. It is also the founding moment for construals of Kant's moral theory as 'formal' or 'formalist.'

The story, as Kant himself tells it, is this. Before Kant, most if not all practical theorists concerned themselves with identifying or discovering that object or set of objects corresponding to a "supreme concept of good" (KpV 5:64). This supreme or ultimate good was presumed to lend value to all other, subordinate goods (KpV 5:64–5). Moral imperatives were to be grounded in – motivated by, justified by – *its* value, and would consist in instructions toward its attainment, whatever it was (see KpV 5:40; VE 29:621–9). A prime example would be Aristotle's eudaemonism, which nominates happiness or well-being as the supreme good, regards all other goods as derivative and subordinate to this good, and urges courses of action just insofar as they contribute to our well-being or 'eudaemonia.'

Kant's objection to all such attempts is that they always ended up grounding morality in the *desirability* of the objects allegedly constitutive of the supreme good (KpV 5:41). They sought, in Kant's words,

an object of the will in order to make it into the matter and the ground of a law (which was thus to be the determining ground of the will not immediately but rather by means of that object referred to the feeling of pleasure or displeasure). (KpV 5:64)

By identifying, as the determining ground of a moral will, an object, "referred to the feeling of pleasure and displeasure," such attempts end up making pleasure the true, ultimate ground for morality. But, as we have seen, Kant can never be satisfied with a view that grounds morality in pleasure. Such a view will never be able to distinguish duty from natural inclination – and, as we saw in detail in [Chapter 5](#), distinguishing duty from natural inclination is something Kant thinks we all routinely, and rightly, do.

Kant's insistence on the distinction between duty and natural inclination requires, perhaps most obviously, the rejection of all theories grounded ultimately in happiness, since happiness, for Kant, as we also saw in [Chapter 5](#), consists in the maximal jointly possible satisfaction of inclinations

(KpV 5:64; see also KpV 5:35–41). As Kant reads the history of moral philosophy, it also requires the rejection of views based, “in perfection, in moral feeling, or in the will of God” (KpV 5:64; see also G 4:408–10 and KpV 5:152). What are these views? For Kant, “perfectionist” theories, which he associates with Christian Wolff and the Stoics (KpV 5:40), value “the fitness or adequacy of a thing for all sorts of ends,” and urge development of human talents and skills (KpV 5:41; see also G 4:442–3). But Kant thinks these views, on a little inspection, can be seen *really* to value those (pleasure-producing, and hence sensuously desired) ends in relation to which perfection (adequacy, fitness, efficacy) is defined; perfectionist theories are thus truly, if surreptitiously, grounded in some desire for pleasure (KpV 5:41). Theories grounded in moral feeling, such as the ‘moral sense theory’ Kant associates with Francis Hutcheson (KpV 5:40), would have us aim ultimately at *satisfying an inclination to be virtuous* (KpV 5:38) – which patently makes morality a matter of pleasure. (Such theories also fail, for Kant, to address the important question of the ground of our inclination to virtue [KpV 5:38].)² Religious moral theories that promise everlasting happiness (in an afterlife) in exchange for compliance with the will of God similarly ground (motivate, justify) moral action in sensuous inclination (KpV 5:41; see also REL 6:3): we want to feel good in the hereafter. The situation, Kant seems to argue, will be essentially the same for all other moral theories grounded on a supremely good object: all will be based, ultimately, whether they realize it or not, on pleasure.

But of course susceptibilities to pleasure, like susceptibilities to any feeling, depend on our sensuous constitution. A theory grounded on an object valued because it promises pleasure forfeits a crucial desideratum of Kantian moral theory, viz., the claim to ground necessary and universal imperatives. The worth accorded pleasure-promising objects can never be necessary or universal, but must always remain contingent, for, as Kant writes, “only their mere relation to a specially constituted faculty of desire on the part of the subject gives them their worth” (G 4:428). No matter how frequently or generally desired, such objects can ultimately only give us (in other Kantian words) hypothetical imperatives, advice about how to attain something, should we in fact desire it. But we want to ground a categorical

² Indeed, Kant thinks the moral sense theorist’s ‘inclination to be virtuous’ would, if interrogated, turn out to rest on a rational recognition of the demands of duty – and hence not be a true (sensuous) inclination at all. The moral sense theorists thus not only err in thinking that morality is based in inclination, but also ‘deceive’ by surreptitiously relying on what is in fact a rational recognition and interest, counting it as an inclination for which we have a special feeling (KpV 5:38). Such a move is the sort of thing that offends Kant deeply, as it ‘sensualizes’ (and so both degrades and confuses) something that is in fact rational, which leads, “the insightful, feeling confused and dissatisfied without being able to help themselves, [to] avert their eyes” (G 4:409).

imperative, an imperative that commands *without* a 'should we in fact desire it' clause, an imperative that holds necessarily and universally.

Kant's answer to this challenge is formalist. It consists in his appeal to the difference between the *form* of imperatives and the *material* at which they are directed. Kant asks us to turn our attention away from possible objects of desire, and toward the form of a categorical imperative itself. What sort of 'form' is this? Even without knowing its 'material,' we know that the commands of a categorical imperative are not contingent, but are necessary and universal. To command with necessity and universality is to take *the form of a law*. Now, we know that the will is determined through the adoption of maxims, or subjective principles that guide action. Desirable objects can ground the adoption of particular maxims ('adopt this maxim and attain me!'), and hence determine the will, but the formalist strategy is to see what happens if we leave these objects, this 'material,' aside. Kant's aim is to try to ground the adoption of maxims, and hence determine the will, in another way. But how? If we abstract from possible objects, from all possible material of the will, all we know about a categorical imperative is that it has a law-like form: it commands with necessity and universality. How can this fact recommend – command, ground the adoption of – maxims? Just by insisting that the maxims we adopt themselves 'be fitted' to take the form of law. What will this mean? Just that they be susceptible of themselves being willed *as law* – that they could be *made* laws, that they be *suitable* to be given as law. Hence, Kant's categorical imperative: "So act that the maxim of your will could always hold at the same time as a principle in a giving of universal law" (KpV 5:30).

This is a brilliant and amazing move. Kant makes it in §4 of the first part of the *Critique of Practical Reason* (KpV 5:27), and also in presenting the first formulation of the categorical imperative ("act only in accordance with that maxim through which you can at the same time will that it become a universal law" [G 4:421]) in the *Groundwork*. It is brilliant as it answers the 'common-sense' demands that morality focus on the *way* we will, rather than on the objects at which we aim, and it ties this way directly to universality and necessity – all the while relying strictly on reason and bypassing inclination. It finds a categorical imperative without looking for supremely good objects at all. "It is," as Kant writes:

not a prescription according to which an act should occur in order to make a desired effect possible ... it is, on the contrary, a rule which determines the will a priori only with respect to the form of its maxims. (KpV 5:31)

Kant grounds a categorical imperative in its own form: the categorical imperative commands just that our maxims “conform [to] the universality of a law as such” (G 4:421). Kant has discovered a categorical imperative that can determine the will without reference to objects of the will.

It is on account of this move that Kant’s view can be characterized as ‘formal.’ Kant lends significant support to this characterization, writing, for instance, in the *Critique of Practical Reason*, that:

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. (KpV 5:64)

Only a formal law can command a good will. But then the will, to be good, must not, it would seem, be determined by any matter or content or end or object, but only by form. There is plenty of additional text that promotes such a conclusion. In *Religion within the Boundaries of Mere Reason*, Kant writes that, “morality needs absolutely no material determining ground of the free power of choice, that is, no end” (REL 6:3), and that, “morality can perfectly well abstract from ends altogether, and ought so to do” (REL 6:4); form, not matter (material, ends, objects), is what seems to make a good will good.³ In the *Groundwork*, Kant privileges the ‘universal law’ formulation of the categorical imperative, discussing it first, prefacing his statement of it with “there is, therefore, only a single categorical imperative and it is this” (G 4:421), and recommending it as a “canon” of moral appraisal (G 4:424; see also G 4:436–7). The case for characterizing Kant’s whole view as formalist is mounting. In the *Critique of Practical Reason*, Kant makes a famous remark that underscores the priority of form over potentially good objects:

instead of the concept of the good as an object determining and making possible the moral law, it is on the contrary the moral law that first determines and makes possible the concept of the good, insofar as it deserves this name absolutely. (KpV 5:64)

The law, this passage seems to suggest, does not serve some good end, but itself makes any truly moral concept of the good possible: the formal law makes the good possible, but not vice versa. Kant’s view, one might reasonably conclude, eschews ends, objects, matter, and even ideas about the good, in favor of *the form of lawfulness itself*.

³ Indeed, to look to anything other than form, than lawfulness itself, seems somehow fouling: “one who finds it necessary to look around for some end when [for instance] his testimony is rightfully demanded of him, is in this respect already contemptible” (REL 6:4).

What would be wrong with this conclusion? The above account of Kant's accomplishment in answering past theories and deriving the categorical imperative from the form of lawfulness itself is in order. And it *is* a brilliant and amazing accomplishment. But while brilliant and amazing (for its interest, importance, elegance, and productivity), it can be, and I think too often is, read in a way that distorts and weakens Kant's view.

If we read Kant as arguing that ends – material, matter, goods, objects – must be abstracted from in order to find a law that carries the universality and necessity requisite to morality, then Kant's accomplishment begins to be read as a victory over *content*. It starts to seem as if Kantian morality must refuse attention to content – to ends, material, matter, goods, objects – altogether. But this goes too far. Legitimate criticisms loom: a moral law truly without end or object or matter or content would be pretty useless – would, to use Hegel's language, be an 'empty formalism' if there ever was one.⁴ Adding insult to injury, Kant's claim that a focus on form allows him to derive a purely *rational* principle (e.g., G 4:427) can lead to the thought that *reason itself* eschews 'content.' But a 'purely formal' (read, 'contentless') faculty of reason would seem better suited to be the slave, à la Hume, of the passions than to be in the business of issuing, itself, any calls to action.⁵ It is hard to see how Kant's formal categorical imperative could be in the business of moving us to action at all.

What would be a better way of readings Kant's claims? My suggestion is this: the real opposition in Kant's derivation of the universal law formulation of the categorical imperative is not between form and content, or form and object, or form and matter, or form and end or aim, but between *rational* form and *empirical* content (or object or matter or end or aim).⁶ We should not be misled by Kant's eschewal of principles grounded in empirical objects into thinking that the categorical imperative cannot be grounded by *any* object, or end, or matter, or content, or aim at all – even though Kant sometimes writes as if this is what he means. What Kant *really* means, as I will try to show now, is that the categorical imperative cannot be grounded in *the value of some empirical object* (or end, etc.) as such.

⁴ Hegel, *Elements of the Philosophy of Right*, §135, p. 162.

⁵ The famous passage is: "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them" (David Hume, *A Treatise of Human Nature* [1739], ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1978), Book II, Part III, §3, p. 415).

⁶ Kant does not, alas, make any consistent terminological distinction between empirical content or objects and non-empirical content or objects. Kant might, for instance, have reserved 'matter' for the former – but, as G 4:436 makes clear, he did not.

Formalism as a limited strategy

To see the valuable but nonetheless limited extent to which Kant's view can be understood as formal, we should remember, first, that the demand to will only on universalizable maxims is just one way Kant presents his categorical imperative. In the *Groundwork*, Kant offers two additional formulations (some commentators count more), which he famously claims are "at bottom only so many formulae of the very same law" (G 4:436). The formula of universal law – the 'formal' accomplishment – is not the only way to express the basic principle of Kantian morality. This by itself should make us wary of thinking formalism is all there is.

Why then does Kant seem to privilege the formal universal law formulation? Let us revisit the evidence. The comment preceding the universal law formulation that "there is, therefore, only a single categorical imperative and it is this" (G 4:421) should not be read as claiming that *only the universal law formulation* of the categorical imperative is valid. Kant articulates further formulations just eight pages later (at G 4:429), and presumably knew he would do so. The comment must instead be read as the claim that there is a single moral law – "a single categorical imperative" – that finds first expression here (in the universal law formulation). Kant's later insistence that the universal law formulation be used as a canon (G 4:424) should be read simply as advice – this formulation is the most *useful*, the one we are *most able to apply* in ways that yield determinate, reliable results – not as a claim about its being the only avenue to moral truth. So we see, first off, that there is nothing necessarily 'most true' or 'most Kantian' about the purely formal universal law formulation of the categorical imperative, even if it is a distinctive and dramatic accomplishment.

Second, the same writings that give rise to the formalist interpretation are full of references to specifically moral ends, objects, and aims. Kant's second *Groundwork* formulation of the categorical imperative explicitly adverts to *an end* for morality: "so act that you use humanity, whether in your own person or the person of any other, always at the same time as an end, never merely as a means" (G 4:429). "The ground of this principle," Kant writes, "is: *rational nature exists as an end in itself*" (G 4:429; Kant's italics). In the *Critique of Practical Reason*, Kant's now familiar formal derivation of the categorical imperative appears in a chapter entitled "On the Principles of Pure Practical Reason" (KpV 5:19); the chapter is followed by one entitled "On the Concept of *an Object* of Pure Practical Reason" (KpV 5:57; my italics). In the latter chapter, Kant discusses precisely the relationship between the moral law and the good that is the proper object of moral

willing. Kant, in these two key texts, has strategic reasons to establish the law-like nature of his categorical imperative before he talks about ends. But law by itself is not and cannot be the whole story: the moral law, or categorical imperative, understood in its entirety, is not without a determinate end or object, is not ‘purely formal,’ even on Kant’s account.

But then this stands to reason. Recall from [Chapter 2](#) above that for a will to will at all, it must have an object. And this does not just mean that it must aim at some empirical effect, although this of course it must: “in the absence,” Kant writes, “of all reference to an end no determination of the will can take place in human beings at all, since no such determination can occur without an effect” (REL 6:4; see also KpV 5:34). All willed actions, Kant in other words recognizes, pursue the realization of some empirical effect or effects in the world; otherwise they wouldn’t be actions at all. But as we have seen, a Kantian will must have an object *also* in the sense that it must have an ultimate aim, which is perhaps *realized* by empirical effects, but need not be identical with them. The ultimate aim might be sensuous gratification, or it might be accord with morality. Kant’s objection, of course, is to theories that advocate sensuous gratification. Another way to get at the point is this: a practical law, as [Chapter 3](#) made clear, must contain something to recommend it. Any practical law must, that is, be grounded in something we recognize as good; in order to claim our attention, a practical principle must appeal to some interest. If it didn’t, the possibility of its serving as a *practical* principle would not arise – if it didn’t appeal somehow, to some interest, it could not move us to action.⁷ That which interests us in acting a certain way is our overarching aim, and is what *grounds* the determination of the will. Kant’s worry is just that we will try to pass sensuous ends off as moral ones. He does not claim that a moral will *has no end at all*.

The distinction between local empirical ends and ultimate ends is at work in this claim from the *Groundwork*. “All maxims” must, Kant writes, have a *matter*, namely an end, and in this respect the [second] formula [of the categorical imperative] says that a rational being, as an end by its nature and hence as an end in itself, must in every maxim serve as the limiting condition of all merely relative and arbitrary ends. (G 4:436)

⁷ This is why Kant’s answer to his own oft-repeated question, ‘how is a categorical imperative possible?’ addresses, among other things, the question of our *practical interest* in (or motivation for following) the law (see G 4:460n). (For an excellent extended discussion of this point, see Paul Guyer, “The Possibility of the Categorical Imperative,” in his *Kant on Freedom, Law, and Happiness* [Cambridge and New York: Cambridge University Press, 2000], 172–206.)

We see in this passage a clear distinction between “all merely relative and arbitrary ends” and the kind of ultimate end or aim that a maxim also necessarily serves – and of course we also see Kant here urge that the ultimate end or aim served by moral maxims is “a rational being” itself. We will have more to say shortly about this rational being, but the point should be clear: there are ends from which morality must abstract, namely local empirical ends, but these can and must be distinguished from the end that morality itself can and must serve.

What then about the passages from *Religion* quoted above? They seem most troublesome. They are, however, surrounded by other passages that in fact support the very suggestion being made here, viz., that the ends (or objects or matter or material) Kant rejects are all *empirical*. In a note, Kant makes the way he is using term ‘end’ in these passages explicit: “An *end* is always the object of an *inclination*” (REL 6:6n). A nearby comment, noted above, makes sense only if this is the case: “one who finds it necessary to look around for some end when his testimony is rightfully demanded of him, is in this respect already contemptible” (REL 6:4). The comment only makes sense if we understand the ‘contemptible’ person to be looking around for some *empirical* reward. It is contemptible to seek inclination-satisfying payoffs or material advantages for giving testimony when doing so is one’s duty; the problem is the looking around for an end *other* than that of duty for its own sake. The passages from *Religion* quoted above – “morality needs absolutely no material determining ground of the free power of choice, that is, no end” (REL 6:3) and “morality can perfectly well abstract from ends altogether, and ought so to do” (REL 6:4) – now can be read as insisting just that morality steer clear of empirical determining ends.

Kant’s insistence that “the moral law ... first determines and makes possible the concept of the good” (KpV 5:64) is often taken to prioritize the form of law over moral ends. But Kant’s insistence here is insistence that the special object of moral willing cannot be *conceptualized* except via a priori law – not that there is no such object. As Kant himself writes, “this remark ... concerns only the method of ultimate moral investigation” (KpV 5:64). For reasons that have to do with Kant’s understanding of human cognition, and its default dependence on sensory ‘intuition’ or experience, Kant believes that in the absence of special mental discipline and an articulated and guiding moral principle we will conceive only empirical goods. We are doomed, if we simply begin by searching for ‘the good,’ to be guided in our choice of objects by their “immediate relation to feeling, which is always empirical” (KpV 5:64). So, we do better to start by focusing on the law. But this claim is not a claim that the form of lawfulness

as such, for its own sake, is at the base of Kant's conception of morality. Rather, the form of lawfulness is a way to access, for Kant, the non-empirical good at which his morality aims.

The last point is worth further emphasis. Kant's demand that the principle of morality command in a law-like way is central to the popular formalist interpretation. But how much work should it really be asked to do? Kant *does* demand that any supreme moral principle command with necessity and universality. This is, for him, part and parcel of the 'ordinary rational' demand that moral law command unconditionally – moral law must also command everyone, under all circumstances. But Kant's rejection of strategies for grounding moral law in objects of desire – in empirical ends – depends on much more than a commitment to universality and necessity, to law-likeness, for its own sake. Universality and necessity, after all, characterize *all* laws, be they natural or rational. And it seems clear that, for Kant, if we were somehow to discover an empirical object that was desired universally, as a matter (unlikely as this is) of physical necessity, we should still err in claiming that such an object was suitable to ground moral imperatives. But then morality, for Kant, wants more than universality and necessity as such. For Kant, universality and necessity, or law-likeness, *themselves* are neither unique marks of nor primary desiderata for a moral principle.⁸

I have tried here to trace the central roots of a reading of Kant's moral theory as 'formal,' and to show how these roots can be extrapolated in ways that are misleading or distorting. I have argued that they should instead be understood as part of Kant's general rejection of an empirical basis for morality, and *not* as evidence that Kant would have us develop a morality that eschews attention to ends or aims or objects. I hope hereby to have made it legitimate to ask what Kant cares about, what he values, what his moral theory aims at – and to begin to address those questions. The sense in which Kant's view *is* 'formal,' however, should first be reiterated: Kant's moral law is formal in that, in one of its formulations, it is not grounded in the attractiveness of any empirical matter of will, but rather in the attractiveness of a certain form of willing. That form is the form of law-like-ness itself – or more precisely, as we shall see, of *law-making* or legislation itself. Reason, without need of empirical incentive, has grounds for endorsing law-like willing. But

⁸ We saw a version of this point in [Chapter 5](#) above, when we argued that the trouble, for Kant, with happiness is not that it fails to be a universal or necessary end, but that it is a fundamentally *natural* end. Its naturalness, not its (supposed) inability to generate genuine universality and necessity, is happiness's problem, for Kant.

what are these grounds? In what sense is this an attractive form? Why should we want to will – why should morality consist in willing – on maxims that conform to, or express, or make, law? These are the questions with which the rest of this chapter will be concerned. Answering them constitutes my positive argument that Kant is not concerned with form per se, but rather with a form that is uniquely able to respect and sustain a certain sort of content, viz., the free rational activity of the will.

Let me now turn to the task of making good on this claim in more detail.

KANT'S CATEGORICAL IMPERATIVE: ITS FORM
AND ITS CONTENT

We proceed, as promised earlier, by looking closely at Kant's *Groundwork* discussion of the categorical imperative. The strategy is to keep an eye on discerning the categorical imperative's aim – on discerning, in other words, what good is represented in the moral law. We will see that willing according to the categorical imperative simultaneously *instantiates* and *actively respects* the free rational activity of human will – and that it does so by design. We will see, in other words, that instantiation of and active respect for free rational willing is precisely the good at which Kant's moral law aims. Thus does this chapter seek to make good my (now oft-rehearsed) claim that the categorical imperative aims at the free rational activity of the will itself.

People disagree on the number, but in the *Groundwork* Kant takes himself to be offering three formulations of the categorical imperative. They are:

- (1) "Act only in accordance with that maxim through which you can at the same time will that it become a universal law" (G 4:421);
- (2) "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (G 4:429); and
- (3) Act in accordance with "the idea of *the will of every rational being as a will giving universal law*" (G 4:431; Kant's italics).

