

Cambridge University Press

978-1-107-00851-9 - Groundwork of the Metaphysics of Morals

Immanuel Kant

Frontmatter

[More information](#)

CAMBRIDGE TEXTS IN THE  
HISTORY OF PHILOSOPHY



IMMANUEL KANT  
*Groundwork of the Metaphysics  
of Morals*

Cambridge University Press  
978-1-107-00851-9 - Groundwork of the Metaphysics of Morals  
Immanuel Kant  
Frontmatter  
[More information](#)

---

CAMBRIDGE TEXTS IN THE  
HISTORY OF PHILOSOPHY

*Series editors*

KARL AMERIKS

*Professor of Philosophy, University of Notre Dame*

DESMOND M. CLARKE

*Emeritus Professor of Philosophy, University College Cork*

The main objective of Cambridge Texts in the History of Philosophy is to expand the range, variety, and quality of texts in the history of philosophy which are available in English. The series includes texts by familiar names (such as Descartes and Kant) and also by less well-known authors. Wherever possible, texts are published in complete and unabridged form, and translations are specially commissioned for the series. Each volume contains a critical introduction together with a guide to further reading and any necessary glossaries and textual apparatus. The volumes are designed for student use at undergraduate and postgraduate level, and will be of interest not only to students of philosophy but also to a wider audience of readers in the history of science, the history of theology, and the history of ideas.

*For a list of titles published in the series, please see end of book.*

Cambridge University Press  
978-1-107-00851-9 - Groundwork of the Metaphysics of Morals  
Immanuel Kant  
Frontmatter  
[More information](#)

IMMANUEL KANT

---

*Groundwork of the Metaphysics  
of Morals*

TRANSLATED AND EDITED BY

MARY GREGOR

AND

JENS TIMMERMANN

*University of St. Andrews*

TRANSLATION REVISED BY

JENS TIMMERMANN

WITH AN INTRODUCTION BY

CHRISTINE M. KORSGAARD

*Harvard University*

*Revised Edition*



**CAMBRIDGE**  
UNIVERSITY PRESS

Cambridge University Press  
978-1-107-00851-9 - Groundwork of the Metaphysics of Morals  
Immanuel Kant  
Frontmatter  
[More information](#)

---

**CAMBRIDGE**  
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781107008519](http://www.cambridge.org/9781107008519)

© Cambridge University Press 2012

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

First published 1998 and reprinted eighteen times

Revised edition first published 2012

5th printing 2015

Printed in the United States of America by Sheridan Books, Inc.

*A catalogue record for this publication is available from the British Library*

ISBN 978-1-107-00851-9 Hardback

ISBN 978-1-107-40106-8 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.

## Contents

<i>Preface to the revised edition</i>	<i>page</i> vii
<i>Introduction</i>	ix
<i>Chronology</i>	xxxvii
<i>Further reading</i>	xl
 <i>Groundwork of the Metaphysics of Morals</i>	
Preface	3
First section: Transition from common to philosophical moral rational cognition	9
Second section: Transition from popular moral philosophy to the metaphysics of morals	21
Third section: Transition from the metaphysics of morals to the critique of pure practical reason	56
 <i>Notes</i>	 73
<i>Selected glossary</i>	76
<i>Index</i>	81

Cambridge University Press

978-1-107-00851-9 - Groundwork of the Metaphysics of Morals

Immanuel Kant

Frontmatter

[More information](#)

---

## Preface to the revised edition

The translation of this new edition of Immanuel Kant's *Groundwork of the Metaphysics of Morals* first appeared in my German–English edition (Cambridge, 2011). It is based on Mary Gregor's English version, first published by Cambridge University Press in 1996 and subsequently reprinted in the Cambridge Texts in the History of Philosophy.

Throughout the revision process, care was taken to preserve the familiar feel of Gregor's work. While there were many changes in matters of detail, explained in the introduction and notes to the bilingual edition, the principles of her approach – combining a high degree of faithfulness to Kant's German with readability and fluency – naturally remained intact.

The *Groundwork* was first published in 1785. The translation follows the German text of the German–English volume, which is based on the second original edition of 1786. All major departures of the second original edition from the first are documented in the footnotes of this volume.

For this new edition within the Cambridge Texts in the History of Philosophy, Professor Korsgaard has kindly brought her introduction and her note on further reading up to date to take account of recent developments in Kantian ethics. The notes and the selected glossary have been rewritten. I should like to thank Keith Bustos (St. Andrews) for his work on the revised index.

J.T.

Cambridge University Press

978-1-107-00851-9 - Groundwork of the Metaphysics of Morals

Immanuel Kant

Frontmatter

[More information](#)

---



## Introduction

CHRISTINE M. KORSGAARD

A life devoted to the pursuit of philosophical inquiry may be inwardly as full of drama and event – of obstacle and overcoming, battle and victory, challenge and conquest – as that of any general, politician, or explorer, and yet be outwardly so quiet and routine as to defy biographical narration. Immanuel Kant was born in 1724 in Königsberg, East Prussia, to a Pietist family of modest means.<sup>1</sup> Encouraged by his mother and the family pastor to pursue the career marked out by his intellectual gifts, Kant attended the University of Königsberg, and then worked for a time as a private tutor in the homes of various families in the neighborhood, while pursuing his researches in natural science. Later he got a position as a *Privatdozent*, an unsalaried lecturer who is paid by student fees, at the University. There Kant lectured on logic, metaphysics, ethics, geography, anthropology, mathematics, the foundations of natural science, and physics. In 1770, he finally obtained a regular professorship, the Chair of Logic and Metaphysics, at Königsberg. Destined by limited means and uneven health never to marry or travel, Kant remained in the Königsberg area, a quiet, hardworking scholar and teacher, until his death in 1804.

