

The Form of

Practical Knowledge

Stephen Engstrom

A STUDY
OF THE
CATEGORICAL
IMPERATIVE



THE FORM OF
PRACTICAL KNOWLEDGE

THE FORM OF
PRACTICAL KNOWLEDGE

A Study of the Categorical Imperative



STEPHEN ENGSTROM

HARVARD UNIVERSITY PRESS
Cambridge, Massachusetts
London, England
2009

Copyright © 2009 by the President and Fellows
of Harvard College
All rights reserved
Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Engstrom, Stephen P. (Stephen Philip), 1955–
The form of practical knowledge : a study of the categorical
imperative / Stephen Engstrom.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-674-03287-3 (alk. paper)

1. Kant, Immanuel, 1724–1804. I. Title.

B2798.E54 2009

170—dc22 2008039325



Contents

Preface	vii
Note on Citations	xv
I Introduction	1
WILLING AS PRACTICAL KNOWING	23
II The Will and Practical Judgment	25
III Fundamental Practical Judgments: The Wish for Happiness	66
FROM PRESUPPOSITIONS OF JUDGMENT TO THE IDEA OF A CATEGORICAL IMPERATIVE	95
IV The Formal Presuppositions of Practical Judgment	97
V Constraints on Willing	129
INTERPRETATION	147
VI The Categorical Imperative	149
VII Applications	184
VIII Conclusion	241
Epilogue	249
Index	253



Preface

THAT morality is the first-born child of practical reason is an ancient and venerable idea, implicit in ordinary moral consciousness and endorsed, in one form or another, by a long series of eminent philosophers. Already prominent in the thought of Plato and Aristotle, it continues to flourish in later antiquity and carries forward through Aquinas and his followers into modern times, where it receives its fullest expression in the teachings of Kant. A common thread running through them all, it links these figures in a cognitivist tradition in practical philosophy, which finds in knowledge the pattern for life.

Everyone familiar with these great thinkers will know that their doctrines differ in significant respects, exhibiting many variations in emphasis and proportion, in some instances even the appearance of deep-seated opposition. But shining through these differences are certain unmistakable marks and characters revealing membership in a common family line: the arresting insight that all good things depend for their goodness on knowledge and wisdom; the consequent recognition of these latter as basic to virtue; the associated appreciation that the several virtues are bound together in one, giving unity to a virtuous life as a whole, both for an individual and for society; and above all the recognition that the principles of ethical life lie in form and activity. Together these strokes of insight delineate the features of a single countenance, the face of knowledge.

As cognitivist, this tradition stands in opposition to the more skeptical approach of the empiricists, such as Epicurus, Hobbes, and Hume, who subordinate reason to sense and feeling. But in conceiving of reason as having a practical use, it also distinguishes itself, on the other side, from

the physicalist rationalism of the Stoics and their modern followers, who subordinate ethical thought to the exercise of theoretical reason in the scientific knowledge of nature. Indeed, what is particularly striking about this line is its conception of knowledge as divided into two branches, theoretical and practical. Both kinds of knowledge, just as such, share a common form, yet they are specifically different, and it is in its practical use that reason provides the cognitive form immediately suited for living. Virtue is knowledge, but virtue's knowledge is practical rather than theoretical: it is knowledge of how we should live and what we should do.

This tradition offers such a distinguished lineage of philosophical insight that students who are drawn to the most basic questions in ethics and practical philosophy could find no better starting point for their reflections than the works of its principal exponents. This is especially true today. For aside from the fact that the ideas of practical reason and practical knowledge are in themselves no easy topics of investigation, various well-known developments in modern culture have made it increasingly difficult in philosophy to bring these concepts into focus or even to find them at all. Advances in learning and the expansion of commerce have brought with them an increasingly prevalent awareness of the divergences in practice and outlook among societies and of the seemingly intractable disputes often associated with them, and this awareness has commonly stirred up a haze of skepticism and relativism regarding the role of reason in practical affairs. The growth in commerce and education has been accompanied by a torrent of innovation in technology, which has had an extraordinary impact of its own. The wondrous discoveries and achievements in the natural and technical sciences in recent centuries have not only engendered a climate of deference to these fields of knowledge in much of modern and contemporary culture, but have also made them objects of emulation in philosophy, encouraging attempts to study ethics in the fashion of the theoretical rather than the practical sciences, from the outside in rather than from the inside out. All of these developments represent momentous advances in human culture and exert powerful influences on our thinking today. Yet they constitute neither primitive conditions that must be in place in order for practical knowledge, where possible at all, to arise, nor conditions under which the study of it is first possible. Hence they are