I will refer, throughout, to these as the first, second, and third formulations, respectively, as well as calling them by familiar descriptive names (the formula of universal law, the formula of humanity as an end in itself, the formula of autonomy). As just noted, my aim here is to show that each of these three formulations describes a (single) principle fundamentally aimed at getting us to instantiate and actively respect free rational willing – and that this shows free rational willing to be the good 'represented' in the moral law. I proceed by attending to features of the formulations themselves as well as to the four

familiar illustrations Kant offers, viz., the prohibitions on suicide, false promising, neglect in developing one's talents, and a general refusal to aid others.⁹

The good represented in the formula of universal law

As we saw above, the universal law formulation of the categorical imperative prohibits action on maxims that cannot be willed as universal laws. As we also saw, Kant tells us that this formulation should serve as the "canon" and "strict method" of moral judgment (G 4:424 and 436 respectively): this is the formulation of the imperative best suited to help us in making actual practical judgments. It asks us, essentially, to judge whether a particular maxim, when 'universalized,' that is, imagined as actual universal law, can consistently be willed. Kant walks us through four examples: he looks at maxims of committing suicide when we are in despair (G 4:421–2), of making lying promises (G 4:422), of failing to develop our talents (G 4:422–3), and of refusing to aid others (G 4:422–3).¹⁰ All, according to Kant, fail the 'test' for universalizability, showing us, "a few of the many actual duties ... whose derivation from the one principle cited above is clear" (G 4:423).¹¹ All these examples will be important for us here.

The ins and outs of the procedure for testing maxims using the universalizability requirement have been discussed at length by others.¹² What is crucial for us here is the notion of *consistency* at the heart of the test.¹³ The universal law formulation of the categorical imperative commands that we ask of any maxim whether we could consistently will that that maxim be held universally, that is, whether we could consistently will that it come to describe, in a law-like way, human practice.¹⁴ What does this mean?

⁹ I have learned a lot here from Guyer's "The Possibility of the Categorical Imperative." Guyer argues that each formula plays a distinct but necessary role in Kant's elaboration of the conditions that make a categorical imperative possible.

¹⁰ These examples are also discussed, in relation to Kant's second formulation of the categorical imperative, at G 4:429 (suicide), G 4:429–30 (false promises), G 4:430 (failure to develop talents), and G 4:430 (refusal of aid).

¹¹ The actual duties would be: do not commit suicide when in despair, do not make lying promises, do develop your talents, do offer aid to others.

¹² The best discussion I know is by John Rawls, "Themes in Kant's Moral Philosophy," in *Kant's Transcendental Deductions*, ed. Eckart Förster (Stanford: Stanford University Press, 1989), 81–113. Useful discussion is also to be found in Barbara Herman, "Murder and Mayhem," in her *The Practice of Moral Judgment* (Cambridge, Mass.: Harvard University Press, 1993), 113–31.

¹³ In thinking about the notion of self-contradiction employed by Kant, I have benefited from Christine Korsgaard's "Kant's Formula of Universal Law," *Pacific Philosophical Quarterly* 66 (1985), 24–47 (reprinted in her *Creating the Kingdom of Ends*, 77–105), and from Herman, "Murder and Mayhem."

¹⁴ Or, as Kant puts it, whether you could consistently will that it should "become by your will a universal law of nature" (G 4:421).

Kant tells us that a universalized maxim can be inconsistent with, or contradict, the will in two ways (G 4:424). First, conceiving of a given maxim as a universal law may be somehow self-contradictory. This is the so-called ‘contradiction in conception’ case, and can be illustrated by the ‘lying promise’ example. If we all made lying promises whenever a need arose, Kant argues, the institution of promising would disappear, and hence the maxim, universalized, is not consistent with itself (G 4:422): an order of things in which the maxim were universal law would undermine, rendering nonsensical, any strategy involving trust and promises. Second, universalizing a maxim might result in a state of affairs that, while consistent with action on the maxim itself, nonetheless could not be consistently willed by a willing human agent (G 4:423). This is the ‘contradiction in willing’ case, and can be illustrated by the ‘refusal of aid’ example. Finite creatures, such as ourselves, often want and need the “love and sympathy” of others (G 4:423). To will an order of things in which a maxim of refusing aid were universalized is, Kant argues, to will an order of things in which routine human needs and desires were routinely thwarted. It is, hence, to undermine the interests of the will itself: “a will that decided this would conflict with itself” (G 4:423).

It is important to notice that the contradictions or inconsistencies against which the universalizability test warns are neither logical contradictions nor contradictions with the natural order, but contradictions or inconsistencies with the will’s own basic interests. One can easily go wrong if one misses this. First, it is easy to be tempted, by talk of ‘strict method’ and by the allure of finding a moral algorithm, to think Kant has logical contradiction in mind. The ‘contradiction in conception’ leg of the universalizability test makes it especially tempting to read Kant as prohibiting logical inconsistency. But this is the wrong way to think of the test, and of the contradictions we are after, as attested to by the ease with which one can invent morally acceptable maxims that nonetheless fail the universalizability test, logically construed. One well-known example will make the point: a maxim of giving money to the poor (which we imagine should *pass* the test) seems to fail, since if everyone did this it would (on a plausible set of assumptions¹⁵) eliminate poverty. But then the maxim undermines itself: I cannot give to poor who don’t exist.¹⁶ We might play around with the maxim’s formulation, and, indeed, rules for maxim formulation designed to assist us

¹⁵ That there is enough wealth, all told, to go around; that the giving in question, if everyone did it, is enough to make a difference; etc.

¹⁶ For a discussion of a version of this case, see Barbara Herman, “Mutual Aid and Respect for Persons,” in her *The Practice of Moral Judgment*, 45–72.

in avoiding unpleasant test results abound in the literature.¹⁷ Such reformulations, however, tend to feel ad hoc, and to require frequent recourse to some interpretation of the general aim of the moral law – but once such recourse is required, dreams of a strict logical algorithm must already be abandoned.

Kant sometimes describes the consistency in question as a matter of whether a maxim could be a ‘law of nature’ (G 4:421, 424). This is also open to misinterpretation. Kant can be read here as suggesting that we must guard against conflict with the *natural order*. On this interpretation, the problem with a bad maxim is that, universalized, nature could not accommodate it. This interpretation is also suggested by Kant’s claim that a maxim of suicide in the face of despair is prohibited because:

a nature whose law it would be to destroy life itself by means of the same feeling whose destination is to impel toward the furtherance of life would contradict itself and would therefore not subsist as nature; thus that maxim could not possibly be a law of nature and, accordingly, altogether opposes the supreme principle of all duty. (G 4:422)

It looks, in other words, like suicide is prohibited because the natural world, under a universalized suicide maxim, would somehow convulse (‘could not

¹⁷ I noted many examples in [Chapter 3 \(note 9\)](#), and rehearse the basics again here. For discussion of the morally acceptable maxim, ‘I will play tennis Sundays at 10 a.m.’ (an example Herman borrows from Scanlon), which does *not* seem to pass the test, see Barbara Herman, “Moral Deliberation and the Derivation of Duties,” in her *The Practice of Moral Judgment*, 138; Tim Scanlon, “Kant’s *Groundwork*: From Freedom to Moral Community,” manuscript, cited in Herman, “Moral Deliberation and the Derivation of Duties.” Commentators who have hoped that clarification about what should count as a maxim will head off such untoward universalizability test results include Onora O’Neill, who proposes that we think of maxims as fairly general ‘orchestrating’ principles (Onora O’Neill, “Kant after Virtue,” *Inquiry* 26 [1983], 387–405 and “Consistency in Action,” in *Universality and Morality: Essays on Ethical Universalizability*, ed. by N. Potter and M. Timmons [Dordrecht: D. Reidel Publishing, 1985], 159–86). Otfried Höffe (“Kants kategorischer Imperativ als Kriterium des Sittlichen,” in *Ethik und Politik*, ed. Otfried Höffe [Frankfurt: Suhrkamp Verlag, 1979], 84–119) and Rüdiger Bittner (“Maximen,” *Akten des IV. Internationalen Kant-Kongresses, Mainz, 1974*, ed. G. Funke and J. Kopper [Berlin: Walter de Gruyter, 1974], 485–98) would have us think of maxims as general long-standing ‘life rules.’ Barbara Herman is willing to let levels of specificity vary, but proposes that only ‘generic maxims,’ which justify a *type* of action for a *type* of reason, be thought of as subject to moral review by the universalizability test (“Moral Deliberation and the Derivation of Duties,” 147–8). As I noted earlier, and as will be shown here, it is my view that the real answer lies not in finding a way to specify what a maxim is, such that only bad ones will fail and good ones pass the universalizability test, but in rethinking the meaning and use of the universalizability test itself. Whether or not a given maxim, or subjective practical principle, is universalizable and so morally acceptable, will never be a matter of mechanical application of a test – it will always be a matter that requires subtle investigation into underlying grounds for action, a general sense of what morality demands, of why universalizability matters to it, and of what therefore needs investigation, ‘universalizability-wise.’ The upshot? Maxims may take all shapes and sizes. I point those who want to think more about maxims again to Herman’s extremely useful discussion in her “Leaving Deontology Behind,” in her *The Practice of Moral Judgment*, 217–24.

subsist'). Kant seems to be saying that the basic structures of empirical reality, the laws governing physical objects in space and time, would somehow be violated if a maxim allowing suicide – and like it other immoral maxims – were universalized.

This, however, cannot be what Kant meant. The universalizability test cannot be a test of whether universal willing on a given maxim could be sustained by the natural world; it cannot be a test of potential compatibility with empirical reality. A moment's reflection on the universalized suicide maxim makes this clear.¹⁸ Kant is either offering a very *bad* argument (for surely nature would manage even we all killed ourselves when faced with despair), or else Kant's argument in fact does *not* ultimately turn on a universalized maxim's consistency with empirical reality. It is much more productive to understand Kant as appealing here to, as he writes, "*nature* in the most general sense (as regards its form) – that is, the existence of things insofar as it is determined in accordance with universal laws" (G 4:421; Kant's italics). The complete passage makes clear that nature in this 'general sense' contrasts with nature as empirical reality. It is more productive, in other words, to understand Kant as appealing to an idea of things *hanging together* and being as they are *qua* parts of a systematic law-governed whole. In *this* sense, there is something inconsistent about setting oneself the end of doing away with one's own capacity to set ends, or about a life-force that self-undermines – not because these go against nature as the empirically known, organic, physical world, but because they go against systematicity. A system – 'a nature in the most general sense' – that incorporated the end of ending end-setting, or a self-destructive life-force, would be disheveled at its core.

Thus Kant's claim that a world with a universalized suicide maxim could "not subsist as nature" (G 4:422) cannot be read as a claim about compatibility with empirical, physical, nature. Indeed, Kant explicitly dismisses inconsistency with *this* nature as the real issue when he writes, about a maxim of neglecting talents, that "a nature could indeed always subsist with such a universal law" (G 4:423), and about a maxim of refusing aid that "it is possible that a universal law of nature could very well subsist in accordance with such a maxim" (G 4:423). The notion of consistency at the heart of the universalizability test must lie elsewhere.

So where does it lie? The suggestion with which we began was that Kant's universalizability requirement makes most sense as a test of whether

¹⁸ A nice discussion of some of the puzzling things Kant writes about suicide can be found in Thomas E. Hill, Jr., "Humanity as an End in Itself," *Ethics* 91:1 (Oct. 1980), 84–99. Reprinted in his *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca: Cornell University Press, 1992), 38–57.

maxims, universalized, thwart the *will's interests*, either specifically (the maxim undermines *itself*) or generally (the universalized maxim is inconsistent with the will's projects generally). Bearing this in mind helps with potentially troubling cases. Take our earlier example: in giving money to the poor, surely the *interest* is in alleviating or eliminating poverty. A maxim that expresses this interest can hardly be thought of as 'contradicted' if, universalized, it would lead to poverty's elimination. The maxim, if ultimate interests are the focus, passes. By the same token, the real problem with a maxim of ending my life when I am in despair is that suicide actively forecloses the possibility of having or acting on any further interests whatsoever. Suicide destroys life, which is necessary (at least for us) for agency, and therefore conflicts with the will's inherent interest in itself. Committing suicide, like selling oneself into slavery, is a kind of self-contradictory act not because it cannot be done (both have been done often enough), but because it evinces a will to betray will.¹⁹ As Kant writes, "we must *be able to will* that a maxim of our action become a universal law: this is the canon of moral appraisal of action in general" (G 4:424; Kant's emphasis). The contradictions, 'in conception' and 'in willing,' are contradictions within the system of interests and principles that constitute the will itself.

Now to say this is to say something very important, and is to open a set of very important questions. Kant's 'being able to will' involves having it make sense for a will – constituted by interests and principles – to will the universalized maxim. But what does this amount to? At G 4:424, Kant emphasizes that the contradiction to be avoided is one that arises precisely for the *rational will*: the problem with a bad maxim is that "if we weighed all cases *from one and the same point of view, namely that of reason*, we would find a contradiction in our own will" (G 4:424; my emphasis). But then the categorical imperative does not demand that I ask whether a world in which my maxim were universalized is one I personally, and given the inclinations served by the maxim, would like; universalization does not just test whether my local interests are served. Indeed, were this Kant's request, three of Kant's four examples would fail to illustrate his claim: my suicidal aims are perfectly

¹⁹ See this passage from the *Lectures on Ethics* (VE 27:346):

[Freedom] has to be restricted, not, though, by other properties and faculties, but by itself. Its supreme rule is: In all self-regarding actions, so to behave that any use of powers is compatible with the greatest use of them. For example, if I have drunk too much today, I am incapable of making use of my freedom and my powers; of if I do away with myself, I likewise deprive myself of the ability to use them. So this conflicts with the greatest use of freedom, that it abolishes itself, and all use of it, as the highest *principium* of life. Only under certain conditions can freedom be consistent with itself; otherwise it comes into collision with itself.

well served in a suicidal society; idleness is (lucky me!) par for the course in the 'South Sea Islands' of Kant's imagination; and my disinclination to help is certainly compatible with, and even likely to arise from admiration for, a society of self-sufficient, rugged individualists. Kant's categorical imperative demands consistency with interests that must be more or other than interests in the success of inclination-based projects. If the decision procedure offered by the universal law formulation is to yield the results Kant thinks it must, the relevant interests must be in play *from the point of view of reason*.²⁰

But what are the interests of will from the point of view of reason? They can only be the interests will has *qua* free rational will. Free rational will is will apart from any idiosyncratic sensuous desires – it is the will we all have in common. What interests does it have? With what ought (therefore) universalized maxims to be consistent? My suggestion is that, for Kant, free rational will is supremely interested in its own activity, that is, in the activities of setting its own ends, of devising and adopting rules to govern its own practice, and of being efficacious in the world. The demand to will on maxims the universalizations of which we can consistently will is the demand to will on maxims the universalizations of which are consistent with continued and, where possible, enhanced free rational willing.²¹

What supports this suggestion? Kant's illustrations of the universal law formulation bear it out. The willings prohibited by the demand for universalizable maxims all show disregard and disrespect for constituents of free rational willing itself, or for conditions that, in us, preserve and promote it. Suicide deprives us of the very basis of free rational willing, namely life; physical self-diminution or self-mutilation (which Kant discusses in connection with suicide at G 4:429) eat away at this basis. A lying promise undermines free rational willing by misleading people about how the world will be, skewing the information they need to make

²⁰ This is one way that Kant's actual view differs from John Rawls' 'Kantian' view. Rawls' original position is meant to let (a-rational) self-interest do its work, just under a veil of ignorance that will ensure a just result. Kant's view asks deliberators to abstract themselves from self-interest altogether. They perhaps come to the same results – we can read Rawls' original position as simply an effort to help us imagine what such abstracting would look like – but it seems to me that Rawls does have a harder time knowing what to say to recalcitrant deliberators, that is, to those who insist they 'don't mind' things like extreme environmental degradation, or neighbors who won't speak to them, or censorship, or other things we would hope they would reject. See John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), 150ff.

²¹ A thought along these lines is also to be found in Christine Korsgaard, "Kant's Formula of Universal Law," *Pacific Philosophical Quarterly* 66 (1985), 24–47 (reprinted in her *Creating the Kingdom of Ends*, 77–105). Korsgaard there shows how the demand to treat humanity as an end helps explain the contradiction tests Kant has introduced. She concludes that consistency with the conditions for agency is required.

informed judgments about practical matters. Neglecting talents disregards our own capacities to pursue varied, complicated projects, and to live lives that pursue things other than the immediate satisfaction of inclinations; it shrinks the arena in which we can choose, about which we can reason, and in which we can be efficacious. Finally, a refusal to aid others disrespects the fact that human beings routinely need, as Kant put it, “love and sympathy from others,” in order to carry out our projects (G 4:423); most of us also need love and sympathy, at least from a few people, in order to be able to think at all rationally, to understand our choices, and to venture to act at all.

One can imagine the following skeptical question being asked: don't these demands all rely on facts about our sensuous desires, our actual projects, and our empirical resources? The answer is yes, but not in a way that is problematic. It is true that an accurate and valid weighing up of what will maintain and, even better, enhance free rational agency depends on judicious use of what Kant would call ‘anthropological data.’ Our wills are, after all, human wills, housed in bodies, subject to and in the business of negotiating amongst inclinations, and in command of limited, empirically articulated resources. In making the judgment that, for instance, a world without promising is not one I can will, I draw on knowledge about human communication and about the roles promises play in structuring social reality (that we are not transparent to each other, that professed intentions normally structure action in various ways, etc.). But while the universalization test must always *rely* on local anthropological data – it must rule on maxims meant to operate on and for people in the empirical world – it is not ultimately animated or informed by a concern with ‘empirical success’ (the success of empirical projects in the empirical world).²² It is instead animated by a concern to sustain and promote free rational agency, whatever this means for whoever is asking.

²² For Kant, this is a good thing: a decision procedure whose value rested on accurately assessing which empirical conditions would be most likely to advance some ‘average’ human set of projects would be a decision procedure on very shaky ground. The information needed to make such an assessment simply does not exist; the range and instability of human projects (conceptions of happiness) and the substantive disagreements between people about the conditions required to advance them ensure this fact. John Rawls’s conception of the “original position,” from which decisions about principles are to be made, errs, at least from a Kantian point of view, insofar as it assumes and requires recourse to a non-controversial body of empirical data about the laws of “human society” (“[Parties in the original position] understand political affairs and the principles of economic theory; they know the basis of social organization and the laws of human psychology” [Rawls, *Theory of Justice*, 137]). But as philosophers of social science have been telling us for some time, value-neutral understanding and knowledge are not available when it comes to politics, economics, sociology, and psychology. More gravely (again from a Kantian point of view), agents in Rawls’ original position are asked to assess proposed principles from the point of ‘veiled’ but otherwise ordinary self-interest, and not from a position principally informed by a higher-order value.

There is another way that the demand to will only on universalizable maxims can be seen to express an interest in free rational willing. Willing on universalizable maxims does not only sustain and promote free rational willing – it also *instantiates* it. To subject our wills to an overarching principle that is *not* grounded in inclination is to will in a way at once free and rational, ‘free’ because free of external, here sensuous, determination, and ‘rational’ because determined by reason, apart from inclination. Insofar as I adopt the universal law formulation of the categorical imperative as my practical law, insofar as I have it inform my maxims, I become a free rational will. This is more than just a piece of good luck for Kant; it is part of the ambition of his view that our moral theory will be ‘self-serving’ in the special sense that it will serve the best in all of our selves, in no small part by making us the best selves – the freest, most rational, most efficacious – selves we can be.

To sum up: a close look at the notion of consistency at the heart of Kant’s applications of the universal law formulation reveals that the test insists on consistency with free rational willing itself. The demand to act only on maxims that can be willed as universal law thus is a demand to honor the good of free rational willing. This is what the ‘formal’ categorical imperative tells us to care about; this is what it assumes we have an interest in. Let us turn now to Kant’s second *Groundwork* formulation of the categorical imperative, and see what it adds to the case.

The good represented in the formula of humanity as an end in itself

In offering this second formulation of the categorical imperative, Kant explicitly raises and addresses the question of our interest in the moral law. He reiterates the idea that material – or, we can now say, empirical – ends of action will not be adequate for our purposes, as they can only ground hypothetical imperatives (G 4:428). Kant then writes:

But suppose there were something *the existence of which in itself* has an absolute worth, something which as *an end in itself* could be a ground of determinate laws; then in it, and in it alone, would lie the ground of a possible categorical imperative, that is, of a practical law. (G 4:428; Kant’s emphasis)

Suppose, in other words, that there were something good in itself, something the existence of which all must take an interest in, and for its own sake. Something like this, and only something like this, could serve as an end or ground for a categorical imperative. Kant continues:

Now I say that the human being and in general every rational being *exists* as an end in itself, *not merely as a means* to be used by this or that will at its discretion; instead

he must in all his actions, whether directed to himself or also to other rational beings, always be regarded *at the same time as an end*. (G 4:428; Kant's emphasis)

On the basis of this assertion, that “the human being and in general every rational being *exists* as an end in itself,” Kant offers the second formulation of the categorical imperative: “So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means” (G 4:429). Here we see Kant clearly formulating the categorical imperative as a demand to serve a certain end, namely the ‘end in itself’ that is ‘rational being, or ‘humanity.’ What is this end, and what exactly are we to do to serve it? What, in other words, is ‘humanity’ or ‘rational being,’ and what is involved is regarding it always as an end? The general direction my answer will take is no secret: humanity or rational nature, is, I will argue, equivalent to our capacities for free rational willing, and to treat it always as an end is to think and act in ways that sustain and promote it, in ourselves and in others. The evidence for this, and details that give it some more specificity, will be found by looking again at Kant’s illustrations. Kant returns to the same four cases, and finds that this formula too prohibits suicide, lying promises, neglect of talents, and global refusal of aid. Suicide and neglect of talents fail to respect humanity in ourselves; lying promises and refusal of aid fail to respect it in others. How come, and what do the answers show?