But sometime in the 1770s – we do not know exactly when – Kant began to work out ideas that were destined to challenge our conception of reason's relationship – and so of our own relationship – to the world around us. Kant himself compared his system to that of Copernicus, which explained the ordering of the heavens by turning them inside out, that is, by removing the earth – the human world – from the center, and

---

*Introduction*

---

making it revolve around the sun instead. Kant's own revolution also turns the world inside out, but in a very different way, for it places humanity back in the center. Kant argued that the rational order which the metaphysician looks for in the world is neither something that we discover through experience, nor something that our reason assures us must be there. Instead, it is something which we human beings impose upon the world – in part through the way we construct our knowledge, but also, in a different way, through our actions.

The implications for moral philosophy, first presented in the *Groundwork of the Metaphysics of Morals*, are profound. The *Groundwork* is an acknowledged philosophical classic, an introduction to one of the most influential accounts of our moral nature which the tradition has ever produced. Some of its central themes – that every human being is an end in himself or herself, not to be used as a mere means by others; that respect for your own humanity finds its fullest expression in respect for that of others; and that morality is freedom, and evil a form of enslavement – have become not only well-established themes in moral philosophy, but part of our moral culture.

But the *Groundwork* owes its popularity to its power, not to its accessibility. Like all of Kant's works, it is a difficult book. It is couched in the technical vocabulary which Kant developed for the presentation of his ideas. It presents us with a single, continuous argument, each of whose steps is itself an argument, and which runs the length of the book. But the particular arguments which make up the whole are sufficiently difficult in themselves that their contribution to the larger argument is easy to lose sight of. The main aim of this introduction will be to provide a kind of road map through the book, by showing how the material presented in each of the main sections contributes to the argument as a whole. First, however, we must situate the project of the *Groundwork* within Kant's general project, and explain some of the basic terminology he employs.

### Kant's philosophical project

Kant was led to his revolutionary views about reason through an investigation of the question "What contribution does pure reason make to our knowledge of the world and to the government of our actions?" The empiricists of Kant's day had claimed that all of our knowledge, as well as

---

*Introduction*

---

our moral ideas, is derived from experience. The more extreme of the rationalists, on the other hand, believed that at least in principle all truths could be derived from self-evident rational principles. And all rationalists believe that at least some important truths, such as the existence of God, the immortality of the soul, and truths about what we ought to do, are either self-evident or can be deductively proved. In order to formulate the issue between these two schools of thought more clearly, Kant employed two distinctions that apply to judgments. Since Kant uses these two distinctions in the *Groundwork* in order to formulate the question he wants to raise about morality, it is necessary for the reader to be acquainted with them.

The first is the analytic/synthetic distinction, which concerns what makes a judgment true or false. A judgment is analytic if the predicate is contained in the concept of the subject. Otherwise, the predicate adds something new to our conception of the subject and the judgment is synthetic. Analytic judgments are, roughly, true by definition: when we say that a moon is a satellite of a planet, we are not reporting the results of an astronomical discovery, but explaining the meaning of a term. The second is the a priori/a posteriori distinction, which concerns the way we know that a judgment is true. A judgment is known a posteriori if it is known from experience, while it is known a priori if our knowledge of it is independent of any particular experience. Putting these two distinctions together yields three possible types of judgment. If a judgment is analytically true, we know this a priori, for we do not need experience to tell us what is contained in our concepts. For this reason, there are no analytic a posteriori judgments. If a judgment is known a posteriori, or from experience, it must be synthetic, for the subject and the predicate are “synthesized” in our experience: we learn from experience that the sky is blue, rather than yellow, because we see that the sky and blueness are joined. The remaining kind of judgment, synthetic a priori, would be one which tells us something new about its subject, and yet which is known independently of experience – on the basis of reasoning alone. If pure reason tells us anything substantial and important, either about the world, or about what we ought to do, then what it tells us will take the form of synthetic a priori judgments. So for Kant, the question whether pure reason can guide us, either in metaphysical speculation or in action, amounts to the question whether and how we can establish any synthetic a priori judgments.<sup>2</sup>

---

*Introduction*

---

The Preface, and the project of the *Groundwork*

We can make these abstract ideas more concrete by turning to the Preface of the *Groundwork*. Here Kant divides philosophy into three parts: *logic*, which applies to all thought; *physics*, which deals with the way the world is; and *ethics*, which deals with what we ought to do. Kant thinks of each of these as a domain of laws: logic deals with the laws of thought; physics, with the laws of nature; and ethics, with what Kant calls the laws of freedom, that is, the laws governing the conduct of free beings. Logic is a domain of pure reason, but physics and ethics each have both a pure and an empirical part. For instance, we learn about particular laws of nature, such as the law that viruses are the cause of colds, from experience. But how do we learn that the world in general behaves in a lawlike way – that every event has a cause?<sup>3</sup> This judgment is not based on experience, for we can have no experience of every possible event. Nor is it an analytic judgment, for it is not part of the concept of an event that it has a cause. If we do know, then, that the world in general behaves in a lawlike way, we must have synthetic a priori knowledge. A body of such knowledge is called a “metaphysics.” If it is true that every event has a cause, then this truth is part of the metaphysics of nature.

That there must be a metaphysics of morals is even more obvious. For morality is concerned with practical questions – not with the way things *are*, but with the way things *ought to be*. Since experience tells us only about the way things are, it cannot by itself provide answers to our practical questions. Moral judgments must therefore be a priori. Yet it is clear that moral laws are not analytic, for if they were, we could settle controversial moral questions simply by analyzing our concepts. So if there are any moral requirements, then there must be a metaphysics of morals, a body of synthetic a priori judgments concerning what we ought to do.