not suited to serve as starting points for an account of practical knowledge or as original parameters within which it must be developed. On the contrary, they tend to make it harder to see even the possibility of such a thing as practical knowledge, and to that extent they exacerbate the difficulties we face in gaining a philosophical understanding of ethics as constituting genuine knowledge of what we should do—not to mention that in a corresponding way they can exert a corrosive influence on such knowledge itself, in that, as practical knowledge, it depends from the ground up on an implicit understanding of itself as such. On account of these and other related factors, it is more difficult now than it was in earlier ages to comprehend the idea that morality issues from practical reason itself. To recover and to secure an understanding of it, we must first turn back.

In the teachings of Kant, we find modern philosophy's most systematic and deeply conceived elaboration of this idea. Morality, he holds, is based in a single, though variously formulable, principle of reason, a fundamental law encountered as an imperative in practical consciousness. In advancing this proposition, Kant espouses a doctrine shared by many of his practical-cognitivist predecessors. Aquinas, for example, and prominent school philosophers of Kant's own day, such as Wolff and Baumgarten, also identify a first principle of practical reason. In developing his account of this principle, however, Kant fundamentally rethinks the received conception. According to his predecessors, this principle or law is based on practical reason's apprehension of the good, as what all things seek, and it directs us to do and to pursue good and to avoid what is bad or evil. Kant, in contrast, maintains that this first principle traces not to apprehension of the good, but to practical reason's own autonomy, or self-legislation. Here Kant breaks sharply with the received view that reason's function in the practical sphere is to guide us in action on the basis of its antecedent apprehension of a final end that has an independent footing in nature. Kant rejects this picture, claiming that practical reason's most basic imperatives, those of morality, are categorical rather than hypothetical in form. Human reason must accordingly be conceived as autonomous, as the sole source of its principle of action.

Yet simply to point out that Kant departs in this way from earlier accounts of practical reason's principle does not adequately record the full depth of insight that his reconception is intended to yield. Kant

makes clear that the very idea of autonomy and along with it the notion of a categorical imperative and the doctrine of its several formulas are meant to flow from a proper understanding of practical reason as a source of purely formal requirements on human choice and action, requirements such as the universality and consistency implicated in the bare idea of law. It is from sustained reflection on what reason itself is and on what its practicality must consist in that Kant comes to think it possible—and necessary—to recover practical reason's first principle, and through it morality, without attributing to practical reason any power of apprehension.

In addition to offering a highly developed treatment of practical reason, Kant engages the broader field of approaches in practical philosophy, incorporating in his account the truth in the opposing empiricist and rationalist schools. Though neither of these traditions secures a place in its thinking for reason's practical capacity, they do nevertheless indirectly register the different aspects or faces this capacity displays in practical knowledge. Thus, the stoical rationalists expound a doctrine of natural law, representing such law as existing independently of the conventions of human society yet apprehensible by human reason, whereas more empirically minded modern moralists organize their moral and political views around the idea of a contract, or an agreement between free and independent persons. Recognizing that both of these notions—the ideas of natural law and of social contract—express essential aspects of practical knowledge, Kant develops his account of morality's categorical imperative in such a way as to capture and to display them in their complementary relation.

Thus, Kant's ethics offers a distinctively deep and sustained treatment of the basic ideas and principles of practical philosophy that is also comprehensively developed to incorporate the insights of opposing approaches. Yet Kant's goal of plumbing the ideas of the practical-cognitivist tradition while also taking the measure of the main alternative currents in practical philosophy depends for its attainment on a depth of philosophical penetration that has proved not to be within easy reach for succeeding generations of readers. His own contemporaries misunderstood him from the start, and these early misconceptions laid a foundation for a long history of routine misinterpretation. The standard misreadings have calcified in more recent times, particularly within anglophone philosophy, where

for much of the last century preoccupation with utilitarianism and noncognitivist approaches ensured that comparatively little notice was taken of the practical-cognitivist tradition.