A person may not commit suicide, if guided by the categorical imperative, Kant tells us, because he sees that, “if he destroys himself in order to escape from a trying condition he makes use of a person *merely as a means* to maintain a tolerable condition up to the end of life” (G 4:429). But this is very strange: in killing oneself one ‘makes use of oneself as a means’ only briefly, if at all, and how suicide ‘maintains a tolerable state of affairs’ is pretty obscure. To make sense of the thought we have to notice what Kant takes himself to have shown: “I cannot, therefore, dispose of a human being in my own person by maiming, damaging or killing” myself (G 4:429). Ah: the prohibition is on suicide as a logical extension of any physical self-diminution – “maiming, damaging, or killing” – which, reading back, might be undertaken as a ‘means to maintaining a tolerable condition,’ that is, undertaken for the sake of sensuous relief or escape. Alcoholism, self-mutilation, self-starvation, or other self-destructive ways of soothing the self must be at issue here. Neglecting talents, like committing suicide, also violates a duty to humanity in oneself, Kant tell us. This is so because, “there are in humanity predispositions to greater perfection” and “to neglect these might admittedly be consistent with the *preservation* of humanity as an

end in itself but not with the *furtherance* of this end” (G 4:430). We are therefore obligated to nurture our talents, pushing ourselves to fully develop our gifts. These two illustrations together forbid self-abuse and self-neglect, and require self-preservation and development.

How do they support the claim that free rational willing is really at stake? That, in other words, ‘humanity’ and ‘rational being’ as Kant understands these terms refer to our capacities for free rational willing? First, physically harming myself and neglecting talents both impair Kantian freedom. Freedom of any sort is, of course, wiped out, along with everything else, if I kill myself; this precisely makes the specific harm of suicide difficult to gauge. The other cases are more telling for this reason. Maiming or damaging myself, for the sake of some sort of sensuous relief, seems compulsive; just as we think addicts fail at *full* self-determination, so should we, I think, think of the cases Kant suggests here. Of course, as we saw in [Chapter 4](#), I do *freely choose* self-abuse – freedom of choice, and attendant responsibility, is inevitable for creatures like us – but to ‘give in’ to impulses to sooth myself in ways that ultimately harm me is to use freedom in a way that betrays the hope of achieving a fuller, more expansive and sustained kind of freedom (the ‘achievable’ freedom of [Chapter 4](#)). Failing to develop talents similarly renders me less free in the ‘achievable’ sense. For one, it makes me more dependent on luck and circumstance, and on the kindness of strangers (who may seek to determine my will); I lose a measure of ‘self-determinability.’ For another, the fewer talents I have cultivated, the poorer is the ‘self’ and the more constricted its sphere of self-determination: I am less able to respond to situations flexibly; I accordingly face fewer viable options for action; I am less likely to be able to weigh and choose meaningfully. In neglecting talents, I may also fail to learn virtues like self-control and discipline, which assist in subordinating fleeting urges, which in turn would have helped sustain self-determination. At least for these reasons, keeping myself alive and healthy and developing my talents and abilities enhances the level of ‘achievable’ freedom we achieve. Harming and neglecting myself don’t destroy freedom – nothing (save death) could – but they do dishonor and undermine it.

Self-destructive acts and talent neglect also denigrate Kantian rationality. Damaging or maiming oneself, in the ways Kant has in mind, presumably inflicts harm that make it difficult or impossible to satisfy a range of other desires – holding a desired job, being able to care adequately for loved ones, successful pursuit of other ambitions – and for this reason profoundly disrespects reason’s instrumental or prudential demands. I am not supposed to make it harder for myself to get things I want. Neglecting talents similarly

disrespects the dictates of instrumental rationality: I am supposed to make it easier for myself to get the things I want. Moreover, both self-abuse and self-neglect also ignore and so denigrate reason's capacity to override inclinations in general, since self-abuse and self-neglect are (at least for Kant) things we would only pursue out of inclination. Reason is dismissed and disrespected. Compare this with the special cases Kant mentions, viz., "having limbs amputated in order to preserve myself, or putting my life in danger in order to preserve my life, and so forth" (G 4:429). These cases do not constitute self-abuse because they in fact rationally subordinate impulses for a greater good.

Finally, in both self-abuse and self-neglect cases, the sheer power of will – our power to intentionally change the world, in all sorts of ways – is truncated. Maiming and damaging my body reduces its efficacy, adaptability, and strength; failing to develop talents leaves me with fewer skills at my disposal. The 'humanity' or 'rational being' Kant demands we treat always as an end in itself is thus very plausibly glossed as our capacity for free rational willing itself, which is not just the prerequisite, but also the end of morality insofar as it is ultimately also always the capacity for freer, more rational, more efficacious action in the world.

Notice that bearing this in mind in fact helps interpret demands like "don't harm yourself," or "don't neglect your talents," or even "treat yourself always as an end." By themselves, these demands seem open to arbitrary and idiosyncratic interpretation: some see tattoos and scarification as self-abuse, others as ennobling adornment; some think studying languages is the most important part of a general education, others insist it is studying art. On my reading, we at least have rich Kantian senses of realized freedom and rationality, as desiderata, to help us figure out when, for instance, plastic surgery is self-abuse and when not. We may still not be able to settle all these questions *easily* – they require much searching argument and discussion – and indeed we may not be able to settle them at all – general answers may not be forthcoming, because so many individual differences *are* relevant – but a focus on preserving and promoting free rational will can help inform any practical decisions we try to make in non-arbitrary, non-idiosyncratic ways.

The prohibition on lying promises – the first of the illustrations demanding respect for *others'* humanity – lends further support to the idea that Kant's ultimate interest is in the activity of free rational will, in protecting and promoting the capacity for freer, more rational, more efficacious action in the world. Kant tells us that lying promises are ruled out by the second formulation of the categorical imperative because they engage others in

actions the ends of which they would not endorse if they were privy to them: for an action to avoid treating others merely as means, Kant writes, those others “must also be able to contain in themselves the end of the very same action” (G 4:430).²³ They must, in other words, be able to understand and endorse the enterprise into which they are being drawn. Why? The proper use of our capacities is at stake, and the ruling is: our (human) capacities (of all sorts) are used properly when they are used in the service of ends each of us, individually, freely, rationally could choose. But when I respect the demand to avoid using you (your money, your trust, your efforts) in ways you could not thus choose, I pay homage to the idea that you should act – employ your capacities – only toward ends you yourself could and would freely and rationally will. But if treating humanity as an end in itself means this, means treating others always in ways they could freely, rationally will, then it actively honors the activity of free rational willing.

The duty to aid others goes the next step. It is not just a duty to avoid trampling, so to speak, on the free rational willings of others, but is a duty to positively support them. Everyone must try, Kant writes, “as far as he can, to further the ends of others” (G 4:430). He continues:

For, the ends of a subject who is an end in itself must as far as possible be also *my* ends, if that representation is to have its *full* effect on me. (G 4:430; Kant’s emphasis)

Why must I promote others’ ends if the idea of a subject who is an end in itself is to have its “full effect” on me? If, as I have been arguing, the categorical imperative aims to preserve and promote free rational willing, and if the capacity for free rational willing is what makes humanity an end in itself, we might expect a demand to promote others’ freedom, or rationality, or powers. But this isn’t the demand – the demand is to help them attain their ends. It helps here to notice arguments Kant makes in his later (1797) *Metaphysics of Morals*, where he elaborates on the duty to aid others. There, Kant argues that while I have a duty *to myself* to pursue moral perfection (to discipline and cultivate my own powers to set ends in accordance with morality, i.e., to protect and enhance my own capacities for free rational willing), it would be “self-contradictory to require that I do (make it my

²³ The questions of how to decide what an agent would endorse, of how to avoid offering and/or applying criteria of ‘endorse-ability’ in a way that in fact violates individuals’ self-sovereignty, are not, I would argue, real questions for Kant. This is because the question of whether an end is endorsable is equivalent to the question of whether the maxim directed at it is universalizable – we do not need to see into the souls of others to decide whether they could share such-and-such an end. (One might also say: we look precisely as far as we need to into the souls of others whenever we determine whether a maxim is universalizable, that is, whether it is consistent with free rational agency in general.)

duty to do) something that only the other himself can do” (MS 6:386). “For the *perfection* of another man, as a person, consists just in this: that he *himself* is able to set his end in accordance with his own concepts of duty” (MS 6:386; Kant’s emphasis). The idea is this: I cannot be asked to act in ways that would preserve and enhance another’s free rational willing per se. Rather, I can be, and here am, asked to respect this willing, first, by not undermining it (e.g., by making false promises), but also, here, by smoothing, via aid, the way for its expression. Practical reason is a capacity for setting ends. It is a desiring capacity – it wants things, it wills them. If the value of free rational willing is to have its full effect on me, I must feel protective of its efforts.²⁴ I am thus bound, in a general way, to help others attain the ends they set for themselves.²⁵

In addition, Kant’s discussions of the prohibition on lying promises and the demand to aid others both suggest that the demand to treat humanity as an end expresses respect for the human ability to subordinate inclinations to a purely rational principle. The possibility of actions with ‘sharable’ ends, if such ‘sharability’ is not to be accidental but instead reliable and universal, is equivalent to the possibility of a harmonization of purposes that can only be achieved by human beings insofar as we subordinate our ends to reason itself. I can only assume the ‘sharability’ of my own ends, and I can only myself adopt the ends of others, insofar as I can assume ends set under guidance of a purely rational (and hence shared) principle.²⁶ My and your pursuit of ends leads to conflict whenever our wills are guided, in contrast, by a principle of self-love; or, if harmony is somehow achieved under guidance of self-love, it is achieved only fortuitously. As Kant puts it, the only ‘harmony’ an action-guiding principle of self-love will reliably give rise to is

like that which a certain satirical poem depicts in the unanimity between a married couple bent on going to ruin: “*O marvellous harmony, what he wants she wants too*” and so forth, or like what is said of the pledge of King Francis I to the Emperor Charles V: “What my brother Charles would have (Milan), that I would also have.” (KpV 5:28)²⁷

²⁴ Of course, I am only obligated to help others attain morally acceptable ends (which will be the same as ends set freely and rationally). Kant makes this explicit in the *Metaphysics of Morals*, at 6:450; we can assume he takes this for granted in the *Groundwork* discussion.

²⁵ The duty here is ‘imperfect,’ meaning that I am *not* obligated to help everyone attain all their ends all the time; rather, I am obligated to help others *at my own discretion*. In contrast, the prohibition on lying promises is ‘perfect’: I *am* obligated to avoid making false promises to everyone all the time.

²⁶ The universal law formulation of the categorical imperative is just such a principle; we see here how the formulae work together, or better, are different ways of demanding the same things.

²⁷ The ‘marvellous harmony’ verse, by Johann Nikolaus Goetz, was set to music by Franz Joseph Haydn (“Die Harmonie in der Ehe”). Thanks to Katrin Wolffsohn for drawing this to my attention.

The point here is just that the categorical imperative, insofar as it demands we set ends that others can endorse, demands that we eschew ultimate guidance by an inclination-based principle, and, quite importantly, simultaneously imputes to and vows respect for the ability others have to eschew such guidance – for without this ability there is no sense in which we could hope to have truly shared ends. But then, insistence on actions with sharable ends is equivalent to insistence on actions that both express and respect free rational willing, wherever it is found.

To sum up, we have seen here not only that Kant's second formulation of the categorical imperative expressly commands us to adopt an end, but that that end, named 'humanity' or 'rational nature' by Kant, amounts to our capacity for free rational willing. Free rational willing is the good represented in the formula of humanity as an end in itself.

The good represented in the formula of autonomy

Kant offers a third formulation of the moral principle expressed by the categorical imperative. Kant twice gives the third formulation as simply "the idea of the will of every rational being as a will giving universal law" (G 4:431 and 432); he later characterizes it as "the principle of every human will as *a will giving universal law through all its maxims*" (G 4:432; Kant's emphasis).²⁸ The formula, then, expresses a kind of regulative ideal, asking us to imagine all wills as law-making. Because it emphasizes the activity of making laws for oneself, rather than mere subjection to the law, Kant calls this formula "the principle of autonomy" (*auto*: self; *nomos*: law) (G 4:433; see also G 4:439), and much of the literature refers to it as the 'formula of autonomy.' As Kant writes, his third formula describes a will,

not merely subject to the law but subject to it in such a way that it must be viewed as also giving the law to itself and just because of this as first subject to the law (of which it can regard itself as the author). (G 4:431)

The third formula describes a will that binds itself to its own law. Now because the formula also emphasizes the thought of *all* wills – "the will of every rational being" – and thinks of them as co-legislating, Kant tells us that this formula gives rise to the idea of a "kingdom of ends," or a "systematic union of various rational beings through common laws" (G 4:433). The idea here is of a set of wills, harmonized with each other

²⁸ Above, we put the formulation thus: 'Act in accordance with "the idea of the will of every rational being as a will giving universal law"' (G 4:431; Kant's italics).

because all will (on) the same common overarching laws. Because of its close connection with this idea, the formula of autonomy is also sometimes referred to as the formula of the kingdom of ends.

Kant comes to this third formulation of the categorical imperative following his illustrations of duties derived from the formula of humanity as an end in itself. He is engaged in an effort to explicate and defend the claim that humanity as an end in itself is not an end “borrowed from experience,” but is an end that “ought as law to constitute the supreme limiting condition of all subjective ends,” such that its principle must “arise from pure reason” (G 4:431). Kant writes:

That is to say, the ground of all practical law giving lies (in accordance with the first principle) *objectively in the rule* and the form of universality which makes it fit to be a law (possibly a law of nature); *subjectively*, however, it lies in the *end*; but the subject of all ends is every rational being as an end in itself (in accordance with the second principle); from this there follows now the third practical principle of the will, as supreme condition of its harmony with universal practical reason, the idea *of the will of every rational being as a will giving universal law*. (G 4:431; Kant's italics)

This very dense and rich Kantian sentence claims, first, that the ground or ultimate reason for choosing to act in accordance with universalizable maxims “lies” in – is contained or represented in – the universalizable maxim (“the rule”) itself, in virtue of the maxim’s taking the form it does: a maxim’s being universalizable is a *reason* it is acceptable. This is what it means to say the ground lies objectively in the rule. The claim is, secondly, that this ground – the reason for choosing to act according to universalizable maxims – must also ‘lie in’ subjects: subjects must represent this reason for choosing to themselves. And subjects represent grounds to themselves by adverting to *ends*; “*subjectively*, however, [the ground] lies in the end.”²⁹ These points are meant to make clear that and how a principle based in the end of humanity, “arise[s] from pure reason.” This making clear needs to be done: the upshot of Kant’s argument thus far has been to exclude ends that spring from anywhere but reason, so we have reason to want assurance about ‘humanity’s’ pedigree as a purely rational end in itself. The only end that can serve to ground an enactment of universal law must be an end subjects represent to themselves (i.e., have) in virtue of reason alone. Kant’s

²⁹ This interpretation of what Kant means when he writes that the ground is ‘objective’ in the rule, and ‘subjective’ in the end makes the best sense of the surrounding discussion. The idea is that in the rule the ground is formally represented, available and applicable to all rational beings, independent of particular wills, and is in this sense objective, while subjectively, in terms and ‘inside’ of particular subjects’ wills, the ground must be represented as an end – because this just is how grounds are represented by subjects for themselves.

claim here is that the end subjects have in virtue of reason alone is just what the second formula said it was, namely ‘humanity’ or ‘rational being’ itself, that is, the capacity for freely setting ends and ordering actions according to rational principles – to be able to do this is, after all, precisely to be the “subject of all ends” (G 4:431). Hence, the ground for the enactment of universal law, the ground that lies ‘objectively in the rule’ and ‘subjectively in the end,’ is just the subject, as free rational will, itself. These reflections prompt Kant to introduce the formula of the autonomy itself: “from this there follows now the third practical principle of the will, as supreme condition of its harmony with universal practical reason, the idea of *the will of every rational being as a will giving universal law*” (G 4:431; Kant’s italics).

Now rational being is actively *being* rational being – is doing its valuable, end-in-itself thing – precisely insofar as it is engaged in the activities of formulating and adopting universalizable maxims. We saw this point earlier: the categorical imperative effectively demands not just respect for but also instantiation of free rational willing. The will that enacts law universally (“in accordance with the first principle”) thus both *respects* and *instantiates* those qualities that make humanity an end in itself (“in accordance with the second principle”). But then we see that the ground for enactments of universal law – the guiding idea or “supreme condition of the will’s conformity with universal practical reason” – can indeed be formulated in terms of “the idea of *the will of every rational being as a will giving universal law*” (G 4:431; Kant’s italics).

We can come at this thought another way if we turn again to Kant’s notion of autonomy. A will that makes its own law is autonomous, that is, is self-legislating. The will that authors its own law manifests independence from every external, conditioned interest. This at least means that insofar as we author a law, we are subject to it for no external reason. It seems also to mean that the (internal) reason we are subject to it is that it is our own. But a law’s being our own might just as well be counted as a reason that we are *not* subject to it. Indeed, many before Kant had argued that it is incoherent to speak of agents subjecting themselves to themselves, and Kant himself raises this objection, as an antinomy to be solved, in the *Metaphysics of Morals* (MS 6:417). (See also Kant’s discussion of the ‘paradox’ of self-legislation at G 4:439.)

Kant solves this problem in an incredibly ingenious way. He writes:

In accordance with this principle all maxims are repudiated that are inconsistent with the will’s own giving of universal law. Hence the will is not merely subject to the law but subject to it in such a way that it must be viewed as also giving the law to itself and just because of this as first subject to the law (of which it can regard itself as the author). (G 4:431)

When Kant writes that the will “viewed as also giving the law to itself” is “just because of this ... first subject to the law” the claim is this. In making the law, the will first enacts or instantiates its own value, and at the same time experiences itself as compelled to honor this value. The idea is that the activity of autonomy – free rational activity itself – is sufficient to command the will’s own interest in itself. The formula of autonomy describes the form of willing to which our own inherent interest in autonomy draws us.

This thought is lent support by Kant’s ensuing discussion of the progress he has made in introducing the third formula. The formula of autonomy is the first formulation of the moral law that somehow contains – that is, internally expresses – its own ‘categoricity’:

that there are practical proposition which command categorically could not itself be proved, any more than it could be proved either here or anywhere else in this section; one thing, however, could still have been done: namely, to indicate in the imperative itself the renunciation of all interest, in volition from duty, by means of some determination the imperative contains, as the specific mark distinguishing categorical from hypothetical imperatives; and this is done in the present third formula of the principle. (G 4:431 2)

That there really are effective categorical imperatives (“practical propositions which command categorically”) cannot yet be shown (this must wait for the *Groundwork’s* [third section](#), and Kant’s proof that we are really free to be determined by reason). But that a given imperative *is* categorical, if anything is, can be shown if we find an internal mark distinguishing it as categorical. I interpret Kant here as follows. An internal mark distinguishing an imperative as categorical would just be something in the imperative itself that excluded the possibility of following it for external reasons. Now, the formula of universal law does not, in itself, contain a determination that would exclude its being followed on external reasons. One might, as far as the formula itself goes, follow it on prudential ‘what goes around comes around’ grounds, or out of a belief that it expresses God’s will and will guarantee heaven: there may be lots of good (external) reasons to act only on universalizable maxims. The formula of the end in itself likewise does not indicate in itself why or how it qualifies to serve as a categorical imperative: “always treat humanity as an end in itself” is an imperative that might also be followed on external grounds.³⁰ Each imperative, Kant writes, was “only

³⁰ At least this is what we must imagine Kant meant, in writing that the first two formulae were only ‘assumed’ to be categorical (G 4:431). It seems, however, that if the imperative is taken seriously it excludes compliance on external grounds – to treat humanity as an end in itself would after all seem to demand treating it as an end for its own sake, and not for some other reason. We leave this matter unsettled here.

assumed to be categorical because we had to make such an assumption if we wanted to explain the concept of duty” (G 4:431; Kant’s emphasis).