The *Groundwork*, however, is not Kant’s entire metaphysics of morals, but only its most fundamental part. Kant wrote another book under the title *The Metaphysics of Morals*, in which our duties are categorized and expounded in considerable detail. There the reader may learn what conclusions Kant himself thought could be derived from his theory about a wide variety of issues, ranging from questions of personal morality – such as the legitimacy of suicide, the permissibility of using alcohol and drugs, the proper treatment of animals, and the nature and conduct of

---

*Introduction*

---

friendship and marriage – to larger political questions, such as the proper form of the political state, the legitimacy of revolution, and the permissibility of war.

This book is only a *Groundwork*, and its aim is to establish the most preliminary and fundamental point of the subject: that there is a domain of laws applying to our conduct, that there is such a thing as morality. Its aim is, as Kant himself says, “the identification and corroboration of the supreme principle of morality” (4:392).<sup>4</sup> That supreme principle, which Kant calls the *categorical imperative*, commands simply that our actions should have the *form* of moral conduct; that is, that they should be derivable from universal principles. When we act, we are to ask whether the reasons for which we propose to act could be made universal, embodied in a principle. Kant believed that this formal requirement yields substantive constraints on our conduct – not every proposed reason for action can be made universal, and so not every action can be squared with the requirement of acting on principle. We have already seen that the principle that tells us that nature in general behaves in a lawlike way must be synthetic a priori, if it can be established at all. In the same way, Kant thinks, the principle that tells us that *we ought to* behave in a lawlike way must be synthetic a priori, if ethics exists at all. The project of the *Groundwork* is simply to establish that there is a categorical imperative – and so that we have moral obligations.

First section

In each section of the *Groundwork*, Kant carries out a specific project, which in turn forms part of the argument of the whole. In the Preface, Kant says that his project in the first section will be “to take one’s route analytically from common cognition to the determination of its supreme principle” (4:392). In other words, Kant is going to start from our ordinary ways of thinking about morality and analyze them to discover the principle behind them. It is important to keep in mind that because he is analyzing our ordinary views, Kant is not, in this section, trying to *prove* that human beings have obligations. Instead, he is trying to identify *what* it is that he has to establish in order to prove that. What must we show, in order to show that moral obligation is real?

The “common cognition” from which Kant starts his argument is that morally good actions have a special kind of value. A person who does the

---

*Introduction*

---

right thing for the right reason evinces what Kant calls a good will, and the first section opens with the claim that a good will is the only thing to which we attribute “unconditional worth.” The good will is good “just by its willing” (4:394), which means that it is in actions expressive of a good will that we see this special kind of value realized. Kant does not mean that the good will is the only thing we value for its own sake, or as an end. A number of the things which Kant says have only “conditional” value, such as health and happiness, are things obviously valued for their own sakes. Instead, he means that the good will is the only thing which has a value that is completely independent of its relation to other things, which it therefore has in all circumstances, and which cannot be undercut by external conditions.

A scientist may be brilliant at his work, and yet use his gifts for evil ends. A political leader may achieve fine ends, but be ruthless in the cost she is willing to impose on others in order to carry out her plans. A wealthy aesthete may lead a gracious and happy life, and yet be utterly regardless of the plight of less fortunate people around him. The evil ends of the scientist, the ruthlessness of the politician, and the thoughtlessness of the aesthete undercut or at least detract from what we value in them and their lives. But suppose that someone performs a morally worthy action: say, he hurries to the rescue of an endangered enemy, at considerable risk to himself. Many things may go wrong with his action. Perhaps the rescuer fails in his efforts to save his enemy. Perhaps he himself dies in the attempt. Perhaps the attempt was ill judged; we see that it could not have worked and so was a wasted effort. In spite of all this, we cannot withhold our tribute from this action, and from the rescuer as its author. Nothing can detract from the value of such an action, which is independent of “what it effects, or accomplishes” (4:394).<sup>5</sup>

When we attribute unconditional value to an action, it is because we have a certain conception of the motives from which the person acted. If we found out, for instance, that the rescuer had acted only because he hoped he would get a reward, and had no idea that there was any risk involved, we would feel quite differently. So what gives a morally good action its special value is the motivation behind it, the principle on the basis of which it is chosen, or in Kantian terms, willed. This implies that once we know how actions with unconditional value are willed – once we know what principle a person like the rescuer acts on – we will know what makes them morally good. And when we know what makes actions

---

*Introduction*

---

morally good, we will be able to determine *which* actions are morally good, and so to determine what the moral law tells us to do. This is what Kant means when he says he is going to “unravel the concept” of a good will (4:397): that he is going to find out what principle the person of good will acts on, in order to determine what the moral law tells us to do.

In order to do this, Kant says, he is going to focus on a particular category of morally good actions, namely those which are done “from duty.” Duty is the good will operating under “certain subjective limitations and hindrances, which . . . far from concealing it and making it unrecognizable . . . bring it out by contrast and make it shine forth all the more brightly” (4:397). The hindrance Kant has in mind is that the person of whom we say that he acts “from duty” has other motives which, in the absence of duty, would lead him to avoid the action. When such a person does his duty, not otherwise wanting to, we know that the thought of duty alone has been sufficient to produce the action. Looking at this kind of case, where the motive of duty produces an action without any help from other motives, gives us a clearer view of what that motive is.<sup>6</sup>

Kant proceeds to distinguish three kinds of motivation. You may perform an action *from duty*, that is, do it because you think it is the right thing to do. You may perform it from *immediate inclination*, because you want to do it for its own sake, or because you enjoy doing actions of that kind. Or, finally, you may perform an action because you are “impelled to do so by another inclination,” that is, as a means to some further end (4:397). In order to discover what is distinctive about good-willed actions and so what their principle is, Kant invites us to think about the contrast between right actions done from duty and right actions motivated in these other ways. To illustrate this contrast, he provides some examples.