But in the third formulation, “renunciation of all interest” is indicated “by means of some determination the imperative contains, as the specific mark distinguishing categorical from hypothetical imperatives” (G 4:431). Kant writes that, “among all possible imperatives,” the third formula “can alone be *unconditional*” (that is, can be seen *prima facie* to be unconditional) (G 4:432; Kant’s emphasis). We can think of it, Kant writes, this way:

if there is a categorical imperative (i.e., a law for every will of a rational being) it can only command that everything be done from *the maxim of one’s will as a will that could at the same time have as its object itself as giving universal law*; for only then is the practical principle, and the imperative that the will obeys, unconditional, since it can have no interest as its basis. (G 4:432; my italics)

The third formula of the imperative represents will that makes universal law. Making universal law is something that cannot be done on the basis of empirical or inclination-driven interest. We have already seen, in the discussion of the formula of universal law, that willing universal law precisely requires that we abstract from such interests and will only according to some internal interest of will. Kant makes the point here by noting that an externally motivated or ‘dependent’ will cannot, precisely insofar as it is dependent, enact universal law; its enactments must always answer ultimately to self-love, robbing them of true universality (G 4:432). Only an independent will can make true, and not provisional, conditioned, or otherwise faux, universal law. The imperative therefore represents will that opts to regulate its own choices and actions by taking on the mantle of law-giver, that is, by insisting on its own autonomy, with the result that “all maxims are repudiated that are inconsistent with the will’s own giving of universal law” (G 4:431). An imperative that asks us to “think of a will of this kind” (G 4:432), and to make the idea regulative, is an imperative (‘act in accordance with this idea’) that is, in virtue of itself (or ‘some determination the imperative contains’), unconditioned. Its guiding idea is that of a will that wills according to its own law, not on external grounds, that is, in other words, autonomous; and with this, voilà, we have an imperative that describes a form of willing that can only be undertaken for its own sake.

But then this is extraordinary. Kant has succeeded where all previous moral theories have failed: he has found a moral principle that is not driven by desire for some external end. Instead, his moral principle is grounded in a value inherent in and realized by willing on the principle itself – and this is so because it makes a certain kind of willing the end in itself. As Kant writes,

a categorical imperative “can only command that everything be done from the maxim of one’s will as a will that could at the same time have as its object itself as giving universal law” (G 4:432). We are now in a position to see how apt Hegel’s characterization of Kantian moral theory is: for Hegel, Kantian moral theory is based fundamentally on the idea of “the free will which wills the free will.”³¹ Or, to put the idea in other terms, Kantian morality is founded on the interest free rational willing takes in itself.

Kant’s ensuing discussion of dignity helps make explicit that autonomous activity, or free rational willing, is indeed ‘self-interested’ in just this sense. Dignity, Kant tells us, is the quality a thing has if it is “raised above all price and therefore admits of no equivalent” (G 4:434). Reason, Kant writes,

refers every maxim of the will as giving universal law to every other will [via the idea of a kingdom of ends] and also to every action toward oneself, and does so not for the sake of any other practical motive or any future advantage but from the idea of the *dignity* of a rational being, who obeys no law other than that which he himself at the same times gives. (G 4:434; Kant’s italics)

We author and follow the categorical imperative, “not for the sake of any other practical motive or any future advantage but from the idea of the *dignity* of a rational being” (G 4:434), which is to say, from the idea of ourselves as free rational wills. Free rational will finds itself supremely impressive, looking on its own activity with “immediate favor and delight,” as “the object of an immediate respect” (G 4:435); free rational will is interested, ultimately, in itself.

FREE WILL WILLING ITSELF

Here all the work we have done so far comes together. “For when we think of a will of this kind” (to borrow Kant’s words at G 4:432), when we conceive a will that makes its own law and is inherently interested in subjecting itself to this law, when we think of autonomy, and understand that autonomy, or free rational willing, describes the source and the aim, the origin and the end of moral practice, once we do this the analytic work is done. We have, that is, completed the ‘analysis’ that Kant tells us makes up the [second section](#) of the *Groundwork*: we have unpacked the concept of duty, and learned what its principle must be. The categorical imperative, in all its formulations, calls on us to be free rational wills in the service of free rational willing. We saw this in our discussions of the first two formulations:

³¹ Hegel, *Elements of the Philosophy of Right*, §27, p. 57.

they each amounted to commands to preserve and promote free rational willing. With the formula of autonomy, we make explicit what was implicit in those formulations, namely that respecting and instantiating free rational willing are inseparable activities, and that preserving and promoting free rational willing is something to which the free rational will by its nature is committed. The categorical imperative is categorical, and effective, because it commands us to will in a way that serves the will's own true and only inherent interest, the only interest that is its own and not furnished by inclination, viz., its interest in itself.

It is now a straightforward matter to see that Kant is right when he claims that "the above three ways of representing the principle of morality are at bottom only so many formulae of the very same law, and any one of them of itself unites the other two in it" (G 4:436). The formulae are different, he explains, "subjectively" rather than "objectively," meaning that while they express the same law, subjects experience them as guiding deliberation and judgment in different ways. The first formula addresses itself to the form maxims must take, the second to the matter maxims must respect, the third to the "complete determination" of maxims, offering as it does a formula that simultaneously refers to the form and the matter to which maxims must conform (G 4:436). But all, we now see, express the demand to will in ways that put the value of free rational willing above all other considerations.

This is very helpful. Taken separately, the formulations are incomplete, unsatisfying, and even misleading. But taken together, they provide a complete framework of moral judgment and action, giving us a rule of thumb for assessing courses of action (the universalizability requirement), an end on which to steady our sights (free rational willing itself), and a sense of ourselves as precisely both source and object of morality.³²

The work of Chapter 6 is now essentially complete. We have covered a lot of ground. We looked first at the claims that give rise to an interpretation of Kant's view as 'formal,' and saw that they amount really to a requirement that the categorical imperative represent a non-empirical good. In order to show how Kant understands this good, we looked at Kant's three *Groundwork* formulations of the categorical imperative. We saw that each demands willing that honors – by preserving and promoting, and by

³² The following passage from Kant's lectures on ethics also sums the whole story up nicely:

Anyone who allows his person to be governed by his inclinations is acting contrary to the essential end of mankind, for as a free agent he must not be subject to his inclinations, but should determine them through freedom; for if he is free, he must have a rule; and this rule is the essential end of mankind. (VE 27:345)

instantiating – free rational willing itself. The value of such activity is at work in Kant's arguments for, and illustrations of, each formulation. We saw that the formulae are, at bottom, the same just insofar as they all prescribe free rational willing that aims at free rational willing.

Kant's accounts of the categorical imperative in the *Critique of Practical Reason* and *Metaphysics of Morals* support the same conclusions. The *Critique* relies more exclusively than the *Groundwork* on a formulation demanding universalizable maxims; in the *Metaphysics of Morals*, Kant returns to draw heavily on the notion of humanity as an end in itself. But there is no reason to think that he changed his views in ways that affect the claims we have made so far.

Just after claiming that all three formulae of the categorical imperative are, at bottom, the same, Kant writes:

We can now end where we set out from at the beginning, namely with the concept of a will unconditionally good. *That will is absolutely good* which cannot be evil, hence whose maxim, if made a universal law, can never conflict with itself. (G 4:437)

Kant here himself affirms the argument I have been making. Recall the first sentence of the [first section](#) of Kant's *Groundwork*:

It is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a **good will**. (G 4:393; Kant's emphasis)

This opening sentence now takes on a much richer meaning. The idea of a will that makes universal law, treating rational being in itself and others always with respect – that is, the idea of a fully free, fully rational will – is the idea of a will that wills in a way that conforms to our understanding of a good will. Good willing consists just in willing that is consistent with (instantiates, honors, promotes) good (free rational) willing itself. The idea of the will of every rational being as a will which makes universal law is thus at once the idea of the subject (author, source, origin) and the object (end, aim) of the moral law, or of free rational practice. Free rational willing both generates the moral law, and fulfills its commands.³³ Kantian moral goodness, to recast Hegel's formulation, is realized in the good will that wills the good will.

³³ This reminds us for a final time of how the third formulation follows from, and/or contains the first and second formulations: in describing an autonomous will, it describes both the activity of making universal law (commanded by the first principle) and the feature of humanity that makes it an end in itself (honored by the second principle).

Everything does come together for Kant in free rational willing. In authoring and following the moral law, in (that is) preserving, promoting, and instantiating our capacities to be moral by authoring and following the rational law of freedom, we individually and collectively express our higher nature, manifest reason in the world, participate in and create a world of value, and achieve independence from the burdensome and confusing world of sense. I will try to show all this, and to show how all this makes 'a good will willing itself' much more and other than an empty formalism, in the [next chapter](#). I have succeeded here, in this chapter, if I have shown that Kant's categorical imperative commands us, in an ingenious and profound way, to freely and rationally will to will freely and rationally.

CHAPTER 7

What's so good about the good Kantian will? The appeals of the strange thing¹

INTRODUCTION

Why ask about the good of free rational willing?

We have seen arguments that, for Kant, free rational willing is the good at which practice ought ultimately to aim – that free rational willing is, for Kant, the moral good. We have seen, that is, that the value of free rational willing grounds, for Kant, the commands of morality, and that our interest in free rational willing is what moves us to follow these commands. Here, we need to ask another question, namely, what allows free rational willing to play all these roles? What, that is, is so appealing about it? Kant thinks an interest in free rational willing is strong enough to compete with sensuous desire in moving our wills. Why? How? Wherein lies the pull of free rational willing?

These questions will seem distinctly unKantian to some. For some, insistence that morality eschew incentives, interests, or motives is a hallmark of Kantian thought. However, as we have seen, Kant's real claim is that morality must eschew *empirical* and *external* incentives, interests, and motives. The answers I offer here to these questions will also seem distinctly unKantian to some. I address the question of what pulls us toward free rational will in terms of appealing experiences and lived self-awarenesses. But this can seem all wrong, since free rational willing is noumenal if anything is: how could it appeal, attract, be experienced, be something like a 'phenomenon' with pull, at all? The answer lies in the fact, suggested in [Chapter 1](#), that the free will's 'noumenality' does not, for Kant, mean we have no awareness of ourselves as free rational wills, or that such awareness

¹ Portions of this chapter were heard by very helpful audiences at Barnard College (Willen Faculty Seminar), the University of Miami, and the Florida Philosophical Association Annual Meeting. Special thanks are due Taylor Carman, Peter Lewis, and Corina Vaida for their questions and comments on these occasions.

has no psychological texture. Finally, the questions I am raising here may seem unnecessary. Aren't we done? We have come to the conclusion that free rational willing is the ultimate aim of Kantian moral practice. We have shown that free rational willing is the keystone in Kant's moral architecture. Free rational willing permits Kant to fit what he takes to be his readers' common-sense intuitions about morality together with the philosophical or metaphysical demands he thinks any coherent theory must accommodate. Free rational willing, first as sheer capacity, then as promising potential for greater, more rational human self-determination, emerges as the thing that, in a host of ways, makes the system work together. It fills a gap, it is necessary to the whole, and, as such, it seems gratuitous to ask for further recommendation – isn't it enough, as it were, that free rational willing perform this very important job of satisfying architectonic demands?

Notice that if one takes this view, a different set of concluding questions make sense. It makes sense to ask about the soundness of Kant's arguments for the necessary, universal, unconditional value of free rational will, arguments from the respective demands of moral practice and metaphysics. One might for instance ask: Is it really so that moral demands must be unconditional? Are respect and reverence really only accorded activities and capacities that transcend the natural world? Is there no philosophically respectable alternative to libertarian free will if we are to preserve moral responsibility? Must we (still) subject practical theories to the demands of a Newtonian conception of the natural world? And so on. These are all fine questions, and, indeed, many significant interpretations of and challenges to Kant's view come from theorists who take them up.² But I think we need, in

² One thinks immediately of interpretations offered by Henry Allison (*Kant's Transcendental Idealism: An Interpretation and Defense* [New Haven: Yale University Press, 1983, 2004]) and Kant's *Theory of Freedom* [Cambridge and New York: Cambridge University Press, 1990]; Paul Guyer (*Kant and the Experience of Freedom* [Cambridge and New York: Cambridge University Press, 1993], *Kant on Freedom, Law, and Happiness* [Cambridge and New York: Cambridge University Press, 2000], *Kant's System of Nature and Freedom: Selected Essays* [Oxford: Oxford University Press, 2005]); Christine Korsgaard (*Creating the Kingdom of Ends* [Cambridge and New York: Cambridge University Press, 1996]); Onora O'Neill (*Constructions of Reason: Explorations of Kant's Practical Philosophy* [Cambridge and New York: Cambridge University Press, 1989]); and many others. Influential critics have challenged the moral intuitions Kant assumes and endorses at least since Schiller (see Chapter 1, n10). Contemporary challenges have been mounted by commentators such as Bernard Williams (*Moral Luck: Philosophical Papers 1973–1980* [Cambridge and New York: Cambridge University Press, 1981]); and Annette Baier (*Moral Prejudices: Essays on Ethics* [Cambridge, Mass.: Harvard University Press, 1994]). Many feminist critics (in addition to Baier) attack Kantian intuitions and assumptions about moral practice (see, e.g., the essays collected in Virginia Held, *Justice and Care: Essential Readings in Feminist Ethics* [Boulder, Col.: Westview Press, 1995]), as do communitarians such as Michael Sandel (*Liberalism and the Limits of Justice* [Cambridge and New York: Cambridge University Press, 1982, 1998]). Critics who have challenged Kant's assessment of the metaphysical

concluding, to ask directly about free rational willing's appeal. What about free rational willing, as Kant understands it, qualifies it to ground an entire system of morality?

There are good methodological reasons for asking *this* question. For one, to treat free rational willing as something good merely insofar as it solves a formal problem, insofar as it fills a gap in an architectonic, without allowing or explaining that it exerts a powerful attractive force of its own, renders those eager and willing to act on behalf of Kantian free rational willing motivationally opaque (or subject to charges of caring too much about formalities and architectonics). As Bernard Williams put the charge, "there can be no reason [for Kant] for being moral, and morality itself presents itself as an unmediated demand, a categorical imperative."³ For this reason alone it will be worth trying to say why free rational willing is, as such, attractive.

Furthermore, the most damaging and enduring criticisms of Kant's moral theory have been aimed directly at Kant's conception of free rational willing and the moral ideal it entails.⁴ Many of these charges focus on Kantian *autonomy*, which we have claimed is the same as free rational willing.⁵ Kantian autonomy, goes one familiar charge, denigrates the value of fellow-feeling, sympathy, and other emotions as bases for action, esteeming only rational dictates and actions issuing from them.⁶ An ideal based on Kantian autonomy, others charge, separates us from and constitutes us

terrain to be negotiated are almost too numerous and varied to try to name. Hegel is certainly one of the most significant (see Sally Sedgwick, "Metaphysics and Morality in Kant and Hegel," *Bulletin of the Hegel Society of Great Britain* 37 [1998], reprinted in *The Reception of Kant's Critical Philosophy: Fichte, Schelling, and Hegel*, ed. S. Sedgwick [Cambridge and New York: Cambridge University Press, 2002], 306–23; and Michael Forster, *Hegel's Idea of a Phenomenology of Spirit* [Chicago: University of Chicago Press, 1998], esp. 33–78). Contemporary challenges have been mounted by compatibilists and others who challenge Kant's metaphysics generally. Of course, as is to be expected, objections to Kantian intuitions and to Kantian metaphysics often go hand in hand, since they are mutually supportive.

³ Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge, Mass.: Harvard University Press, 1985), 55.

⁴ Schiller's well-known dig at the Kantian host, who regrets his duty-confounding warm feelings toward his guests, begins the tradition (Friedrich Schiller, *Xenien* [1797], collected in Goethe, *Werke I*, ed. Erich Trunz [Hamburg: Christian Wegner Verlag, 1949], 221; for text of dig, see [Chapter 1, n10](#)). Contemporary critiques of Kant's autonomy ideal come from many quarters, but perhaps especially from feminists (see, e.g., works cited in [note 2](#) above) and communitarians (see, e.g., Sandel, *Liberalism*).

⁵ The autonomous Kantian will is both free and rational. As noted earlier, I prefer to use the expressions "free rational activity of the will" or "free rational willing," because, though less felicitous, they keep the three components (freedom, rationality, an active faculty of desire and action) distinct while emphasizing that in human beings they always, for Kant, come together.

⁶ This sort of charge is leveled, for instance, by Humeans and some feminists. See work by Annette Baier and feminist work cited in [note 2](#) above.

(delusionally) as masters of the natural world.⁷ Others find reason to object in the fact that such an ideal will insist on radical freedom and choice over acceptance of luck and situated-ness as part of its basic moral strategy; others criticize it as an ideal that promotes the 'virtues' of self-sufficiency and independence over interconnectedness and relatedness.⁸

As criticisms of Kant's moral theory, claims such as these cut to the chase. Moral theories, after all, are theories that aim to inform our conceptions of ourselves, indeed to inform reality, by describing possibilities – possible ways of being, ways of interacting, ways of valuing – that we are urged to make real. Moral theories can serve as self-fulfilling prophecies; those with any sympathy at all for the idea that we are free and can guide our own actions via our representations must admit this.⁹ Not everything is possible, of course, and simply wishing a moral theory (or any other theory) true will not make it so. Nonetheless, and whatever the metaphysics, it is a fact that people, collectively and individually, can and do change, bringing themselves more in line with their ideals. Choosing a moral theory, aligning oneself with a certain view, is therefore not only a matter of weighing up which view seems most conceptually coherent, or seems to have philosophy on its side. It is also a matter of deciding which moral theory we would most *like* to be true, which one fits best with what we know and feel about ourselves, and with our aspirations for ourselves and for each other. People are therefore right to eye moral theories with questions about who and how they want to be uppermost; people are right to press questions about the vision of self and its relations that the theory promotes.

For these reasons, it will be useful to make explicit the virtues, the appeals, and the attractions that are internal to free rational willing as Kant understands it. Lest there be misunderstanding, it should be said that what I aim to do here is not to say why Kant thought free rational willing was good, *as if it were good for some reason other than its being free rational willing itself*. The aim rather is to unpack what makes it good in

⁷ This charge is leveled by some feminists, environmental philosophers, and critics of modern technophilia such as Horkheimer and Adorno (see, e.g., Max Horkheimer and Theodor Adorno, *Dialectic of Enlightenment: Philosophical Fragments*, ed. G.S. Noerr, trans. E. Jephcott [Stanford: Stanford University Press, 2002]).

⁸ These strands are to be found in many of the critical works already mentioned, for example Sandel, *Liberalism*.

⁹ In this connection, see Charles Taylor's arguments on behalf of a view of human beings as 'self-interpreting' and 'self-defining.' Charles Taylor, "Interpretation and the Sciences of Man," in his *Philosophy and the Human Sciences: Philosophical Papers*, 2 (Cambridge and New York: Cambridge University Press, 1985), 15–57 and "Self-Interpreting Animals," in his *Human Agency and Language: Philosophical Papers*, 1 (Cambridge and New York: Cambridge University Press, 1985), 45–76.

itself – to describe *the ways that* it is good, one might say, rather than to provide an (independent) *why*. My elaboration will look at the experiences of oneself, of one's agency, and of one's relationships – of the 'lived reality' – made available if one understands oneself as a free rational willing subject à la Kant. With such an elaboration in hand, along with a grasp of the conceptual/philosophical web that, for Kant, supports it, we can think seriously and productively about what we want to insist on, and what perhaps rethink, in Kantian moral theory.

A brief description of the free rational willing subject à la Kant

Before we proceed, it will be useful to review a few key points. To be a free rational will, for Kant, is *to follow an action-guiding principle one has given oneself*. Because to do this is to give oneself a law, to do this is also to be *autonomous*. Now, in order to truly give oneself an action-guiding principle, in order for an action-guiding principle to be truly one's own, and hence in order for one to be free in adopting and following it, it must be grounded in something internal to oneself. That which is internal to oneself in the relevant way for Kant is *reason* (see, e.g., G 4:446–7, 458, 460). How can an action-guiding principle be grounded in reason? We have seen the answer in preceding chapters: it must serve an end that is reason's own. This end, we have argued, is the free rational activity of the will itself. Reason, in its relevant form (as 'pure practical reason') is, as we have seen for Kant, *universal* – it is the same in all human beings. The action-guiding law free rational wills give themselves will therefore be the same in all cases: it will be the moral law that Kant identifies with the categorical imperative (see G 4:432; KpV 5:28–31). There is of course much more that might be said about all of this. Much of it has been said in earlier chapters. We therefore proceed.

THE GOOD OF FREE RATIONAL WILLING

Kant's texts are rich with material that describes, in glowing terms, selves and lives characterized by free rational willing. I have singled out eight such features. They are (1) elevation above nature and participation in a 'higher' order of being, (2) power and agency, (3) self-sufficiency, (4) self-contentment, (5) regularity or orderliness, (6) universality and necessity, (7) infinity, and finally (8) creativity. All count for Kant as reasons to aspire to free rational self-hood, that is, as reasons for reason to make free rational activity of will its end. They are interrelated, and one might parse them differently – I don't

think it matters a lot. What matters is that, taken together, they present us with a conception of our selves to which Kant thought we would aspire.

Elevation above nature and participation in a higher order of being

If we look around in Kant's texts for clues about why he takes free rational willing to be good, the first thing we notice is that, for Kant, free rational activity elevates a person above nature and transports her into a *higher order of being*.

We may begin with a famous passage from the *Critique of Practical Reason* (one we also looked at in [Chapter 5](#)). Duty's root or origin, Kant writes,

can be nothing less than what elevates a human being above himself (as a part of the sensible world), what connects him with an order of things that only the understanding can think and that at the same time has under it the whole sensible world and with it the empirically determinable existence of human beings in time. (KpV 5:86 7)

The 'nothing less' which elevates us above our 'sensible' selves is, Kant writes, 'personality' (KpV 5:87):

It is nothing other than *personality*, that is, freedom and independence from the mechanism of the whole of nature, regarded nevertheless as also a capacity of a being subject to special laws – namely pure practical laws given by his own reason, so that a person as belonging to the sensible world is subject to his own personality insofar as he also belongs to the intelligible world. (KpV 5:87; Kant's italics)

In willing freely and rationally, we are loosed of nature's mechanism and experience ourselves as part of an order of things higher than the sensible world. Kant's passage continues:

it is then not to be wondered at that a human being, as belonging to both worlds, must regard his own nature in reference to this second and highest vocation only with reverence, and its laws with the highest respect. (KpV 5:87)

Kant writes further on of the "sublimity of our own supersensible existence," which, "subjectively effects respect for their higher vocation in human beings" (KpV 5:88). Kant repeatedly writes that our "intelligible" or "rational" nature – our capacity for free rational activity – suits us for a "higher purpose" (e.g., KpV 5:61, 62; G 4:395, 430). And Kant refers often to the goodness of being apart from the world of sense – of being above and independent of it; he does this in many of the passages already mentioned, as well as at, for example, KpV 5:158, 159, 162, and G 4:457.

The world of sense, above which free rational willing places us, is identified for Kant with nature, or the world of appearances or phenomena; the world to which free rational willing gains us membership is the world of freedom, the world structured by the moral law, wherein we find Kantian intelligibilities or noumena. What is so attractive about leaving the world of sense and entering the world of freedom? This question may hardly seem worth asking. But Kant's critics charge precisely that leaving the world of sense is nothing so wonderful, or is a strange fantasy. So let us try to unpack Kant's motivations for making the claim, the meanings it had for him, would have had for many of his contemporaries, and probably has for many of us.

'Elevation above the sensible world' places us, Kant writes, "above animality" (KpV 5:160).

[T]he moral law reveals to me a life independent of animality and even of the whole sensible world. (KpV 5:162)

Man has a duty to raise himself from the crude state of his nature, from his animality (*quoad actum*), more and more toward humanity, by which he alone is capable of setting himself ends. (MS 6:387)

What is so attractive about leaving animality behind? Elevation above animality presumably exempts us from the forms of human behavior people routinely (if problematically¹⁰) describe as 'animalistic' or 'brutish' – wild-eyed, hungry scavenging, unkempt gluttony, out-of-control passion, other ways of being that seem driven exclusively and powerfully by 'basic instincts.' In elevating us above animality, participation in the higher order revealed and governed by the moral law also exempts us from the blind repetitions of animal life – as individuals and as a species we learn, and we change ourselves. Kant sounds this theme in many of his essays, as well as in the *Critique of the Power of Judgment* (see KU 5:429–34).¹¹

A related reason to want to leave the sensible world is that it is a world in which we are pushed around, blinkered and willy-nilly. "Inclination," Kant writes, "is blind and servile" (KpV 5:118). It is 'blind,' presumably, in not being able to see past its own nose – inclinations clamor 'me, me, me!' but are rarely far-sighted; reason, in contrast, confers on us a long view. Inclination is 'servile' in being ordered around by nature; inclination is

¹⁰ Anyone interested in some of the ways intellectuals have misunderstood animals must read Vicki Hearne's wonderful book, *Adam's Task: Calling Animals by Name* (Pleasantville, NY: Common Reader Press, 2000; first published in 1982).

¹¹ This idea is picked up on and put to impressive use by Hannah Arendt in *The Human Condition* (Chicago: University of Chicago Press, 1958).

tethered to the gears and springs that constitute “the mechanism of the whole of nature” (KpV 5:87), of which we’d just as soon fly free. Participation in a higher order relieves us, at least in significant part, of being animate cogs in nature’s machine, reacting blindly and mechanically to an external environment that is indifferent to us.

Power and agency

Becoming a free rational will promises us all sorts of otherwise unavailable power. It reveals to us that we can subdue warring and intemperate inclinations. Of the would-be moral subject, Kant writes: “The vices, the brood of dispositions opposing the law, are the monsters he has to fight” (MS 6:405). Fighting this brood – exercising one’s capacity for rational self-determination – is not easy but it is rewarding. It delivers a sense of oneself as a special kind of powerful and potent agent.

As Kant conceives them, the inclinations (or, here, “acts of sensibility”) stand in need of strong governance.¹²

There is in man a certain rabble element which must be subject to control, and which a vigilant government must keep under regulation, and where there must even be force to compel this rabble under the rule in accordance with ordinance and regulation. (VE 27:360)

Sensibility is not only rebellious, but also mischievous, attempting to ‘dupe and outwit’ understanding. “With regard to the senses in general, since they dupe and also outwit the understanding, we can do nothing else but outwit them in turn” (VE 27:364). When we outwit them, we exercise “strength of soul” (MS 6:384), “self-constraint, that is, inner necessitation” (KpV 5:83), and “power” (KpV 5:152). Of course, it is permissible to act on some inclinations, but only where *we* decide it is right to act on them – we must not be swayed by their own “impetuous importunity” (KpV 5:161). Tenacity in resisting their importunity Kant describes as *virtue*:

Now the capacity and considered resolve to withstand a strong but unjust opponent is *fortitude* (*fortitudo*), and, with respect to what opposes the moral disposition *within us*, *virtue* (*virtus, fortitudo moralis*). (MS 6:380; Kant’s emphases)

¹² This is of course not to say that for Kant, inclination needs to be *extinguished* or *destroyed*. Barbara Herman and others have been quite right to point out that inclination has an indispensable role to play in good Kantian moral judgment (see, e.g., Barbara Herman, “Integrity and Impartiality,” in her *The Practice of Moral Judgment* [Cambridge, Mass.: Harvard University Press, 1993], 23–44). But it would betray the texts to suggest otherwise than that Kant’s attitude toward inclination is often antagonistic.

Where resolve falters, where we fail the moral law, we exhibit weakness, and display a “slack, or indeed ignoble, attitude” (G 4:426), betraying “weariness,” and embracing “a dream of sweet illusions” (G 4:426). We should not let this happen: morality demands constant *watchfulness* or *vigilance* (VE 27:348). “We first have to discipline ourselves,” Kant writes, “i.e., to root out, in regard to ourselves, by repeated actions, the tendency that arises from the sensory motive” (VE 27:361). For Kant there is thus a problem if we find ourselves having *inclinations* toward behavior that looks like self-control, or if we acquire temperate *habits*. Virtuous power over inclinations is no power at all to the extent that it becomes, we might say, second nature. Under these circumstances, we are apt to let the reins of internal vigilance slack (MS 6:396, 383–4) and to find ourselves “neither armed for all situations nor adequately secured against the changes that new temptations could bring about” (MS 6:384).

But I am always capable of willing freely and rationally; nothing can prevent me from willing thus. The power or strength available here is thus *always* available, and is worthwhile:

For by strength of soul we mean strength of resolution in a man as a being endowed with freedom, hence his strength insofar as he is in control of himself (in his senses) and so in the state of *health* proper to a man. (MS 6:384; Kant's emphasis)¹³

In what does this ‘health’ consist? Self-control delivers us ourselves as powerful, coherent *agents*. “In self-mastery there lies an immediate worth, for to be master over oneself demonstrates an independence of all things” (VE 27:361–2). Without this, the agent, “becomes a plaything of other forces and impressions, against his choice, and is dependent on chance and the arbitrary course of circumstances” (VE 27:362). Under these circumstances, Kant writes, a person's “thinking state” becomes “arbitrary” (VE 27:362), and her energies lose focus. In contrast, self-control enables us to organize and actively pursue a coherent set of self-chosen ends, rather than be a mere passive conduit of natural impulse and energy.

In the following note from the *Groundwork*, Kant makes clear how appealing he takes the power of self-control to be.

For the most ordinary observation shows that if we represent, on the one hand, an action of integrity done with steadfast soul, apart from every view to advantage of

¹³ Kant here points out that strength of soul proper is always directed at subduing passions *in the service of the moral law* (which after all is the only time passions are truly overcome, and not merely juggled): “it is not only unnecessary but even improper to ask whether great *crimes* might not require more strength of soul than do great virtues ... great crimes are paroxysms, the sight of which makes a man whose soul is healthy shudder” (MS 6:384).

any kind in this world or another and even under the greatest temptations of need or allurements, it leaves far behind and eclipses any similar act that was affected in the least by an extraneous incentive; it elevates the soul and awakens a wish to be able to act in like manner oneself. (G 4:411n)

The impressiveness of self-control may even embolden us into thinking that it constitutes moral worth itself.

Moderation in affects and passions, self control, and calm reflection are not only good for all sorts of purposes but even seem to constitute a part of the *inner* worth of a person. (G 4:394; Kant's emphasis)

Kant of course is here ready to remind us that "without the basic principles of a good will they [moderation, self-control] can become extremely evil" (G 4:394). This is so, but does not change the fact that self-control is part and parcel of willing freely and rationally, and that the appeals of self-control deliver much of free rational willing's appeal. The free rational willing Kantian subject, who exercises vigilance and self-control, can take satisfaction not only in this power itself, but can also anticipate, according to Kant, freedom from the confusing and chaotic incursions of sensibility, a more coherent set of projects and purposes, and a more focused efficacy. In conclusion here, a long quotation from the *Critique of Practical Reason*:

the pure moral motive must be brought to bear on the soul ... because it teaches the human being to feel his own dignity gives his mind power, unexpected even by himself, to tear himself away from all sensible attachments so far as they want to rule over him and to find a rich compensation for the sacrifice he makes in the independence of his rational nature and the greatness of soul to which he sees that he is called. (KpV 5:152)

Self-sufficiency

In experiencing ourselves as free rational agents, and so as capable of choosing and acting on grounds that come from our own reason, we experience self-sufficiency in several forms. The first is intellectual: Kantian practical reason acts independently of sensibility, disregarding sensibility's input in framing fundamental principles and in forming its 'elementary concepts,' viz., concepts of good and evil:

instead of the form of intuition (space and time), which does not lie in reason itself but must be drawn from elsewhere, namely from sensibility, these, as practical elementary concepts, have as their basis the *form of a pure will* as given within reason and therefore within the thinking faculty itself. (KpV 5:65 6)

We do not need to look outside ourselves to come to understand good and evil. Practical reason functions, in a priori essentials, unaided.

Second, *moral* self-sufficiency is promised by free rational willing. As free rational wills, we are self-sufficient in our *goodness*. In acting on the moral law, we instantiate moral goodness; this goodness is something we can 'effect' without the cooperation of nature or anything in the external world. "To satisfy the categorical command of morality is within everyone's power at all times" (KpV 5:36–7). Kant famously claims that even a person who "by a special disfavor of fortune or by the paltry provision of a step-motherly nature" is unable to effect the outcomes he intends can nonetheless be morally good, since it is a question not of the effects he produces but of the quality of his will (G 4:394).¹⁴ Moral value or worth is something that cannot be taken from us, rendering our worth and moral status thereby invulnerable to, as Kant writes below, "chance or fate." Kant is here discussing what he calls "practical wisdom":

Only in its possession is a man "free," "healthy," "rich," "a king," and so forth, and can suffer no loss by chance or fate, since he is in possession of himself and the virtuous man cannot lose his virtue. (MS 6:405)

My practical wisdom, my virtue, my good will, and hence my moral goodness, is – if I adopt Kant's view of things – mine to keep or lose.

Self-sufficiency enters the picture in a third way. Here, free rational willing does not promise but rather *requires* a form of self-sufficiency, with salutary effects:

he who accepts favors creates debts that he cannot repay; he can never even the score with his benefactor, since the latter first did him the kindness of his own accord; if he returns the favor, he does it only insofar as the other preceded him in this, and thus remains forever owing thanks to him; but who will incur such debts? A debtor is at all times under the constraint of having to treat the person he is obliged to with politeness and flattery; if he does not, the benefactor soon lets him know of it, and often he has to circumvent the latter with many detours and greatly burden himself. But he who pays promptly for everything can act freely and nobody will hamper him in doing so. (VE 27:341 2)

This passage stands in some tension with Kant's clear understanding that human beings often need help – witness the fourth of his famous illustrations of the categorical imperative, which insists that one may

¹⁴ I have here altered Gregor's translation. The German is "*gleich durch eine besondere Ungunst des Schicksals, oder durch kärgliche Ausstattung einer stiefmütterlichen Natur.*" See also this comment from the lectures on ethics: "Socrates was in a sorry state, which had no value at all, but his person, in this condition, was of the greatest worth" (VE 27:344).

not will the general withholding of aid (G 4:423). This tension is not so easily resolved, but we may at least note that this passage's aim is to stress the unacceptability of placing oneself in a perpetually beholden position. One is not, Kant might argue, required to be sufficient to every situation life throws up – it is acceptable to need help. But one must, Kant does seem to be saying, be sufficient, over the long haul, to the help one has received – that is, one must be able to pay for it, to repay it, to have earned it, in at least some rough sense. Free rational willing demands *this* kind of self-sufficiency because without it, one becomes too vulnerable to the demands and expectations of those to whom one is indebted. This sort of vulnerability, like any relation with others that compromises our capacity to think and act as reason dictates, is inimical to free rational willing.¹⁵

Each of these forms of self-sufficiency, whether benefits of free rational willing (intellectual and moral self-sufficiency) or aids to it (personal self-sufficiency), are connected intimately enough with it to be counted among potential appeals of free rational self-hood. All, we may plausibly imagine, can serve as grounds of a kind of pride, the satisfaction of the successful do-it-yourselfer. (Satisfaction of this sort is of course connected to feelings of agency and power as well.) Moreover, as the avenue to intellectual and especially moral self-sufficiency, free rational willing protects one from the slings and arrows of fortune, constituting a self that is imperturbable. This brings us to the next set of reasons for finding the experience of free rational Kantian subjectivity appealing.

Self-contentment

“It is up to us,” Kant writes, “to put ourselves into a certain mood, which is a voluntarily chosen disposition, whereby we contemplate the world and its destinies, and from which we pass judgment upon them” (VE 27:368). To do otherwise is to fall prey to the moods our feelings suggest. And this is tantamount to leaving our fates – our happiness, our well-being – to, well, fate.

¹⁵ A similar thought is behind Kant's exclusion of propertyless workers from enfranchisement (MS 6:314–15); in Kant's view, workers' votes would merely reflect the wills of those upon whom they were dependent, namely their employers. Truly free rational decision-making demands that I be my own person, and not someone else's. There is much to question in this line of Kantian thought; for an excellent discussion, see Susan Mendus, “Kant: ‘An Honest But Narrow-Minded Bourgeois?’” in *Essays on Kant's Political Philosophy*, ed. Howard Williams (Chicago: University of Chicago Press, 1990), 166–90.

Choosing one's moods may be an exercise in free rational willing for the fairly advanced. But in general Kant is confident that the happy disposition of *self-contentment* is reliably available to free rational subjects, and is part of what makes free rational willing attractive. Please bear here with a lengthy quotation:

Have we not, however, a word that does not denote enjoyment, as the word happiness does, but that nevertheless indicates a satisfaction with one's existence, an analogue of happiness that must necessarily accompany consciousness of virtue? Yes! This word is *contentment with oneself* [*Selbstzufriedenheit*], which in its strict meaning always designates only a negative satisfaction with one's existence, in which one is conscious of needing nothing. Freedom, and the consciousness of freedom as an ability to follow the moral law with an unyielding disposition, is *independence from the inclinations*, at least as motives determining (even if not as *affecting*) our desire, and so far as I am conscious of this freedom in following my moral maxims, it is the sole source of an unchangeable contentment, necessarily combined with it and resting on no special feeling, and this can be called intellectual contentment. (KpV 5:117 18)

The prospect, offered by free rational willing, of "satisfaction with one's existence," of being, "conscious of needing nothing" – indeed of "unchangeable contentment" – is attractive indeed. Kant contrasts the real availability of such contentment with the chimera of sensuous contentment. Sensuous contentment is in fact impossible:

For the inclinations change, grow with the indulgence one allows them, and always leave behind a still greater void than one had thought to fill. (KpV 5:118)

Now, it is of course true that free rational willing sometimes rankles: the self-constraint involved in placing ourselves "under a *discipline* of reason" (KpV 5:82; Kant's emphasis) brings with it "sacrifice" (KpV 5:83, 152), even "pain" (KpV 5:160). But submission to such discipline also leads precisely to self-contentment. Here is Kant describing an exercise in moral pedagogy (this is where he mentions the 'pain' of renouncing inclination):

the pupil's attention is fixed on the consciousness of his *freedom* and, though this renunciation excites an initial feeling of pain, nevertheless, by its withdrawing the pupil from the constraint of even true needs, there is made known to him at the same time a deliverance from the manifold dissatisfaction in which all those needs entangle him and his mind is made receptive to the feeling of satisfaction from other sources. (KpV 5:160 1)

The feeling of contentment "from other sources" is due not only to the freedom from the 'manifold dissatisfaction' of inclination, but also to some of the other benefits of free rational willing we have mentioned here. There

is contentment to be found in confidence that we are not just brutish animals, or cogs in nature's machine; in an awareness of our own potency, our power and dominion at least over our inner lives; and in the thought that we are significantly in control of our fates, free from the whims of nature in important matters and fully sufficient to tasks we deem key. As the next set of considerations will show, free rational willing also promises to bring another likely source of contentment, namely a pleasing order and organization, to our lives.

Regularity or orderliness

An advantage to accepting governance by the moral law often advertised by Kant is that it affords certainty, sureness, and reliability. The moral law *is* reliable; it is a rule that can be counted on, that promises certain advice, and not just risk assessments or lists of pros and cons; the moral law tells us how we *must* order our ends, not just how we might like to given x , y , or z – which may or not in fact be given. And it tells us according to a rule that will always hold for us, as well as for all other agents. This, for Kant, is a clear advantage over any guidance available from inclinations, which cannot “furnish a uniform standard of good and evil” (G 4:442).

The moral law is thus uniquely able to prevent internal conflict and help us ‘find rest’ (see G 4:405). Inclinations make competing claims on our wills, recommending a changing multitude of ends; duty makes claims as well. The law is clear, however, about what trumps what, and tames our “propensity to rationalize” (G 4:405). “Common human reason,” Kant writes, is impelled to step into practical philosophy not only in order to learn the “source” and “correct determination” of the moral law, but also “that it may escape from its predicament about claims” (G 4:405). In a dramatic claim about the burdensome confusion of even pleasant feelings (a claim often seized on by critics), Kant writes,

Even this feeling of compassion and tender sympathy, if it precedes consideration of what is duty and becomes the determining ground, is itself burdensome to right thinking persons, brings their considered maxims into confusion, and produces the wish to be freed from them and subject to lawgiving reason alone. (KpV 5:118)

The moral law frees us from internal turmoil: the law “is the sole condition under which a will can never be in conflict with itself” (G 4:437).

One could indeed be forgiven for imagining that Kant finds goodness in non-contradiction itself:

That will is absolutely good which cannot be evil, hence whose maxim, if made a universal law, can never conflict with itself. (G 4:437; Kant's emphasis)

Of course, Kant's aim in this passage is to distinguish absolutely from not-absolutely good wills – not to define goodness as non-contradiction. But as we have seen, moral goodness *is* coextensive with non-self-contradictory willing.¹⁶ And the point here is that part of the appeal of willing in ways both free and rational is that doing so will resolve internal conflict.

A related effect of submitting inclinations to governance by the moral law is that such submission *unifies* consciousness, at least what we might call 'practical' consciousness, or will. Sound practical reasoning serves "to subject a priori the manifold of *desires* to the unity of consciousness" (KpV 5:65). Adherence to the moral law contributes to "the unity and harmony of the mental powers," helping one retain "presence of mind" and aiding in "carrying out one's business" (VE 27:366). (A related point about unification of the 'theoretical' manifold, and the pleasure we take in this unification, is to be found at KU 5:187–8.) Free rational willing lets us keep it together, press ahead, and remain ourselves despite the blooming buzzing confusion of inclinations pressing in.¹⁷

Just as attention and responsiveness to the moral law can order our wills, so can it order – make systematic, unconfused – our ideas about morality. Kant writes that we need to identify and adhere to the moral law in order to avoid the

disgusting hodge podge of patchwork observations and half rationalized principle, in which shallow pates revel because it is something useful for everyday chitchat. (G 4:409)

Such a "hodge-podge" confuses "the insightful," who "feeling confused and dissatisfied without being able to help themselves, avert their eyes" (G 4:409). "Morals themselves," Kant writes, "remain subject to all sorts of corruption as long as we are without that clue and supreme norm" (G 4:390).¹⁸ Pleasing systematicity, order, regularity, and rational coherence are, in contrast, delivered by the law grounded in pure reason.

¹⁶ This is most clearly seen by looking at Kant's 'contradiction in conception' and 'contradiction in willing' tests for maxims (G 4:424), discussed in Chapter 6 above. It is also interesting to compare REL 6:35.

¹⁷ A related suggestion is made by Thomas E. Hill, Jr., "The Importance of Autonomy," in his *Autonomy and Self-Respect* (Cambridge and New York: Cambridge University Press, 1991), 51.

¹⁸ Also see Kant's comment in the *Critique of Practical Reason* that "Principles must be built on concepts; on any other foundation there can be only seizures [*Anwandelungen*], which can give a person no moral worth and not even confidence in himself" (KpV 5:157).