The first one involves a merchant who refrains from overcharging gullible customers, because this gives him a good reputation which helps his business. This is an example of the third kind of motivation – doing what is right, but only as a means to some further end – and Kant mentions it only to lay it aside. The difference between doing the right thing from duty and doing it to promote some other end is obvious, for someone who does the right thing from duty does it for its own sake, and not for any ulterior motive. Yet in order that an action should evince a good will, it is not enough that it should be done for its own sake. This is the point of the other three examples, in which Kant contrasts someone

---

*Introduction*

---

who does an action from immediate inclination with someone who does the same action from duty. For instance, Kant says, there are people

so attuned to compassion that, even without another motivating ground of vanity, or self-interest, they find an inner gratification in spreading joy around them, and can relish the contentment of others, in so far as it is their work. (4:398)

A person like this helps others when they are in need, and, unlike the prudent merchant, but *like* the dutiful person, does so for its own sake. A sympathetic person has no ulterior purpose in helping; he just enjoys “spreading joy around him.” The lesson Kant wants us to draw from this is that the difference between the sympathetic person, and the person who helps from the motive of duty, does not rest in their purposes. They have the same purpose, which is to help others. Yet the sympathetic person’s action does not have the moral worth of the action done from duty. According to Kant, reflection on this fact leads us to see that the moral worth of an action does not lie in its purpose, but rather in the “maxim” on which it is done, that is, the principle on which the agent acts (4:399).

In order to understand these claims it is necessary to understand the psychology behind them: the way that, as Kant sees it, human beings decide to act. According to Kant, our nature presents us with “incentives” which prompt or tempt us to act in certain ways. Among these incentives are the psychological roots of our ordinary desires and inclinations (as sympathy is the root of the desire to help); later, we will learn that moral thoughts – thoughts about what is required of us – also provide us with incentives. These incentives do not operate on us directly as causes of decision and action. Instead, they provide considerations which we take into account when we decide what to do. When you decide to act on an incentive, you “make it your maxim” to act in the way suggested by the incentive. For instance, when you decide to do something simply because you want to, you “make it your maxim” to act as desire prompts.

Kant claims that the difference between the naturally sympathetic person and the dutiful person rests in their maxims. The sympathetic person decides to help because helping is something he enjoys. His maxim, therefore, is to do those things he likes doing. The point here is not that his *purpose* is simply to please himself. His purpose is to help, but he adopts that purpose – he makes it his maxim to pursue that end – because he enjoys helping. The reason his action lacks moral worth is not



### *Introduction*

---

that he *wants* to help only because it *pleases* him. The reason his action lacks moral worth is that he *chooses* to help *only* because he *wants* to: he allows himself to be guided by his desires in the selection of his ends. The person who acts from duty, by contrast, makes it her maxim to help because she conceives helping as something that is required of her. Again we must understand this in the right way. The point is not that her *purpose* is “to do her duty.” Her purpose is to help, but she chooses helping as her purpose *because* she thinks that is what she is required to do: she thinks that the needs of others make a claim on her.

Kant thinks that performing an action because you regard the action or its end as one that is required of you is equivalent to being moved by the thought of the maxim of the action as a kind of law. The dutiful person takes the maxim of helping others to *express* or *embody* a requirement, just as a law does. In Kant’s terminology, she sees the maxim of helping others as having *the form of a law*.<sup>7</sup> When we think that a certain maxim expresses a requirement, or has the form of a law, that thought itself is an incentive to perform the action. Kant calls this incentive “respect for law.”

We now know what gives actions done from duty their special moral worth. They get their moral worth from the fact that the person who does them acts from respect for law. A good person is moved by the thought that his or her maxim has the form of a law. The principle of a good will, therefore, is to do only those actions whose maxims can be conceived as having the form of a law. If there is such a thing as moral obligation – if, as Kant himself says, “duty is not to be as such an empty delusion and a chimerical concept” (4:402) – then we must establish that our wills are governed by this principle: “I ought never to act except in such a way that I could also will that my maxim should become a universal law.”

### Second section

Although the argument of the first section proceeded from our ordinary ideas about morality, and involved the consideration of examples, it is not therefore an empirical argument. The examples do not serve as a kind of data from which conclusions about moral motivation are inductively drawn. Instead, the argument is based on our rational appraisal of the people in the examples, taking the facts about their motivation as given: if these people act from respect for law, as the examples stipulate, then their actions have moral worth. Whether anyone has ever actually acted from

### *Introduction*

---

respect for law is a question about which moral philosophy must remain silent. So demonstrating that the categorical imperative governs our wills is not a matter of showing that we actually act on it. Instead, it is a matter of showing that we act on it in so far as we are rational. A comparison will help here. Showing that the principle of non-contradiction governs our beliefs is not a matter of showing that no one ever in fact holds contradictory beliefs, for people surely do. Nor is it a matter of showing that people are sometimes moved, say, to give up cherished beliefs when they realize those beliefs will embroil them in contradiction. Instead, it is a matter of showing that in so far as they are rational, that's what they do. Kant's project in the second section therefore is to:

trace and distinctly present the practical rational faculty from its general rules of determination up to where there arises from it the concept of duty. (4:412)

In other words, in the second section Kant lays out a theory of practical reason, in which the moral law appears as one of the principles of practical reason.

It is a law of nature, very roughly speaking, that what goes up must come down. Toss this book into the air, and it will obey that law. But it will not, when it reaches its highest point, say to itself, "I ought to go back down now, for gravity requires it." As rational beings, however, we do in this way reflect on, and sometimes even announce to ourselves, the principles on which we act. In Kant's words, we act not merely in accordance with laws, but in accordance with our representations or conceptions of laws (4:412).

Yet we human beings are not perfectly rational, since our desires, fears, and weaknesses may tempt us to act in irrational ways. This opens up the possibility of a gap between the principles upon which we actually act – our maxims or subjective principles – and the objective laws of practical reason. For this reason, we conceive the objective laws of practical reason as imperatives, telling us what we *ought* to do. The theory of practical reason is therefore a theory of imperatives.