One final set of comments about the ways free rational willing brings with it appealing regularity and orderliness. In the *Lectures on Ethics*, Kant contrasts his two lawful systems (that of reason, or freedom, on the one hand, and that of nature on the other) with *anarchy*. Anarchy, or the absence of rules, in nature *or* in freedom, leads to destruction.

If freedom is not restricted by objective rules, the result is much savage disorder. (VE 27:344)

If there were no order in nature, everything would come to an end, and so it is too with unbridled freedom. (VE 27:346)

Since we have the freedom and the ability to satisfy our inclinations by all manner of devices, men would bring about their downfall, in the absence of restraint. (VE 27:347)

“Unbridled freedom” is the incomprehensible sort of freedom that follows no rule. Its incomprehensibility does not deter Kant in warning against it. The disorder it brings is infinitely destructive. A subject possessed of it would be “a diabolical being” (REL 6:35). Nature would “come to an end” without its own rules; likewise humanity will destroy itself if inclinations and actions directed at their satisfaction are not organized and governed. Free rational willing, here, is pretty obviously good.

Universality and necessity

We turn now to two experiences, available to the Kantian subject with a free rational will, that derive from the fact that such a subject must give itself a law. These are the experiences of universality and necessity.

First, let us recall that, when I will freely and rationally, I will on the moral law, which, being a law, is a principle that carries with it universality and necessity. In the case of the moral law, focus on the joint demands of universality and necessity can, as we have seen, reveal much about the law itself. The moral law’s universality is foregrounded and translated into content in its ‘universal law’ formulation (G 4:402, 421), according to which I may will only in ways that are universalizable. The moral law’s necessity derives from its grounding in a necessary end, which Kant asserts must be humanity itself, giving rise to the ‘humanity as an end in itself’ formulation of the law (G 4:429).

For these reasons, when I will freely and rationally, I will ‘with universality and necessity’ in several senses. Of course, whenever I will, my will itself is *law-governed*. But this of course is true even if I will heteronomously: if

I will heteronomously I am simply governed by laws not my own. However, when I will freely and rationally, I will, first of all, on universalizable maxims, and am thus willing in a universal, because universalizable, way – I will maxims that are endorsable from a ‘universal point of view.’ Second, I am ‘necessitated’ or obligated to will as I do because such willing is directed at a necessary end. Third, insofar as I understand myself as the *author* of the moral law, I “cognize that my connection with that [invisible] world [of freedom] (and thereby with all those visible worlds as well) is not merely contingent, as in the first case, but universal and necessary” (KpV 5:162). In legislating a universal and necessary law, I ‘link up’ to the universal space of reason and contribute activity that is necessary if there is to be moral law, and hence a morally structured realm, at all.

Why or how might such involvement with universality and necessity *appeal*? To what self-conceptions or experiences does this give rise, and what is attractive about it? Take universality first. The moral law I make is the same law all would make – it would be universally endorsed. For this reason, I can and do make it for all – I legislate universally. I am entitled to this because the faculty I use, reason, and the perspective I adopt in using it are universal – the law does not come from a private, idiosyncratic place. To what experience of self does this give rise? I abandon a view of myself as a mere accident of self-interest or positionality or personal preference, experiencing myself as ‘bigger’ than all that. In experiencing the ‘universal in me’ – my rationality, my capacity for autonomy, my commitment to respecting the same – I also experience, for Kant, my “proper self” (G 4:458), and am simultaneously most connected to, most ‘in common with,’ others. Satisfactions we might call those of authenticity and community seem made available here.¹⁹

What about necessity? The moral law carries necessity. In Kant’s expression, it ‘necessitates,’ it unconditionally demands our allegiance, because it is grounded in, it aims at, free rational willing, which is a necessary end for us. Moreover, it is necessary *that the law be* – without it, there would be no

¹⁹ A recent paper by Paul Guyer brought the following note from the *Reflexionen* (probably 1776–8) to my attention:

Freedom is the original life and in its connection the condition of the coherence of all life; hence that which promotes the feeling of universal life or the feeling of the promotion of universal life causes a pleasure. Do we feel good in universal life? The universality makes all our feelings agree with one another, although prior to this universality there is no special kind of sensation. It is a form of *consensus*. (R 6862, 19:183; Guyer’s translation)

Although strange in many ways, the note is certainly suggestive of a special pleasure in universality/universalization. For further discussion, see Guyer, “Kant on the Theory and Practice of Autonomy,” in his *Kant’s System of Nature and Freedom*, 115–45, *Reflexion* on p. 131.

'moral realm,' no way of thinking about the world in terms of any but contingent values. This renders our law-making activity necessary. Without us, there would be no moral realm. What is attractive about this? The necessity integral to free rational willing ensures that if I will in this way, I transcend contingency: in adopting a necessary end, I leave aside accidental needs, interests, projects, etc., and pursue something the value of which is unqualified. Furthermore, I understand my own law-making activity as itself necessary, as crucial to the fiber of the world of value. Indeed, it is in virtue of this capacity for law-making activity that I myself instantiate unqualified, necessary value. My way of being itself can be seen as necessary – and therefore as something significantly more impressive than if I could only conceive myself as a contingent coming together of passive atoms in a impersonal Newtonian void.

Infinity

Perhaps even more grandly (if this is possible), autonomy permits me to participate in *infinity*. The moral law, Kant writes,

begins from my invisible self, my personality, and presents me in a world which has true infinity ... [It] infinitely raises my worth as an *intelligence* by my personality, in which the moral law reveals to me a life independent of animality and even of the whole sensible world, at least so far as this may be inferred from the purposive determination of my existence by this law, a determination not restricted to the conditions and boundaries of this life but reaching into the infinite. (KpV 5:162)

As a free rational will, I am unbounded. How so? As a free rational will, I am not limited by space or time, by foreign desires, by empirical facts. My self is not my phenomenal self, determined by natural law, and it is therefore not a self that is usefully thought of in terms of phenomenal space and time. I can aim at whatever I can imagine, set ends and design projects that go beyond what exists in the world. What I, as a natural being, can actually do in the world of space and time may of course be limited, but my will itself, insofar as it is free, is not.

There is something in this reminiscent of a feeling described by Hans Castorp in Thomas Mann's *The Magic Mountain*. Castorp has been at the spa in the Alps for many months, and has begun to forget his life back in Hamburg – life “in the flatlands,” as spa residents call it. One of the activities Castorp has come most to enjoy is what he calls “playing king.” Here is his description:

And then to the sounds of the rushing brook and amid blue blossoming columbine, he would lean back against the crude wooden bench, cross his arms, tilt his head to one shoulder, and begin to reminisce about it “all.”

That sublime image of organic life, the human body, hovered before him just as it had on that frosty, starlit night when he had pursued his learned studies; and in contemplating its inner aspect now, young Hans Castorp was caught up in a great many questions and distinctions . . . even though down in the flatlands he had never noticed such questions, probably never would have noticed them, but certainly did here, where one looked down on the world and its creatures from the contemplative retreat of five thousand feet and thought one's thoughts . . . He called up the images of the two grandfathers, placing them side by side, the rebel and the faithful servant, who both wore black but for different reasons, and he considered their merits. He went on to deliberate such vast complexities as form and freedom, mind and body, honor and disgrace, time and eternity – and was overcome by a brief, but frantic dizzy spell at the thought that the columbine was blooming again and the year had come full circle.

He had a special term for this responsible preoccupation with his thoughts as he sat at his picturesque secluded spot: he called it “playing king” – a childish term taken from the games of his boyhood, and by it he meant that this was a kind of entertainment that he loved, although with it came fear, dizziness, and all sorts of heart palpitations that made his face flush even hotter. And he found it not unfitting that the strain of all this required him to prop his chin – and the old method seemed perfectly appropriate to the dignity he felt when “playing king” and gazing at that hovering sublime image.²⁰

What Castorp enjoys, and what gives him such a sense of well-being, is the play of his mind over various possibilities and ideas, the combination of limitlessness and of mastery, of responsibility only to himself for his path and conclusions. “Playing king” is not exactly the same as playing Kantian universal legislator – playing king is not an activity of pure practical reason and is not focused on moral law,²¹ and Castorp here is a bit absurd – but the description nonetheless I think captures an experience of self as simultaneously above and unconstrained by the given. Castorp involves himself here with “a determination not restricted to the conditions and boundaries of this life but reaching into the infinite” (KpV 5:162).

We have pointed, so far, to the free rational Kantian subject's elevation above nature, which appeals both as release from animality and from blind mechanism, and which includes us in a calmer, freer, more far-sighted order of being. We have seen that willing freely and rationally is willing that both

²⁰ Thomas Mann, *The Magic Mountain*, trans. John E. Woods (New York: Random House, 1996), 383.

²¹ The difference in political metaphor is also thus significant: to play king is to play at willing whatever one wants, where to play Kantian universal legislator is to play at the more restrained game of willing whatever everyone could want.

requires and produces heightened power and agency. Free rational willing brings with it several forms of self-sufficiency, deep self-contentment, conflict-resolving regularity or orderliness, and entitles us to claim for ourselves kinds of universality, necessity, and infinity. The argument here is that these are all part of the experience and self-conception available to agents who adopt Kant's understanding of morality and who, accordingly, will in ways that respect – by preserving, promoting, and instantiating – free rational willing. These experiences are not accidental by-products. They are not due, for instance, to some contingent set of connotations suggested by 'free,' 'rational,' and 'will'; they are not attributable just to the fact that Kant was thinking in the eighteenth century or grew up Pietist; they are inseparable from Kant's considered conception of free rational willing, and are deeply implicated in the metaphysical and moral pictures that support this conception. And, the argument is, drawing them out as we have helps make clear what Kant takes to be so attractive about free rational willing.

We should now start to have a feel for those attractions. Many of them – the satisfactions, benefits, assurances, and related pleasures – that Kantian moral subjectivity promises will strike familiar chords with readers today, as will key components of a Kantian view of the ideal subject. Such components include the ideas that human beings are qualitatively distinct from and higher than animals, that desires are in need of discipline, that certain forms of reason are self-sufficient (that is, can reach sound conclusions independent of empirical evidence or input), that a life spent satisfying sensuous desires will never be a contented life, that coherence, unity, harmony, and consistency are desiderata for a healthy self, that particularity, contingency, and limitation are all best transcended, insofar as such transcendence is possible.

It is of course true that these ideals and promises are not *unique* to Kantian morality – many have found homes in moral theories and ideals of the self before and after Kant. It is of course also true that these ideas and promises will not speak to everyone. But the continued currency and widespread appeal of many of the components of Kantian free rational willing, together with the coherent package Kant gives them, makes Kant's ideal a particularly potent one for proponents and detractors alike.

I argued at the beginning of this chapter (as well as in [Chapter 1](#)) that the best defense of Kantian moral theory will have to be a defense of the ideal it promotes. Such a defense would have to persuade its audience that the ideal deserves their allegiance, and on various sorts of grounds: that it fits with existing aspirations, or with the best of them, that it delivers a picture of agency in the world – a metaphysics – that is plausible and satisfying, that it

organizes the self and its relations in ways that appeal. Accordingly, it will have to proceed by drawing on experience, history, psychology, literature, on what human beings know and feel about ourselves, and tell a good story about why free rational willing describes something we either already do or should value.

The *final section* of this chapter makes one small but I hope significant stab in this direction. My stab will not amount to a full-blown defense of free rational Kantian willing, but will rather suggest a way to think about it that highlights and brings together many of its appeals. Some of us may still find that there are parts of the Kantian picture we want to revise, but I want at least to show here how an appealing complex of parts hangs together. I think we can also see that, whether he understood his view this way or not, Kant's own attachment to the free, rational activity of human wills is deeply motivated by an appreciation of precisely the complex of virtues I suggest here.

The possibility of creativity

The suggestion is this. Kantian free rational willing is closely related to human creativity – to innovation, to the introduction of the radically new into the world, to the design and realization of new things and new ways of doing things. Recall the features of free rational willing we have looked at already. Creativity is implied in the exemption from the blind repetitions of nature. Creativity is a, if not the, primordial form of power or potency; creativity, as a form of authorship, is a prime exercise of agency. Cognitive, personal, and perhaps even a kind of 'moral' self-sufficiency are at once part of the experience of creating and in certain respects prerequisite to it. Creative effort can be deeply satisfying in a way that the satisfaction of sensuous desires cannot; the self-contentment free rational willing makes available thus bears resemblance to the satisfactions of creativity. Peculiar features of the experience of creativity can be captured by talk of universality, necessity, and infinity. This is, at any rate, the suggestion.

Before I go on to say more, let me clarify what I mean by 'creativity,' and notice its straightforward relationship to free rational willing. First, by 'creativity,' I have in mind all the ways human beings bring new things into the world. Artistic creation is a species of this activity, but is not the only species. New ways of building roads, or treating anxiety, or sentencing misdemeanors, or cooking halibut are all creative. Random mutation is not creative, but sudden solutions to problems on which someone has been working for a long time are: creativity is doing something in a way that is

new and different on purpose. “The practical is everything that is possible through freedom,” Kant writes (A800/B828).²² And it is just so that in willing and acting freely and rationally, we are creative, that is, we introduce new aspects and arrangements into the world. Kant writes in many places about the natural world gaining “intelligible” or “supersensible” form through the free rational activity of moral agents (see, e.g., KpV 5:43–8; G 4:438). The natural world is rationally ‘informed,’ shaped, configured, whenever we can say of something that it was produced by us under the guidance of reason (rather than ‘instinct’). Under this guidance, we have created not just the many physical artifacts that are peculiarly human products, but also the institutions and relationships within which our lives occur. We have created and will continue to create a ‘world’ full of and informed by ideas and meanings that are not from nature.²³

It should be clear that Kant values the activity that generates this ‘intelligible’ world. The fact that we devise plans and carry them out, aspire to and are desirous of things we make up, and are continually reinventing ourselves and the world around us is what leads Kant to call us ends in ourselves.²⁴ The creative power of free rational activity is also what makes Kant’s view profoundly humanist. Kant places us, as free wills and as rational cognizers, at the authorial center of a very important universe, namely our own. We may not create the raw matter, but our cognitive faculties are, for Kant, responsible for the organization of the phenomenal world. The noumenal world is, of course, our creature as well: the laws of freedom, and the objects and concepts that derive from them, are the inventions of reason, acting independent of nature, that is, are the inventions of free rational will. The valuable creativity of such will on such a picture is hard to deny.

These claims – that free rational will is creative, that the creative power of free rational will is valued by Kant – seem to me uncontroversial, but are

²² This is my own translation. The German is: “*Praktisch ist alles, was durch Freiheit möglich ist*” (A800/B828).

²³ I have tried to formulate the claim here in a way that is agnostic about whether free rational agency alters what would ‘otherwise’ have been the course of events, or merely contributes another way of parsing and organizing descriptions of those events and the things that issue from them. Allen Wood seems to favor the first option (Wood, *Kant's Ethical Thought* [Cambridge and New York: Cambridge University Press, 1999], 180–2); Henry Allison might be regarded as an advocate of the second (Allison, *Kant's Theory of Freedom*). Further pursuit of this issue would require addressing, among other things, Kant’s claims about God’s creativity, e.g., the following point made in the *Critique of Practical Reason*: “the concept of **creation** does not belong to the sensuous mode of conceiving of existence or to causality but can be referred only to noumena” (KpV 5:102).

²⁴ On this point, see [Chapter 6](#); see also Christine Korsgaard, “Kant’s Formula of Humanity,” in her *Creating the Kingdom of Ends*, 106–32.

nonetheless rarely made as such.²⁵ I don't think their significance can be overstated. One indication of their significance may be found in how disturbing some of Kant's critics find the creative power of free rational will. Free rational will grants human beings creative independence from nature and from God, neither of which can, on Kant's view, be understood as the source or origin of everything that is, or even, for that matter, of much that is very interesting. Charges of 'secular humanism' – charges leveled at Kant by Christian fundamentalists (and others) – are in this case right on the money.²⁶

What I want to do next is supplement what I take to be the more or less uncontroversial (if underreported and extremely significant) claim that Kantian will *is* creative with a few observations on the experience of creativity. My aim in doing so is to return to the task of describing the lived self-understanding and particular satisfactions promised to us if we choose Kantian moral theory. I look here at what experiences of creativity have in common with the experience of free rational willing described above, drawing on the testimony of two very different thinkers, Virginia Woolf and Adrian Piper.

Woolf is of course the well-known author of *Mrs Dalloway*, *To the Lighthouse*, and other novels. She is also author of the important essay (itself significantly about creativity), *A Room of One's Own*. Adrian Piper is an important, internationally known conceptual artist as well as a philosophy professor and a Kant scholar. Much of Piper's artwork is on race and racial identity. Both Woolf and Piper are interested in, among other things, collective failures to promote or even acknowledge the possibility of full free rational subjectivity in some people (in women, in African-Americans). Both may be read as suggesting, again among other things, that such collective failures disrupt, at least temporarily, the subjective prerequisites of creative activity, particularly artistic and intellectual activity, for those whose subjectivity is at issue. This, I think, makes each a particularly interesting thinker on the relationship between creativity and free rational willing of the Kantian sort.

²⁵ Friedrich Schiller is one thinker who has made this connection explicitly (Schiller, *On the Aesthetic Education of Man in a Series of Letters* [1794], trans. E. Wilkinson and L. Willoughby [Oxford: Oxford University Press, 1967]). More recently, the point has been made by Susan Neiman, *The Unity of Reason: Rereading Kant* (Oxford: Oxford University Press, 1994).

²⁶ It is not accidental that the intellectual (and curricular) battles Christian fundamentalists wage with most vehemence are those against the big bang theory and the theory of evolution. One's answer to the question of who or what created the universe and its occupants has deep and immediate implications for both metaphysics and morality.

I focus particularly on the ways some of Woolf's and Piper's comments resonate with the above-described experiences of Kantian universality, necessity, and infinity. I focus on these as they are perhaps hardest to grasp as forms of experience. If we looked for them, we would also easily find Woolf and Piper treating relationships between creativity and power, self-sufficiency, and self-contentment. (Regularity and self-mastery make less obvious appearances in their reflections on creativity.) And of course, many of these experiences contribute to and support each other – as mentioned earlier, separating the experiences I describe out from each other is somewhat artificial (and could be done differently); they are in any case mostly mutually implicating.

Let us turn first to Virginia Woolf. In *To the Lighthouse*, Woolf writes brilliantly about how creative activity is disrupted when a person is made to feel an object watched, perhaps judged, perhaps desired, reminded in so many ways of her role and place and needs and debts in the mundane world. I'm thinking here especially of Woolf's depiction of Lily Briscoe.²⁷ Lily's determined if insecure efforts to paint are disrupted by the presence of fellow holiday guest, Mr. Tansley, who has disparaged 'lady artists,' and Mr. Carmichael, an eccentric neighbor – both make Lily uncomfortably aware of her appearance, her unmarriedness, and her dependence on their hosts. To be made to feel these things is precisely to be reminded of oneself insofar as one is a particular, contingent, finite object in the phenomenal world, subject to all the heteronomy that comes with that territory. One loses, in being so reminded, a sense of oneself as expansive mind, and loses, until a sense of autonomy is regained, the freedom and power to do, or to make, as one pleases.

As Lily at last begins to paint, finding her brush "fallen in with some rhythm," Woolf's description shifts:

Certainly she was losing consciousness of outer things. And as she lost consciousness of outer things, and her name and her personality and her appearance, and whether Mr. Carmichael was there or not, her mind kept throwing up from its depths, scenes and names, and sayings, and memories and ideas, like a fountain spurting over that glaring hideously difficult white space, while she modeled it with greens and blues.²⁸

In *A Room of One's Own*, Woolf takes up the question of the requirements of creative work explicitly. "Returning again to my original enquiry into what state of mind is most propitious for creative work," she writes,

²⁷ See Virginia Woolf, *To the Lighthouse* [1927] (New York: Harcourt Brace Jovanovich, 1981), esp. 156–62 ("The Lighthouse," section III).

²⁸ Woolf, *To the Lighthouse*, 159 ("The Lighthouse," section III).

the mind of an artist, in order to achieve the prodigious effort of freeing whole and entire the work that is in him, must be incandescent ... There must be no obstacle in it, no foreign matter unconsumed.²⁹

I suggest that we imagine the “state of mind ... most propitious for creative work” that Woolf has in mind here as the state of a self that – at least for a time – does not experience itself as particular, contingent, and limited but rather as universal, necessary, and infinite.

How so? Such a state of mind seems *universal* insofar as it aims at communication, as all artistic and intellectual effort must. A mind or self engaged in such effort strives for, and in happy moments feels it has achieved, communion with the universe of other minds. Moreover, such a self operates and experiences itself as transcending, as moving above and outside, personal ruts and preoccupied self-awareness. Woolf's creative self is a big self, a self that can understand things from points of view not exclusively its own.³⁰

What about *necessity*? We might say that the experience of creativity exemplifies a sense of necessity insofar as creativity often carries with it a sense of urgency – a sense of having found something that *needs* to be said or done (why else undertake the ‘prodigious effort’?). Think of this famous passage, where Woolf argues that Mary Carmichael's writing, is, if not strictly speaking *necessary*, at least very, very important:

For if Chloe likes Olivia and Mary Carmichael knows how to express it she will light a torch in that vast chamber where nobody has yet been ... That is a sight that has never been seen since the world began, I exclaimed.³¹

A related thought is found in Nietzsche, who, in a notebook entry entitled “On the Pathos of Truth,” describes a philosopher struck by illumination as experiencing therein *his own* sudden and incontrovertible necessity.³² We also experience ourselves as necessary conditions – sources – of the things we create. And we suffer acute disappointment at being ‘scooped’ not just

²⁹ Woolf, *A Room of One's Own* [1929] (New York: Harcourt Brace Jovanovich, 1981), 56.

³⁰ Here's Woolf's complex comments on what is meant by “the unity of the mind,” and on Coleridge's suggestion that great minds are androgynous, are very interesting (Woolf, *A Room of One's Own*, 96–9).

³¹ Woolf, *A Room of One's Own*, 84. The new sight here, of course, is friendship (some think more) between professional women. The novelist Mary Carmichael, it is worth noting, is Woolf's fictional creature.

³² Friedrich Nietzsche, “On the Pathos of Truth,” in *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s*, trans. and ed. Daniel Breazale (Atlantic Highlands, NJ: Humanities Press, 1979), 61–6.

because we forfeit the title of originator, but because we feel our hours of work suddenly drained of their claim to necessity.

Finally, in Woolf's incandescent state of mind, as I think we can imaginatively reconstruct it, we feel infinite. The theme of *A Room of One's Own* is of course the fact that, as she writes, "intellectual freedom depends on material things."³³ In this context, she writes about the experience of inheriting a trust that provides enough to live on. The effects take time to sink in. She writes:

in a year or two ... the greatest release of all came, which is freedom to think of things in themselves. That building, for example, do I like it or not? Is that picture beautiful or not? Is that in my opinion a good book or a bad? Indeed my aunt's legacy unveiled the sky to me, and substituted for the large and imposing figure of a gentleman, which Milton recommended for my perpetual adoration, a view of the open sky.³⁴

To have a view of the open sky is to be able, at least mentally, to stretch out in all directions, free from obstacles and 'foreign matter.' We feel ourselves infinite in not being limited by what is; bringing something new into the world is always an experience of transcending the given, and this is an experience of a kind of infinite possibility. Woolf, herself of course a writer, and as it happened one surrounded by artists, writes about what we would most often understand as artistic creativity. But I think the experiences of self she describes belong to anyone who is caught up in working on something, in solving the problems one needs to solve to realize some end, to make something real happen and work, whether the something is a piece of software code or a new porch or a philosophy paper.

Let us turn now to some comments of Adrian Piper's. In introducing a collection of her writings on art, she muses on "the sense of entitlement of an upper-middle-class heterosexual WASP male, the pampered only son of doting parents."³⁵ We understand immediately that Piper is not a huge fan of the pampered only son: her annoyance is there in the description itself. He is "entitled," and she is, she writes, "skeptical of [his] attitude,"³⁶ that is, of his sense of his own normalcy and impartiality. Piper continues:

³³ Woolf, *A Room of One's Own*, 108.

³⁴ Woolf, *A Room of One's Own*, 39. See also the following: "The whole of the mind must lie wide open if we are to get the sense that the writer is communicating his experience with perfect fullness. There must be freedom and there must be peace." (*ibid.*, 104)

³⁵ Adrian Piper, *Out of Order, Out of Sight* (Cambridge, Mass.: Massachusetts Institute of Technology Press, 1996), vol. I, xxxiv.

³⁶ *Ibid.*

To have that sense on this particular planet is to have no bounded sense of self at all, because all of one's subjective tastes, prejudices, and impulses are equated with objective truth. To have this sense is to presume that what one believes and perceives to be the case just simply is the case. One expresses one's thoughts in categorical declaratives without appending intentional operators, as though they communicated simple matters of fact. This manner of expression conveys an exquisite comfort and familiarity with the essential course and character of the universe.³⁷

Piper's subject, her "upper-middle-class heterosexual WASP male, the pampered only son of doting parents," experiences himself as unbounded, his perceptions as objective, his claims as categorical, himself as in harmony with the world. He is not quite a Kantian will: Piper's subject is *not* truly free or rational, though he believes himself to be, precisely because his conception of what is universal is skewed. What is interesting is that despite her annoyance and skepticism at his attitude and sense of himself in the world, Piper writes of this sense, "I prize it too."³⁸ What does she mean? Why would Piper prize this attitude, this self-conception? Because under this sense of self, this self-conception, Piper seems to me to be suggesting, creative work gets done.

The idea, I think (Piper does not elaborate), is that feeling the kind of "comfort and familiarity with the essential course and character of the universe" that her subject feels – feeling unbounded, objective, categorical, and in harmony with the world – is integral to creative work. The claim cannot be that one must *always* feel this way – indeed, anyone who *did* always feel this way would rightly seem a bit creepy. But however positioned one feels, however cramped or lucky in one's position, and however much one's work is informed by one's position, there is still, the claim seems to be, something important and necessary in this set of feelings if one is to work creatively. This seems plausible. Creative activity must reach out, urge on, and transcend the given. My suggestion is that this thought helps us understand the experience of creativity as an experience of oneself, however fleeting, as universal, necessary, and infinite.

By coming at it slantwise, creativity may help light up the value of free rational willing in a useful way. There is something exciting about creation, about innovation, about the human capacity for newness. In difference guises, this excitement has captured the attention and imaginations of many thinkers. Hannah Arendt's emphasis on natality is one example;³⁹ John

³⁷ *Ibid.* ³⁸ *Ibid.* ³⁹ Arendt, *The Human Condition*.

Stuart Mill's emphasis on the value of human experimentation is another.⁴⁰ Concerts, plays, exhibitions, shows, put creativity on display in intentional ways, and part of why we go is to see what people come up with. Great cities move us; the small parts of ingenious machines make us smile; good policies restore faith that reason is operating somewhere. When the creativity is my own, the feeling is quite different but no less great. The expansiveness of which both Woolf and Piper write captures an experience I suspect many will recognize, and that many of us love. It is a feeling of having ideas that go somewhere, that 'flow,' of having what you want to do be the same as what you can do and in fact are doing. And it is, I am suggesting here, a feeling of being freed of external determination, of having power, of an in-gathering of agency, of self-sufficiency; it is contenting or satisfying; one feels coherent, rightly ordered and organized, and, yes, even universal, necessary, and infinite, in the senses in which one can feel these things.

It seems to me important to point out that these experiences *are* commonly valued for their own sakes. One might, I suppose, try to argue for the value of free rational willing (especially when cast as creativity) on utilitarian grounds. One might cite the advances of modern medicine, etc., or argue along Millian lines.⁴¹ But I am with Kant in insisting that we (both do and should) see the free rational activity of the will, whatever its products, as valuable for its own sake. I side here with Kant not out of anti-utilitarian, anti-consequentialist, Kantian orthodoxy, but because I am not convinced, generally, that human life is the kind of thing that reliably gets made better by the products of innovation. And so it is not on utilitarian grounds that we should respect and promote free rational willing and the creativity that comes with it. It must be rather because there is something splendid and admirable about free rational willing and creativity themselves.

Creativity may be a way into thinking about the merits of free rational willing for those who have been suspicious of Kantian autonomy. It seems to me that even those wary of experiences and self-conceptions that involve transcending nature, affirming self-sufficiency, or positioning oneself as universal, necessary, and infinite may find themselves attracted to versions of the same insofar as they are cast as part of an experience of creativity. The question for those of us who are wary in these ways becomes whether we can offer an alternative account that nonetheless allows for what we value in freedom, reason, and creative willful activity. This question is not one

⁴⁰ John Stuart Mill, *On Liberty* [1859] (Indianapolis: Hackett Publishing Company, 1978).

⁴¹ In *On Liberty*, Mill argues that liberty is valuable because it creates happiness and advances the cause of human progress.

I pursue here, but is, I think, the question to be faced in debates about Kantian moral philosophy.

It is time to conclude. I hope here to have presented the beginnings of a plausible account of the feel of Kantian free rational willing, that is, of what the experience of being a free rational Kantian self might be expected to be like. I hope also to have plausibly suggested that this experience bears important relation to an experience of creativity. These seem to me crucial accounts to have in place if we are to think seriously about what we are signing on to if we sign on to Kant's moral theory, and thinking about this, as I indicated earlier, seems to me the thing to do.

*Conclusion: Kant and the goodness
of the good will*

There are times in philosophy, as in life, when the way a problem is framed, the way the alternatives are formulated, makes you feel like your head is going to explode. This was the case for me when, in college, I first encountered John Rawls' effort to separate the political from the moral, and found others echoing, as though it were unproblematic and even helpful, a distinction between 'the right' and 'the good.' The former was meant to be impersonal and somehow suited for public institutions, a matter of ground rules and shared principles, and of duties to which there could not be exceptions, while the latter was best articulated and pursued in ways that were personal and private, an ideal suited to, say, intimate relationships where sensitivity and particularity were called for. What were these people talking about? Connections were drawn to 'liberal neutrality,' to the attractions of a conception of right able to accommodate 'competing conceptions of the good'; those critical of abstract descriptions of right championed Habermas' 'more grounded' approach. But, I thought, isn't the right only right because it is in some important sense good? How could there be a political (public, juridical, institutional) vision that isn't based on (designed to accommodate, realize, and/or protect) some moral conception of human flourishing?

I realize that blame for mapping the terrain of practical philosophy in ways that segregated 'good' and 'right' probably should not all be heaped on Rawls. A significant portion should be laid at the feet of Sir David Ross, author of an influential ordinary-language investigation into moral theory, *The Right and the Good*.¹ And of course I also realize that there was a moment, and there were a set of problems, philosophical and moral and political, which are not exactly ours, which the efforts to hold public values apart from private commitments were meant to address. I don't doubt the good faith, or intelligence, of any of the work, by Ross, or

¹ Sir David Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930).

Rawls, or many others, done in this vein. But still, the idea that there could be rules of rightness that swing in some significant way independent of notions of goodness seemed to me then and seems to me still a *prima facie* impossible compartmentalization.

Some who compartmentalize thus have claimed to be the heirs of Kant's distinction between the form and the matter of practical principles. On this view, full-blown, 'thick,' personal value systems are the 'matter' where bare-bones, 'thin,' basic principles are the 'form'; 'good' is material where 'right' is formal. Such claims are not entirely unjust, or uninteresting, but, as must be clear, I think they lead us significantly astray. The real value and interest in Kant's view, I have tried to argue here, lies in his effort to articulate and philosophically ground a powerful conception of the good.

It is good, I think, for a moral theory to have a conception of the good. Construing Kant's project as primarily formal, or imagining that Kant's motivation stems ultimately from a 'neutral' commitment to rationality or rational agency, are moves that in the end weaken Kant's view, making it more vulnerable to objections and obscuring its real power. When proponents claim that Kantian theory accommodates divergent conceptions of the good, they promote a Kant who lacks any fire of his own, any vision beyond a kind of philosophically principled toleration. When proponents claim that Kantian theory captures norms incumbent on all rational agents, it condescends, consigning dissenters to the realms of the a- or irrational. A characterization of Kant's view as concerned first and foremost with 'the right,' or with duty as articulated by universal, impersonal, law-giving reason, does little to mitigate the caricature of Kant as hyper-rational moralizer, a caricature that still circulates widely. Many don't get as far as they might with Kant because they bristle at the idea that a truly viable morality could ever emerge from something as (seemingly) cold and formal as pure reason – and they end up gravitating toward theories that are Aristotelian, Humean, Hegelian, Romantic, Nietzschean, or something else altogether. The real power of Kant's view – and, I would argue in an effort to keep my head from exploding, the real power of any view – lies in its ideal, its conception of the good.

The powerful conception of the human good, of human flourishing, that Kant articulates ought, he thinks, to inform the ways we treat ourselves, the ways we interact with others, and the ways we organize our collective lives: it is thus at once ethical and political, on any construal, and though Kant makes some principled distinctions between spheres and rules appropriate to them,² *all* spheres of human practice should for Kant be informed by the

² See MS 6:218ff.

same ultimate conception of the good.³ This is an Enlightenment conception, which has human reason, human freedom, and creative human activity at its heart. As readers have by now seen me argue many times, Kant thinks the activity of the free rational will is the most impressive thing in the world: Kant's conception of the good is a conception of the active will animated by respect for free rational will itself. Such a conception of the good is surely strange – it is not a list of particular things (food, shelter, healthcare, education, etc.) that we might find good – but it is, as I have tried to show, a coherent and appealing conception nonetheless.

We can now look at this conception in a way that connects it to several of the broader themes of the Enlightenment, themes Kant developed more fully than any other Enlightenment thinker. To be committed to the value of human reason, in an Enlightenment context, is to value a universal set of resources, possessed by each individual, adequate to adjudicate significant truth claims and to identify essential human ends, without relying on external authority. To value human freedom, in an Enlightenment context, is to value human independence from God and from nature, that is, to value our ability to operate in significant senses independently of the dictates of either. To value creative human activity is to value the production of things – selves, interactions, practices, institutions, etc. – that takes place under the auspices of our own rational freedom. The view is deeply humanist, deeply committed to self-determination, both individual and collective, and deeply committed to realizing a rational world. It is perhaps thus not so strange, given this picture, that a human faculty of desire, which is also rational and free and active, would end up, on careful philosophical inspection, desiring itself, that is, would end up wanting to honor itself above all other things in organizing the energies it directs.⁴

Of course, there are now as there were in Kant's time thinkers who will charge that valuing reason and freedom and creative activity in these ways is to worship false idols. Now, as then, there are what Kant called 'misologists' (see G 4:395), or haters of reason, who think instinct and intuition will serve us better than cold rationality. There is that species of vulgar 'Rousseauianism' that is committed to the idea of the noble savage; there

³ For a defense of this claim, see Jennifer Uleman, "External Freedom in Kant's *Rechtslehre*: Political, Metaphysical," *Philosophy and Phenomenological Research* 68:3 (May 2004), 578–601.

⁴ See the role God's love for himself plays in Leibniz's effort to make sense of the determinants of God's will. The fact that God loves himself helps explain why he always does the right thing, even though he doesn't have to (even though he is free not to); to love himself is, at least in part, to hew to the standards of goodness that are internal to his roles as creator and sustainer of the world. (G. W. Leibniz, "On Freedom and Possibility" [1680–82?], in *Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett Publishing, 1989), 19–23, p. 20.

are Humeans. There are also those today who rail against a host of things Enlightenment rationality is thought to promote: technology that alienates us from nature; calculative casts of mind that squelch warmth and fellow-feeling; specialized knowledge that alienates us from more holistic understanding; logocentrism that obscures productive facts of irrationality and internal contradictions. And now, as then, freedom is suspect in various quarters. Since Hobbes (at least), there are people who worry that freedom will only breed anarchy, that it is at best a good to be weighed against other goods (like security or, in others' minds, social equality). Various philosophers have raised doubts about whether people really do, or should, desire freedom: existentialists identify (though they urge against) a fear of freedom; both left- and right-wing communalisms that oppose atomism and the sovereignty of the individual caution that freedom is an individualistic and divisive ideal; some feminist critiques charge that autonomy ideals elide real, inevitable, and sometimes richly rewarding dependencies. Connected to these hesitations are worries that we overvalue activity, investing in busy-ness and cultivating an almost bullying relationship to reality. Religious traditions that see God as creator bristle at what they see as hubris in our claiming creative prerogative.

The question these objections raise and that faces us is a question about which ways we want to understand ourselves and why. Questions about moral value are questions about what I should *endorse* – not just what I should assent to, or believe, or accept as fact, but what I should *endorse and promote as good*. Facts of course are not irrelevant. But there is a point beyond which they cannot guide us, and beyond which we have room to become more this way or more that way. As I suggested in [Chapter 1](#), moral theories can become self-fulfilling prophecies – we can become, to a significant degree, the hedonists, or the utilitarians, or the free rational Kantian wills we admire. One can appreciate Marx's claims about the material bases of ideology and still attribute the noticeable range of views among people, where material bases are relatively similar, to the effects of ideas. And ideas, concocted and promoted in thought and speech, can account for at least some of the variation one sees in actual social reality: variations in gender equality, rates of poverty, family organization, sexual attitudes and practices, relationships to the natural world (both inert and living), relationships to the human body (e.g., in conceptions of health), beliefs in the divine, attitudes toward fate, commitments to racism and racial stratification, and so on, all vary thanks at least in part to the circulation of ideas. It is fair to think that the ideals we endorse have something to do with the lives we end up leading.

The argument in this book has been that the will that wills itself, that loves itself, and that wants therefore to move in the world in ways that honor itself, doing so just because it wants to, not because it has to, is at the heart of Kant's moral theory – this is Kant's 'strange thing,' and is the thing that awes him. Against those who understand the contours of this strange thing as mere formal, technical requirements posed by the metaphysics of Kantian moral theory, I would argue that Kant's own strategy in trying to persuade us to adopt his view is to hold out this strange thing and with it the promise of certain experiences of ourselves. It is, I think, a good strategy, and we should do it the favor of thoughtful response: we should, in philosophy and in life, look carefully at the kinds of selves and self-understandings Kant proposes, and we should look at some alternatives, and we should choose wisely.

Bibliography

WORKS BY KANT

References to Kant's works are by title initials (from the German original, listed below) and volume and page number from the standard Akademie edition (*Kants gesammelte Schriften*, ed. Königlichen Preußischen [later Deutschen] Akademie der Wissenschaften, Berlin: Georg Reimer [later Walter de Gruyter], 1900). The only exceptions are citations to Kant's *Critique of Pure Reason*, which give page numbers for both 'A' (1781) and 'B' (1787) editions, for example, 'A301/B358' (where a passage occurs in only one edition, only its page will be given, e.g., 'B131'). Translations are from the English editions listed here unless otherwise noted. The square bracketed code that follows publication date gives the Akademie volume number.

- A/B *Kritik der reinen Vernunft* (1781/1787) [Ak. 3, B edition]
Critique of Pure Reason, trans. Paul Guyer and Allan Wood
(Cambridge and New York: Cambridge University Press, 2000).
- ApH *Anthropologie in pragmatischer Hinsicht* (1798) [Ak. 6]
Anthropology from a Pragmatic Point of View, trans. Victor L. Dowdell
(Carbondale and Edwardsville: Southern Illinois University Press,
1978).
- G *Grundlegung zur Metaphysik der Sitten* (1785) [Ak. 4]
Groundwork of the Metaphysics of Morals, trans. Mary Gregor
(Cambridge and New York: Cambridge University Press, 1997).
- KpV *Kritik der praktischen Vernunft* (1788) [Ak. 5]
Critique of Practical Reason, trans. Mary Gregor (Cambridge and New
York: Cambridge University Press, 1997).
- KU *Kritik der Urtheilskraft* (1790) [Ak. 5]
Critique of the Power of Judgment, trans. Paul Guyer and Eric
Matthews (Cambridge and New York: Cambridge University Press,
2000).
- MAM *Muthmaßlicher Anfang der Menschengeschichte* (1786) [Ak. 8]
"Conjectures on the Beginning of Human History," in *Kant: Political
Writings*, ed. Hans Reiss, trans. H. B. Nisbet (Cambridge: Cambridge
University Press, 1970).

- MAN *Metaphysische Anfangsgründe der Naturwissenschaft* (1786) [Ak. 4]
Metaphysical Foundations of Natural Science, trans. James Ellington
 (Indianapolis and New York: Bobbs Merrill, 1970).
- MS *Metaphysik der Sitten* (1797) [Ak. 6]
Metaphysics of Morals, trans. Mary Gregor (Cambridge and New York:
 Cambridge University Press, 1991).
- PPC *Principium primorum cognitionis metaphysicae nova dilucidatio* (1755)
 [Ak. 1].
 “New Elucidation of the First Principles of Metaphysical Cognition”
- R *Reflexionen* (posthumous) [Ak. 19].
Notes and Fragments, ed. Paul Guyer, trans. Curtis Bowman, Paul Guyer,
 and Fred Rauscher (Cambridge and New York: Cambridge University
 Press, 2005).
- REL *Religion innerhalb der Grenzen der bloßen Vernunft* (1793 4) [Ak. 6]
Religion within the Boundaries of Mere Reason, trans. Allen Wood
 and George de Giovanni (Cambridge and New York: Cambridge
 University Press, 1998).
- VE *Vorlesungen über Ethik* (posthumous) [Ak. 27 and 29]
Lectures on Ethics, ed. Peter Heath and J. B. Schneewind, trans. Peter
 Heath (Cambridge and New York: Cambridge University Press, 1997).
- WiA *Beantwortung der Frage: Was ist Aufklärung?* (1784) [Ak. 8]
 “An Answer to the Question: What is Enlightenment?” trans. H. B. Nisbet,
 in *Kant: Political Writings* (2nd enlarged edn.), ed. Hans Reiss (Cambridge
 and New York: Cambridge University Press, 1991), 54–60.

NEWER TITLES OF NOTE

In addition to the many works referred to in notes and listed below, I would urge readers to look at the following titles, which were published as this book was in the final stages of preparation:

- Sedgwick, Sally, *Kant's Groundwork of the Metaphysics of Morals: An Introduction* (Cambridge and New York: Cambridge University Press, 2008).
- Timmerman, Jens, *Kant's Groundwork of the Metaphysics of Morals* (Cambridge and New York: Cambridge University Press, 2007).
- Wood, Allen, *Kantian Ethics* (Cambridge and New York: Cambridge University Press, 2008).