Imperatives may be either hypothetical or categorical. A hypothetical imperative tells you that if you will something, you ought also to will something else: for example, if you will to be healthy, then you ought to exercise. That is an imperative of skill, telling you how to achieve some particular end. Kant believes that there are also hypothetical imperatives of prudence, suggesting what we must do given that we all will to be happy.

### *Introduction*

A categorical imperative, by contrast, simply tells us what we ought to do, not on condition that we will something else, but unconditionally.

Kant asks how all these imperatives are “possible” (4:417), that is, how we can establish that they are legitimate requirements of reason, binding on the rational will. He thinks that in the case of hypothetical imperatives the answer is easy. A hypothetical imperative is based on the principle that whoever wills an end, in so far as he is rational, also wills the means to that end. This principle is analytic, since *willing* an end, as opposed to merely wanting it or wishing for it or thinking it would be nice if it were so, is setting yourself to bring it about, to cause it. And setting yourself to cause something just is setting yourself to use the means to it. Since willing the means is conceptually contained in willing the end, if you will an end and yet fail to will the means to that end, you are guilty of a kind of practical contradiction.

Since a categorical imperative is unconditional, however, there is no condition given, like the prior willing of an end, which we can simply analyze to derive the “ought” statement. The categorical imperative must therefore be *synthetic*, so morality depends on the possibility of establishing a *synthetic a priori* practical principle.

#### *The Formula of Universal Law*

Kant does not, however, move immediately to that task; in fact, he will not be in a position to take that up until the third section. The second section, like the first, proceeds “analytically.” Kant is still working towards uncovering *what* we have to prove *in order* to establish that moral requirements really bind our wills. The first step is to analyze the very idea of a categorical imperative in order to see what it “contains.” Kant says:

when I think of a categorical imperative I know at once what it contains. For since besides the law the imperative contains only the necessity of the maxim to conform with this law, whereas the law contains no condition to which it was limited, nothing is left but the universality of a law as such, with which the maxim of the action ought to conform, and it is this conformity alone that the imperative actually represents as necessary. (4:420–421)

This is the sort of thing that makes even practiced readers of Kant gnash their teeth. A rough translation might go like this: the categorical

### *Introduction*

imperative is a law, to which our maxims must conform. But the reason they must do so cannot be that there is some *further* condition they must meet, or some *other* law to which they must conform. For instance, suppose someone proposed that we must keep our promises because it is the will of God that we should do so – the law would then “contain the condition” that our maxims should conform to the will of God. This would yield only a conditional requirement to keep our promises – “if you would obey the will of God, then you must keep your promises” – whereas the categorical imperative must give us an *unconditional* requirement. Since if the imperative is to be categorical there can be no such condition, all that remains is that the categorical imperative should tell us that our maxims themselves must be laws – that is, that they must be universal, that being the characteristic of laws.

There is a simpler way to make this point. What could make it true that we must keep our promises because it is the will of God? That would be true only if it were true that we must indeed obey the will of God, that is, if “obey the will of God” were itself a *categorical* imperative. Conditional requirements give rise to a regress; if there are unconditional requirements, we must at some point arrive at principles on which we are required to act, not because we are commanded to do so by some yet higher law, but because they are laws in themselves. The categorical imperative, in the most general sense, tells us to act on *those* principles, principles which are laws in themselves. Kant continues:

There is, therefore, only a single categorical imperative and it is this:  
*act only according to that maxim through which you can at the same time will that it become a universal law.* (4:421)

Kant next shows us how this principle serves to identify our duties, by showing us that there are maxims which it rules out – maxims which we could not possibly will to become universal laws. He suggests that the way to test whether you can will your maxim as a universal law is by performing a kind of thought experiment, namely, asking whether you could will your maxim to be a law of nature in a world of which you yourself were going to be a part. He illustrates this with four examples, the clearest of which is the second.

A person in financial difficulties is considering “borrowing” money on the strength of a false promise. He needs money, and knows he will get it only if he says to another person, “I promise you I will pay you back next

### *Introduction*

week.” He also knows perfectly well that he will not be able to repay the money when next week comes. His question is whether he can will that the maxim of making a false promise in order to get some money should become a law of nature. Although Kant does not do this, it helps to set out the test in a series of steps.

The first step is to formulate the maxim. In most cases, the person is considering doing a certain act for a certain end, so the basic form of the maxim is “I will do Act-A in order to achieve Purpose-P.” Suppose then that your maxim is:

I will make a false promise in order to get some ready cash.

Next we formulate the corresponding “law of nature.” It would be:

Everyone who needs some ready cash makes a false promise.

At least where duties to others are concerned, Kant’s test may be regarded as a formalization of the familiar moral challenge: “What if everybody did that?” In order to answer this question, you are to imagine a world where everybody does indeed do that. We might call this the “World of the Universalized Maxim.” At this point it is important to notice that Kant says the categorical imperative tells you to act on a maxim which you can *at the same time* will to be a universal law: he means at the same time as you will the maxim itself. So you are to imagine that you are in the World of the Universalized Maxim, seeing whether you can will to act on your maxim in that world. For instance, you imagine that you are asking whether you could will to secure some ready cash by means of a false promise in a world where everyone who needs a little ready cash (tries to) secure it by means of a false promise. In particular, you are asking whether any contradiction arises when you try to do that. Kant, says, in the example at hand, that it does, because:

the universality of a law that everyone, once he believes himself to be in need, could promise whatever he fancies with the intention not to keep it, would make the promise and the end one may pursue with it itself impossible, as no one would believe he was being promised anything but would laugh about any such utterance, as a vain pretense. (4:422)

Why is this a contradiction? This question has attracted an enormous amount of philosophical attention and many interpretations have been

---

*Introduction*

---

proposed. The views that have been suggested may be divided into three broad categories.