WORKS CITED

- Allison, Henry, *Kant's Theory of Freedom* (Cambridge and New York: Cambridge University Press, 1990).
- Kant's Transcendental Idealism: An Interpretation and Defense* (New Haven: Yale University Press, 1983/2004). (First published in 1983; revised and enlarged version published 2004.)

- Ameriks, Karl, "Kant and Hegel on Freedom: Two New Interpretations," *Inquiry: An Interdisciplinary Journal of Philosophy* 35:2 (1992), 219–32.
- Anderson, R. Lanier, "The Wolffian Paradigm and Its Discontents: Kant's Containment Definition of Analyticity in Historical Context," *Archiv für Geschichte der Philosophie* 87 (2005), 22–74.
- Arendt, Hannah, *The Human Condition* (Chicago: University of Chicago Press, 1958).
- Baier, Annette, *Moral Prejudices: Essays on Ethics* (Cambridge, Mass.: Harvard University Press, 1994).
- 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).
- Bencivenga, Ermanno, "Kant's Sadism," *Philosophy and Literature* 20:1 (1996), 39–40.
- Bittner, Rüdiger, "Maximen," *Akten des IV. Internationalen Kant-Kongresses, Mainz, 1974*, ed. G. Funke and J. Kopper (Berlin: Walter de Gruyter, 1974), 485–98.
- Cummiskey, David, *Kantian Consequentialism* (New York: Oxford University Press, 1996).
- Dear, Peter, *Revolutionizing the Sciences: European Knowledge and Its Ambitions, 1500–1700* (Princeton: Princeton University Press, 2001).
- Descartes, René, *Meditations on First Philosophy with Selections from the Objections and Replies [1641]*, trans. John Cottingham (Cambridge and New York: Cambridge University Press, 1986).
- Engstrom, Stephen, "Allison on Rational Agency," *Inquiry: An Interdisciplinary Journal of Philosophy* 36:4 (1993), 405–18.
- Eze, Emmanuel, "The Color of Reason: The Idea of 'Race' in Kant's Anthropology," in *Anthropology and the German Enlightenment*, ed. Katherine M. Faull (Lewisburg, Pa.: Bucknell University Press, 1995), 196–237.
- Förster, Eckart, ed., *Kant's Transcendental Deductions* (Stanford: Stanford University Press, 1989).
- Forster, Michael, *Hegel's Idea of a Phenomenology of Spirit* (Chicago: University of Chicago Press, 1998).
- Gadamer, Hans Georg, *Truth and Method*, trans. and rev. Joel Weinsheimer and Donald G. Marshall (New York: Crossroad Publishing, 1989). (Originally published in Tübingen as *Wahrheit und Methode*, 1960.)
- Gilligan, Carol, *In a Different Voice: Psychological Theory and Women's Development* (Cambridge, Mass.: Harvard University Press, 1982). (*In A Different Voice* was republished in 1993.)
- Guyer, Paul, "Duty and Inclination," in his *Kant and the Experience of Freedom*, 335–93.
- "Freedom as the Inner Value of the World," in his *Kant on Freedom, Law, and Happiness*, 96–125.
- Kant and the Experience of Freedom* (Cambridge and New York: Cambridge University Press, 1993).

- Kant on Freedom, Law, and Happiness* (Cambridge and New York: Cambridge University Press, 2000).
- “Kant on the Theory and Practice of Autonomy,” in his *Kant’s System of Nature and Freedom*, 115–45.
- “Kant’s Morality of Law and Morality of Freedom,” in his *Kant on Freedom, Law, and Happiness*, 129–71.
- Kant’s System of Nature and Freedom: Selected Essays* (Oxford: Oxford University Press, 2005).
- “Review of Allison’s *Kant’s Theory of Freedom*,” *The Journal of Philosophy* 89:2 (Feb. 1992), 99–110.
- “The Possibility of the Categorical Imperative,” in his *Kant on Freedom, Law, and Happiness*, 172–206.
- “The Strategy of Kant’s *Groundwork*,” in his *Kant on Freedom, Law, and Happiness*, 207–231.
- Hatfield, Gary, *Routledge Philosophy Guidebook to Descartes and the Meditations* (New York: Routledge, 2003).
- Hearne, Vicki, *Adam’s Task: Calling Animals by Name* (Pleasantville, NY: Common Reader Press, 2000. (First published in 1982.))
- Hegel, G. W. F., *Grundlinien der Philosophie des Rechts (Werke 7) [1821]* (Frankfurt: Suhrkamp Verlag, 1970). Translation: *Elements of the Philosophy of Right*, trans. H. B. Nisbet, ed. Allen Wood (Cambridge: Cambridge University Press, 1991).
- Phänomenologie des Geistes (Werke 3) [1807]* (Frankfurt: Suhrkamp Verlag, 1970). Translation: *Phenomenology of Spirit*, trans. A. V. Miller (Oxford: Oxford University Press, 1979).
- Held, Virginia, “Moral Subjects: The Natural and the Normative,” *Proceedings and Addresses of the American Philosophical Association* 76:2 (Nov. 2002), 7–24.
- Held, Virginia, ed., *Justice and Care: Essential Readings in Feminist Ethics* (Boulder, Col.: Westview Press, 1995).
- Herman, Barbara, “Integrity and Impartiality,” in her *The Practice of Moral Judgment*, 23–44.
- “Leaving Deontology Behind,” in her *The Practice of Moral Judgment*, 208–40.
- “Moral Deliberation and the Derivation of Duties,” in her *The Practice of Moral Judgment*, 132–58.
- “Murder and Mayhem,” in her *The Practice of Moral Judgment*, 113–31.
- “Mutual Aid and Respect for Persons,” in her *The Practice of Moral Judgment*, 45–72.
- The Practice of Moral Judgment* (Cambridge, Mass.: Harvard University Press, 1993).
- Hill, Thomas E., Jr., “Humanity as an End in Itself,” *Ethics* 91:1 (Oct. 1980), 84–99. Reprinted in his *Dignity and Practical Reason in Kant’s Moral Theory* (Ithaca: Cornell University Press, 1992), 38–57.
- “The Importance of Autonomy,” in his *Autonomy and Self-Respect* (Cambridge and New York: Cambridge University Press, 1991), 43–51.
- Hobbes, Thomas, *Leviathan [1651]*, ed. Richard Tuck (Cambridge and New York: Cambridge University Press, 1991).

- Höffe, Otfried, "Kants kategorischer Imperativ als Kriterium des Sittlichen," in *Ethik und Politik*, ed. Otfried Höffe (Frankfurt: Suhrkamp Verlag, 1979), 84–119.
- Horkheimer, Max and Theodor Adorno, *Dialectic of Enlightenment: Philosophical Fragments*, ed. G. S. Noerr, trans. E. Jephcott (Stanford: Stanford University Press, 2002). (Translation of vol. v of Max Horkheimer, *Gesammelte Schriften: Dialektik der Aufklärung und Schriften 1940–50*, Frankfurt: S. Fishcher Verlag, 1987.)
- Hume, David, *A Treatise of Human Nature* [1739], ed. L. A. Selby Bigge (Oxford: Clarendon Press, 1978).
- An Enquiry Concerning Human Understanding* [posthumous], ed. L. A. Selby Bigge (Oxford: Clarendon Press, 1975). Hume's first edition of the *Enquiry* was published in 1748; the posthumous text used by Selby Bigge dates from 1777.
- Kittay, Eva, *Love's Labor: Essays on Women, Equality, and Dependency* (New York: Routledge, 1999).
- Korsgaard, Christine, *Creating the Kingdom of Ends* (Cambridge and New York: Cambridge University Press, 1996).
- "Kant's Formula of Humanity," in her *Creating the Kingdom of Ends*, 106–32.
- "Kant's Formula of Universal Law," *Pacific Philosophical Quarterly* 66 (1985), 24–47. Reprinted in her *Creating the Kingdom of Ends*, 77–105.
- The Sources of Normativity*, with G. A. Cohen, Raymond Geuss, Thomas Nagel, and Bernard Williams, ed. Onora O'Neill (Cambridge and New York: Cambridge University Press, 1996).
- Kuhn, Thomas, "Objectivity, Value Judgment, and Theory Choice," in his *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), 320–39.
- Leibniz, G. W., *Discourse on Metaphysics* [1686], in *G. W. Leibniz: Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett Publishing, 1989), 35–68.
- "From the Letter to Des Bosses (1712–16)," in *G. W. Leibniz: Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett Publishing, 1989), 197–206.
- "On Freedom and Possibility (1680–82?)," in *G. W. Leibniz: Philosophical Essays*, trans. Roger Ariew and Daniel Garber (Indianapolis: Hackett Publishing, 1989), 19–23.
- Locke, John, *An Essay Concerning Human Understanding* [1689], ed. Peter H. Nidditch (Oxford: Oxford University Press, 1975).
- Mann, Thomas, *The Magic Mountain*, trans. John E. Woods (New York: Random House, 1996).
- Mendus, Susan, "Kant: 'An Honest But Narrow Minded Bourgeois?'" in *Essays on Kant's Political Philosophy*, ed. Howard Williams (Chicago: University of Chicago Press, 1990), 166–90.
- Mill, John Stuart, *On Liberty* [1859] (Indianapolis: Hackett Publishing Company, 1978).

- Mills, Charles W., "Dark Ontologies: Blacks, Jews, and White Supremacy," in *Autonomy and Community: Readings in Contemporary Kantian Social Philosophy*, ed. Jane Kneller and Sidney Axinn (Albany, NY: SUNY Press, 1998), 131–68.
- Neiman, Susan, *The Unity of Reason: Rereading Kant* (Oxford: Oxford University Press, 1994).
- Nietzsche, Friedrich, "On the Pathos of Truth," in *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s*, trans. and ed. Daniel Breazale (Atlantic Highlands, NJ: Humanities Press, 1979), 61–6.
The Gay Science [1887], trans. Walter Kaufman (New York: Random House, 1974). (Originally published as *Die fröhliche Wissenschaft*, 2nd edn.)
- O'Neill, Onora, "Consistency in Action," in *Universality and Morality: Essays on Ethical Universalizability*, ed. by N. Potter and M. Timmons (Dordrecht: D. Reidel Publishing, 1985), 159–86. Reprinted in her *Constructions of Reason*, 81–104.
Constructions of Reason: Explorations of Kant's Practical Philosophy (Cambridge and New York: Cambridge University Press, 1989).
- "Kant after Virtue," *Inquiry* 26 (1983), 387–405. Reprinted in her *Constructions of Reason*, 145–62.
- Paton, H. J., *Groundwork of the Metaphysic of Morals* (New York: Harper & Row, 1964). This is a re-titled version of Paton's translation of Kant's *Groundwork*, originally published in 1948 as *The Moral Law*.
The Categorical Imperative: A Study in Kant's Moral Philosophy (London: Hutchinson & Co., 1947).
- Piper, Adrian, *Out of Order, Out of Sight* (Cambridge, Mass.: Massachusetts Institute of Technology Press, 1996), vol. I.
- Rachels, James, and Stuart Rachels, *The Elements of Moral Philosophy*, 5th edn. (New York: McGraw Hill, 2006).
- Rawls, John, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971).
"Justice as Fairness: Political, Not Metaphysical," *Philosophy and Public Affairs* 14:3 (1985), 223–51.
Political Liberalism (New York: Columbia University Press, 1996).
- "Themes in Kant's Moral Philosophy," in *Kant's Transcendental Deductions*, ed. Eckart Förster (Stanford: Stanford University Press, 1989), 81–113.
- Reath, Andrews, "Intelligible Character and the Reciprocity Thesis," *Inquiry: An Interdisciplinary Journal of Philosophy* 36:4 (1993), 419–29.
- Ross, Sir David, *The Right and the Good* (Oxford: Clarendon Press, 1930).
- Sandel, Michael, *Liberalism and the Limits of Justice* (Cambridge and New York: Cambridge University Press, 1982). (Reprinted with a new introduction in 1998.)
- Scanlon, Tim, "Kant's *Groundwork*: From Freedom to Moral Community," manuscript, lecture 2, 1983. Cited in Barbara Herman, *The Practice of Moral Judgment*, 138.

- Schiller, Friedrich, *On the Aesthetic Education of Man in a Series of Letters [1794]*, trans. E. Wilkinson and L. Willoughby (Oxford: Oxford University Press, 1967).
- Xenien* [1797], collected in Goethe, *Werke I*, ed. Erich Trunz (Hamburg: Christian Wegner Verlag, 1949). The *Xenien* were published jointly by Goethe and Schiller.
- Schönfeld, Martin, *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford University Press, 2000).
- Schott, Robin May, *Feminist Interpretations of Kant* (University Park, Pa.: Pennsylvania State University Press, 1997).
- Sedgwick, Sally, "Metaphysics and Morality in Kant and Hegel," *Bulletin of the Hegel Society of Great Britain* 37 (1998). (Reprinted in *The Reception of Kant's Critical Philosophy: Fichte, Schelling, and Hegel*, ed. S. Sedgwick [Cambridge and New York: Cambridge University Press, 2002], 306–23.)
- Stumpf, Samuel Enoch and James Fieser, *Socrates to Sartre and Beyond: A History of Philosophy*, 8th edn. (New York: McGraw Hill, 2008).
- Taylor, Charles, *Human Agency and Language: Philosophical Papers, 1* (Cambridge and New York: Cambridge University Press, 1985).
- "Interpretation and the Sciences of Man," in his *Philosophy and the Human Sciences*, 15–57.
- Philosophy and the Human Sciences: Philosophical Papers, 2* (Cambridge and New York: Cambridge University Press, 1985).
- "Self Interpreting Animals," in his *Human Agency and Language*, 45–76.
- The Ethics of Authenticity* (Cambridge, Mass.: Harvard University Press, 1992).
- Uleman, Jennifer, "Everyday Noumena: The Fact and Significance of Ordinary Intelligible Objects," unpublished manuscript.
- "External Freedom in Kant's *Rechtslehre*: Political, Metaphysical," *Philosophy and Phenomenological Research* 68:3 (May 2004), 578–601.
- Weber, Max, *From Max Weber: Essays in Sociology*, trans. H. H. Girth and C. Wright Mills (Oxford: Oxford University Press, 1946).
- Williams, Bernard, *Ethics and the Limits of Philosophy* (Cambridge, Mass.: Harvard University Press, 1985).
- Moral Luck: Philosophical Papers 1973–1980* (Cambridge and New York: Cambridge University Press, 1981).
- "Persons, Character and Morality," in his *Moral Luck*, 1–19.
- Wood, Allen W., *Kant's Ethical Thought* (Cambridge and New York: Cambridge University Press, 1999).
- Woolf, Virginia, *A Room of One's Own* [1929] (New York: Harcourt Brace Jovanovich, 1981).
- To the Lighthouse* [1927] (New York: Harcourt Brace Jovanovich, 1981).
- Zupančič, Alenka, *Ethics of the Real: Kant, Lacan* (London and New York: Verso, 2000).

Index

- a priori, definition, 32, 40
action, Kantian theory of, 31–2
Adorno, Theodor, 148
aid, 128, 133–4
algorithm, no moral, 22, 123
Allison, Henry, 9, 31, 36, 51, 65, 73, 146, 166
Ameriks, Karl, 9
Anderson, R. Lanier, 27
animality, 151
Arendt, Hannah, 151, 171
Aristotle, 19, 112
autonomy, 2, 4, 17, 70–1
 formula of, 135–40
- Baier, Annette, 3, 21, 75, 146, 147
Beck, Lewis White, 31
Bencivenga, Ermanno, 21
Bittner, Rüdiger, 43, 124
bondage, 16
- cards, cheating at, 101
causality
 as always lawful, 30–1, 33
 choice as form of, 30–1
choice, capacity for, 14, 26–31
 animal [*thierische*], 28
 See also Willkür
Churchland, Paul, 75
cognition, as technical term, 94
consciousness
 of laws governing action, 48
 of maxims, 45–6
 of motives, 34–5
consequentialism, 19
creativity, 165–73
Cummiskey, David, 19
- Dear, Peter, 76
Dennett, Daniel, 75
Descartes, René, 10, 109
desire, 25–7, 36–7
- disenchantment, 76
duty, analysis of, 83–7
- effects, not good in themselves, 82
empirical, meaning of, 78, 79
ends, 49–50
 purely rational, 57–9
Engstrom, Stephen, 9
Enlightenment, the, 20, 176–7
eudaemonism, 112
experience, as technical term, 17–18, 93–4
Eze, Emmanuel, 21
- Fieser, James, 3
flowers, 82–3
formalism, 3–4, 112–21
Forster, Michael, 147
freedom, 9–15, 23–4, 29, 50, 63–4, 109, 177–8
 as achievable, 63–4, 67
 as inevitable, 63–6
 lawless, 160
 negative and positive characterizations of,
 67–70
 radical, 65
 transcendental, 65–6
fundamentalism, religious, 20
- Gadamer, Hans-Georg, 9
gecko, 26
Gibbard, Allan, 75
Gilligan, Carol, 21
God, 4, 11, 20, 77, 106, 109, 113, 138, 167, 177
 see also will, God's
Guyer, Paul, 3, 9, 19, 80, 118, 122, 146, 161
- habit, 153
happiness, 54, 81–2, 99–100, 104–7
Hatfield, Gary, 10
Hearne, Vicki, 151
Hegel, G. W. F., 1, 59, 111, 116, 140, 142, 147
Held, Virginia, 21, 75–6, 110, 146

- Herman, Barbara, 3, 36, 43, 47, 65, 122, 123, 124, 152
- Hill, Thomas E., Jr., 125, 159
- Hobbes, Thomas, 10
- Höffe, Otfried, 43, 124
- Horkheimer, Max, 148
- humanism, 19, 20, 166, 167, 177
- humanity, as end in itself, 122–35
- Hume, David, 40, 57, 116
- Hutcheson, Francis, 113
- imperatives, 51–5
 - categorical, 1, 52–4
 - hypothetical, 52–4
- impotence, 16
- infinity, experiences of, 162–5, 170
- interest, rational, 13, 57–8
- justice, 99–100
- Kittay, Eva, 21
- knowledge, as technical term, 94
- Korsgaard, Christine, 3, 122, 127, 146, 166
- Kuhn, Thomas, 5, 58
- laws, 40–1, 69
 - of freedom, 11, 15, 72, 79, 166
 - of nature, 34, 47, 54–5, 72, 79
- Leibniz, G. W., 12–14, 109, 177
- life, 26
- Locke, John, 11
- MacIntyre, Alasdair, 21
- Mann, Thomas, 162–3
- maxims, 41–6
 - as containing laws, 46–8
- Mendus, Susan, 156
- Mill, John Stuart, 171, 172
- Mills, Charles, 21
- money, 18
 - giving to the poor, 123, 126
- monsters, 152
- moral sense theory, 113
- motivation, 98–9
- nature, 79–80
 - as threat to freedom, 10–14
 - see also* Chapter 5 *passim*
- Neiman, Susan, 167
- Nietzsche, Friedrich, 4, 76, 169
- nonsense, 16
- noumena, 11, 71–2, 76, 79, 98
- O'Neill, Onora, 3, 43, 124, 146
- objects, empirical (vs. grounds), 79–80
- Paton, H. J., 88
- perfectionism, 113
- perfections, God's, 16
- phenomena, 11, 71–2, 76, 79
- Piper, Adrian, 167, 168, 170–1, 172
- power, 132, 152–4
- principles, 40
 - objective in strong sense, 52
 - objective vs. subjective, 44, 51
- promises, lying, 127–8, 130, 132–3, 134
- prudence, 100–1
- Rachels, James, 3
- rationalism, 3–4, 21
- Rawls, John, 3, 4, 122, 127, 128, 175
- reason
 - as always informing action, 35–6, 49
 - as threat to freedom, 10–14
 - minimal sense of, 28–9
 - practical vs. theoretical, 33
 - pure practical, 14, 32–3
- Reath, Andrews, 9
- regularity, 158–60
- respect, 92
- responsibility, 16, 103
- Ross, Sir David, 175
- rules, 40
- Sandel, Michael, 21, 146, 147, 148
- Scanlon, Tim, 43, 124
- Schiller, Friedrich, 6, 146, 147, 167
- Schönfeld, Martin, 65, 76
- Schott, Robin May, 21
- Sedgwick, Sally, 147
- self-abuse, 130–2
- self-contentment, 156–8
- self-sufficiency, 154–6
- South Sea Islands, 127
- space, 80
- steward, the self-interested, 101
- Stoics, 113
- Stumpf, Samuel Enoch, 3
- suicide, 124–5, 126, 127
- talents, neglect of, 130–2
- Taylor, Charles, 9, 21, 148
- time, 80
- Uleman, Jennifer, 9, 45, 66, 98, 177
- universality and necessity, experiences of, 160–2, 169–70
- utilitarianism, 19
- voluntarism, 10–13, 89

- Weber, Max, 76
- will, *see* Chapters 2–4 *passim*
as free and rational, 23–4
free, 9–14, 23–4
free will willing itself, 2
God's, 7, 10, 12–14, 23
see also God
- Wille*, 14–15, 17, 31–3
- Williams, Bernard, 3, 46, 146, 147
- Willkür*, 14, 26–31
thierische [animal], 28
see also choice, capacity for
- wish, 30
- Wolff, Christian, 113
- Wood, Allen, 3, 28, 44, 58, 111, 166
- Woolf, Virginia, 167–70
- Zupančič, Alenka, 19, 58