Proponents of a logical contradiction interpretation think Kant means there is a straightforward logical contraction in the proposed law of nature. One might argue, for instance, that the universalization of the maxim of false promising would undercut the very practice of making and accepting promises, thus making promises impossible and the maxim literally inconceivable.<sup>8</sup>

Kant's use of teleological language in some of the examples has suggested to proponents of the teleological contradiction interpretation that the contradiction emerges only when the maxim is conceived as a possible teleological law of nature. False promising violates the "natural purpose" of promising, which is to create trust and cooperation, so that a universal law of false promising could not serve as part of a teleological system of natural laws.

According to proponents of the practical contradiction interpretation, the maxim's efficacy in achieving its purpose would be undercut by its universalization. In willing its universalization, therefore, the agent would be guilty of the same sort of practical contradiction that is involved in the violation of a hypothetical imperative. In fact, the maxim in the example is derived from a hypothetical imperative – "if you need some ready cash, you ought to make a false promise" – which in turn is derived from a "law of nature" or "causal law" – namely that false promising is a cause of, and so a means to, the possession of ready cash. In the World of the Universalized Maxim, however, this law no longer obtains. So in willing the World of the Universalized Maxim the agent undercuts the causal law behind the hypothetical imperative from which his own maxim is derived, making his own method of getting the money ineffective. Language supporting all three of these interpretations can be found in Kant's texts, and different interpretations fit different examples better. The problem of finding a single account of the contradiction test that produces the right answers in all cases is one on which Kantians are still at work.

The question is complicated by the fact that Kant himself thinks contradictions may arise in two different ways (4:421, 4:424). In some cases, he says, the maxim cannot even be thought as a universal law of nature: the contradiction is in the very conception of the universalized maxim as a law. The example we have been considering is of that kind: there could not *be* a law that everyone who needs money should make

---

*Introduction*

---

false promises, so the maxim fails what is often called “the contradiction in conception test.” Maxims which fail this test are in violation of strict or perfect duties, particular actions or omissions we owe to particular people, such as the duty to keep a promise, tell the truth, or respect someone’s rights. But there are also maxims which we can conceive as universal laws, but which it would still be contradictory to *will* as laws: these maxims fail what is often called “the contradiction in the will test.” They violate wide or imperfect duties, such as the duty to help others when they are in need, or to make worthwhile use of your talents.<sup>9</sup> Here again, there is disagreement about exactly what the contradiction is. Kant suggests that “all sorts of possible purposes” (4:423) would have to go unfulfilled in a world in which we had neglected our abilities and in which we could not count on the help of others when we are in need. Since rationality commits us to willing the means to our ends, we must will a world in which these most general means – our own abilities and the help of others – would be available to us.

These examples are offered simply as a few illustrations to show how the categorical imperative works to establish the moral status of our actions. Generally, if a maxim passes the categorical imperative test, the action is permissible; if it fails, the action is forbidden, and, in that case, the opposite action or omission is required. The maxims in the examples fail the test, showing, for instance, that making a false promise is forbidden, and that a commitment to helping others when they are in need is required. For a more complete account of what Kant thinks morality requires of us, however, the reader must look to the *Metaphysics of Morals*.

The thought experiment we have just considered shows us *how* to determine whether a maxim can be willed as a universal law, not *why* we should will only maxims that can be universal laws. Kant is not claiming that it is irrational to perform immoral actions because it actually embroils us in contradictions. The contradictions emerge only when we attempt to universalize our maxims, and the question why we must will our maxims as if they were to become universal laws remains to be answered. It is to this question Kant turns next.

*The Formula of Humanity*

We have now seen what the categorical imperative says. In order to show that we actually have unconditional requirements, and so that moral

---

*Introduction*

---

obligation is real, we have to show that this principle is one that necessarily governs our wills. This investigation is in part a motivational one, since no law can truly govern our wills unless we can be motivated by our awareness of its authority. Although Kant denies that we can ever know for certain that someone has been morally motivated, the moral law cannot have authority over our wills unless it is *possible* for us to be motivated by it. But Kant warns us that we cannot appeal to any empirical and contingent sources of motivation when making this argument. As we saw earlier, the sense in which we are trying to show that the moral law governs our wills is not that it actually moves us, either always or sometimes, but that it moves us in so far as we are rational. So the argument must show that the moral law has an authority capable of moving any rational being, and this means it must appeal only to the principles of pure rational psychology.

As rational beings, as Kant said before, we act in accordance with our representations or conceptions of laws. But what inspires us to formulate a maxim or a law (“what serves the will as the objective ground of its self-determination”) is an end (4:427). Whenever we actually decide to take action, it is always with some end in view: either we regard the action as good in itself, or we are doing it as a means to some further end. If there are unconditional requirements, incumbent on all rational beings, then there must be ends that are necessarily shared by all rational beings – objective ends. Are there any such ends?

The ends that we set before ourselves in our ordinary actions, Kant urges, do not have absolute but only relative value: “merely their relation to a particular kind of desiderative faculty of the subject gives them their worth” (4:427). The point here is that most objects of human endeavor get the value that we assign them *from* the way in which they serve our needs, desires, and interests. Just as we value technology because it serves our needs, so we value pure science because we human beings, as Aristotle says, desire to know; we value the visual arts and music because of the way they arouse the human capacity for the disinterested enjoyment of sensory experience; we value literature and philosophy because they serve our thirst for self-understanding, and so forth. Although these other things are not mere means like technology, yet still the value that we assign them is not absolute or intrinsic, but relative to *our* nature. Yet, since we are rational beings, and we do pursue these things, we must think that they really are important, that there is reason



---

*Introduction*

---

to pursue them, that they are good. If their value does not rest in themselves, but rather in the fact that they are important to us, then in pursuing them, we are in effect taking ourselves to be important. In that sense, Kant says, it is a “subjective principle of human actions” that we treat *ourselves* as ends (4:429).

This suggests that the objective end which we need in order to explain why the moral law has authority for us is “the human being, and in general every rational being.” Accordingly, the categorical imperative can now be reformulated as a law instructing us to respect the value of this objective end:

*So act that you use humanity, in your own person as well as in the person of any other, always at the same time as an end, never merely as a means.*  
 (4:429)

Using the same examples he used before, Kant proceeds to demonstrate how this principle can serve as a moral guide. Being of absolute value, human beings should not sacrifice themselves or one another for merely relatively valuable ends. Since it is in so far as we are rational beings that we accord ourselves this absolute value, the formula enjoins us to respect ourselves and each other *as* rational beings. So Kant thinks we should develop our own rational capacities, and promote one another’s rationally chosen ends. Respecting someone as a rational being also means respecting her right to make her own decisions about her own life and actions. This leads to particularly strong injunctions against coercion and deception, since these involve attempts to take other people’s decisions out of their own hands, to manipulate their wills for one’s own ends. Someone who makes you a false promise in order to get some money, for instance, wants you to decide to give him the money. He predicts that you will not decide to give him the money unless he says he will pay it back, and therefore he says that he will pay it back, even though he cannot do so. His decision about what to say to you is entirely determined by what he thinks will *work* to get the result he wants. In that sense he treats your reason, your capacity for making decisions, as if it were merely an instrument for his own use. This is a violation of the respect he owes to you and your humanity.

This example brings out something important about Kant’s conception of morality. What is wrong with the false promiser is not merely that he does not tell the truth. What is wrong with him is the *reason* that he does not tell the truth – because he thinks it will not get the result he wants –

---

*Introduction*

---

and the attitude towards you which that reason embodies. Even if he told you the truth, if it were *only* because he thought it would get the result he wanted, he would *still* be regarding you as a mere means. Instead, we must tell others the truth so that they may exercise their own reason freely. And that means, that in telling them the truth, we are inviting them to reason together with us, to share in our deliberations. When we need the cooperation of others, we must be prepared to give them a voice in the decision about what it is to be done. These ideas lead Kant to a vision of an ideal human community, in which people reason together about what to do. Because this is the community of people who regard themselves and one another as ends in themselves, Kant calls it the kingdom of ends.

*Autonomy and the kingdom of ends*

To be rational is, *formally* speaking, to act on your representation of a law, whatever that law might be; but we have now seen that the content or *material* of the maxims or laws on which we act is given by the value we necessarily set upon our own humanity or rational nature. Putting these two ideas together leads us to a third idea, which is that as rational beings we make the law, we legislate it, for ourselves and each other. Suppose, for instance, I undertake a program of scientific research. I am curious, and wish to know; in treating my curiosity as a reason to undertake the research, I am in effect taking it to be good that I should know. Furthermore, since we have a duty to pursue one another's ends, my decision to pursue scientific research involves a claim on others: that they should recognize the value of my pursuit of this end, should not hinder it, and perhaps, under certain conditions, even offer help with it when I am in need. Thus my choice is an act of legislation: I lay it down, for myself and all others, that this research is a good thing, and shall be pursued. We may say that I *confer a value* upon scientific research, when I choose to pursue it. At the same time, however, the very fact that I make this claim on others whose humanity must be respected serves as a "*limiting condition*" on my own choice (4:431). If the end that I choose, or the means by which I choose to pursue it, is inconsistent with the value of humanity, then I cannot legislate it, and my choice is null and void: my maxim is not a law. Pulling these ideas together leads to what Kant describes as "the *principle* of every human will as a *will universally legislating through all its maxims*" (4:432).

---

*Introduction*

---

This principle, Kant tells us, “would be very *well fitted* to be the categorical imperative” (4:432), because it suggests that the reason we are bound to obey the laws of morality is that we legislate these laws ourselves, that they are our own laws. According to Kant there are two ways in which we may be bound to conform to a law. Sometimes, we conform to a law because of some interest we have that is served by such conformity – for instance, when the law is supported by a sanction. If disobedience to the law will lead to our being fined, socially ostracized, thrown into prison, or dispatched to hell; or if obedience means we will be loved, saved, rewarded, or well-pleasing to God, we may consider ourselves bound to obey it for those reasons. At other times, however, we regard ourselves as bound to obey a law because we endorse the law itself, considered as a law: we think that this is indeed how people in general ought to act, and so we act that way ourselves. Kant calls the first sort of motivation heteronomous, because we are bound to the law by something outside of ourselves – God, the state, or nature – that attaches the sanction to the law. The second kind of motivation is autonomous, because we bind *ourselves* to the law. The principle that we give universal law through our maxims suggests that moral motivation is autonomous.<sup>10</sup>

And on reflection it seems that moral motivation must be autonomous. For if we are motivated to obey a law heteronomously, by a sanction, then the imperative we follow in obeying that law is a hypothetical imperative: *if* you would stay out of prison, or go to heaven, or whatever, *then* you must obey this law. And in that case, of course, the requirement is not unconditional after all. If categorical imperatives exist, then, it must also be true that human beings are capable of autonomous motivation. There can be only one reason why we must do what duty demands, and that is that we demand it of ourselves.

Earlier we saw that according to Kant’s Copernican Revolution, the laws of reason are not something we find in the world, but rather something we human beings impose upon the world. We have now come around to the practical expression of that idea. Kant’s predecessors, he believes, failed to discover the principle of morality, because they looked outside of the human will for the source of obligation, whereas obligation arises from, and so can only be traced to, the human capacity for self-government. Morality, on Kant’s conception, is a kind of metaphysics in practice. We ourselves impose the laws of

### *Introduction*

reason on our actions, and through our actions, on the world, when we act morally.

The principle of autonomy provides us with a third way of formulating the moral law: we should so act that we may think of ourselves as legislating universal laws through our maxims.<sup>11</sup> When we follow this principle we conceive ourselves as legislative citizens in the kingdom of ends. The kingdom of ends may be conceived either as a kind of democratic republic, “the systematic union of several rational beings through common laws” which the citizens make themselves; or as a system of all good ends, “a whole of all ends (of rational beings as ends in themselves, as well as the ends of its own that each of them may set for itself)” (4:433). The laws of the kingdom of ends are the laws of freedom, both because it is the mark of free citizens to make their own laws, and because the content of those laws directs us to respect each citizen’s free use of his or her own reason. The conception of ourselves as legislative citizens is the source of the dignity we accord to human beings, a dignity which Kant, bringing the argument full circle, now equates with the unconditional value of a good will. We now know what gives the good will its unconditional value:

It is nothing less than the *share* it obtains for a rational being *in universal legislation*, by which it makes it fit to be a member of a possible kingdom of ends. (4:435)

But we also now know what we need to do in order to complete the argument. Recall that morality is real if the moral law has authority for our wills. The argument of the second section has yet not shown this, but it has prepared the way, for we now know what has to be true of us if the moral law is to have authority for our wills. We must be autonomous beings, capable of being motivated by the conception of ourselves as legislative citizens in the kingdom of ends, citizens who are bound only by the laws that we give to ourselves. If Kant can show that we are autonomous, he will have shown that we are obligated by the moral law. This is the project of the third section.

### Third section

Up until now, the argument has proceeded “analytically” (4:392). By analyzing our ordinary conception of moral value, and our conception of rational action, we have arrived at an idea of what the moral law says – it says to act only on a maxim you can will as a universal law – and at an idea

---

*Introduction*

---

of the characteristic in virtue of which a person is governed by the moral law – autonomy of the will. To complete the argument, Kant has to show that we and all rational beings really have the kind of autonomous wills for which the moral law is authoritative. This is not an analytic claim, yet if it is to hold for all rational beings it must be an a priori one. When a proposition is synthetic a priori, Kant now tells us, its two terms must be “bound together by their connection with a third thing in which they are both to be found”; that is, it must be deduced (4:447).

Kant opens the third section by making one of the two connections that his argument requires. The will is the causality of a rational being, for our will determines our actions, and it is through our actions that we have effects in the world. If the will’s actions – its choices and decisions – were in turn determined by the laws of nature, then it would not be a *free* will. Suppose that all your choices were determined by a psychological law of nature, say, “a person’s will is always determined by the strength of his desires.” Although you would always do what you most strongly desire, your will would not, according to Kant’s definition, be free. A free person is one whose actions are not determined by any external force, not even by his own desires.

This is merely a negative conception of freedom. But Kant thinks it points us towards a more positive conception of freedom. The will is a cause, and the concept of causality includes the idea of acting according to laws: since we identify something as a cause by observing the regularity of its effects, the idea of a cause which functions randomly is a contradiction. To put it another way, the will is practical reason, and we cannot conceive a practical reason which chooses and acts for no reason. Since reasons are derived from principles, the will must have a principle. A free will must therefore have its own law or principle, which it gives to itself. It must be an autonomous will. But the moral law just is the law of an autonomous will. Kant concludes that “a free will and a will under moral laws are one and the same” (4:447).

Readers are often taken aback by the ease with which Kant draws this conclusion. In the previous section, Kant showed that the authority of morality must be grounded in our autonomy – that moral laws must be laws which we give to ourselves. So any being who is governed by the moral law must be autonomous. But this argument depends on a reciprocal claim that looks at first as if it were stronger – namely, that any autonomous being must be governed by the moral law. Why does Kant think he has shown

---

*Introduction*

---

this? To see why, consider what the categorical imperative, in particular the Formula of Universal Law, says. The Formula of Universal Law tells us to choose a maxim that we can will as a law. The *only* condition that it imposes on our choices is that they have the form of law. *Nothing determines any content for that law; all that it has to be is a law.* As we have just seen, Kant thinks that a will, as a cause, must operate according to a law. If the will is free, then *nothing determines any content for that law; all that it has to be is a law.* What this shows is that the moral law just is the principle of a free will: to have a free will and to operate in accordance with the Formula of Universal Law are, as Kant puts it, “one and the same.”

Freedom and morality are therefore analytically connected. A free will is one governed by the moral law, so if we have free wills, we are governed by the moral law. But do we have free wills? Kant points out that in so far as we are rational, we necessarily act “under the idea of freedom” (4:448). When you act rationally, you take yourself to *choose* your actions, not to be impelled into them, and you think that you could have chosen otherwise. Even if you act on a desire, you do not take the desire to impel you into the action – you think, rather, that you *choose* to satisfy it, that you adopt a maxim of satisfying it. Rational choices are therefore undertaken under a kind of presupposition of freedom. And this being so, Kant proposes, we must, when we make such choices, see ourselves as being bound by the laws of freedom. Rationality requires that we act under the idea of freedom, and freedom is government by the moral law, so rationality requires that we regard ourselves as governed by the moral law. Kant’s argument seems complete.

But Kant is not satisfied with the argument.<sup>12</sup> He complains that the argument does not explain the interest we take in the ideas of morality. He reminds us of a conclusion already established: if we are morally motivated, we cannot be moved by any interest outside of morality, for if we do our duty for the sake of something else, we are acting on a hypothetical, rather than a categorical, imperative. But now Kant points out that we must nevertheless *take an interest* in moral ideas if we are to act on them. This is clearest when morality demands that we do something contrary to our happiness. Here, on the one hand, is something you badly want to do, something on which your happiness depends; but you find, on reflection, that it would be wrong. If you are to be moved by this reflection to refrain from the action, the very thought that you cannot will your maxim as a universal law must be capable of motivating you to refrain from performing the action. You must assign a worth to autonomous action, and to