In part, however, the difficulty stems from the fact that the more skeptical mind-set of later generations has helped make Kant's revolutionary doctrines harder to understand even while facilitating their influence. Kant's identification of autonomy as the basis of morality has played a large role in reshaping the way modern philosophers think about reason and morality, and the impact of this transformation has been pervasive. The idea of autonomy has an obvious appeal to the modern and contemporary outlook, which has grown increasingly skeptical of the old assumption that an independent order of value can be found in nature. Yet Kant's claim that morality traces to practical reason alone has equally been a source of puzzlement and doubt, even in sympathetic interpreters. If reason does not apprehend any independently subsisting good, but furnishes merely formal constraints such as universality, how can it be the source of morality's evidently substantive demands? Kant's own explanations are laconic and obscure, yet the problem runs deep, threatening the viability of his entire approach.

Happily, however, Kant's philosophy has received more extensive exploration in recent decades, as part of anglophone analytic philosophy's salutary reengagement with figures from earlier periods. As a result, conditions are now more favorable for gaining both a deeper understanding of the central ideas expounded in his practical philosophy and a better appreciation of the fundamental place they occupy in the tradition to which they belong. We can now see that standard interpretations have exaggerated the extent of Kant's break with his precursors and misunderstood the way in which he reconceives reason's practical use. In fact, Kant never departs from the idea that practical reason is a capacity for knowledge of the good. Indeed, once we see both that and how he not only retains this traditional understanding of practical reason but even directly exploits it in developing his account of the latter, we will be able both to appreciate the full depth of his ethical doctrines and also to gain a systematic understanding of the practical-cognitivist approach that animates practical philosophy's core tradition. To recover this traditional understanding of morality's relation to reason, at the most basic level and in the full articulation it receives in Kant's practical philosophy, is the object of the present study.

As advice to the reader, it may be added that the pursuit of this aim will necessitate certain departures from the path typically followed in investigations of Kant's ethics. Thus, in order to bring into full view the practical-cognitivist character of his account of the categorical imperative, the direct interpretation of that account will be preceded by an extended examination of the ideas basic to his practical philosophy, especially his identification of the will with practical reason. The preliminary examination will comprise two stages. The first will investigate Kant's conception of the will, situating it in relation to the practical-cognitivist tradition. The second will outline a general account of cognition and judgment in order to elucidate the cognitive character of willing. Moreover, in dealing with these preliminary matters it will be necessary to emphasize certain practical-psychological and epistemic concepts and distinctions that are not normally taken up for consideration in standard contemporary interpretations. It will prove crucial, for instance, in highlighting Kant's practical-cognitivist heritage, to explore not only his often discussed distinction between will and choice, but also his less commonly marked contrast between choice and wish. And to articulate the cognitivist character of his ethics, we will need to examine a number of epistemic notions, including those of thought, judgment, and validity, not to mention the idea of practical knowledge itself. For the same reason, it will also be necessary to investigate his conception of knowledge at a high level of generality. Indeed, the reader may find that in places the abstractness of the exposition, coupled with its reliance on ideas not expressible through the contemporary interpreter's standard catalogue of terms, imparts a certain austerity to the argument and situates it in seemingly unfamiliar conceptual terrain. If the study is successful, however, these superficial hindrances to easy comprehension should open the way to a greater illumination in the end.

The plan of this book took shape gradually, over the course of nearly two decades of reflection on practical philosophy, primarily in connection with the study and teaching of Kant. During that time I presented the ideas I was developing in a variety of settings, and I thank the audiences for many helpful comments and suggestions. In addition, I am very grateful to many individuals from whom I have learned over the years in discussions relating to topics treated here, especially Barbara Herman, Thomas Hill, Christine Korsgaard, Andrews Reath, Michael Thompson,

and Allen Wood. I am also particularly grateful to Karin Boxer for suggestions leading to numerous improvements in the final draft. Thanks are due as well to the two anonymous readers for Harvard University Press for their helpful comments. Finally, I owe a special debt of gratitude to the students who over the years have participated in my graduate seminars. My dependence on these teachers, colleagues, and students is complete. Nothing of worth in these pages is my own, and whatever is sound they have helped me to see.



Note on Citations

EXCEPT for page references to the *Critique of Pure Reason*, which use the numbers of the first (A) and second (B) editions, all references to Kant's major writings and lectures are given by abbreviated title—*Anth* (*Anthropology from a Pragmatic Point of View*); *G* (*Groundwork for the Metaphysics of Morals*); *KpV* (*Critique of Practical Reason*); *KU* (*Critique of Judgment*); *L* ([*Jäsche*] *Logic*); *MAN* (*Metaphysical Foundations of Natural Science*); *MS* (*Metaphysics of Morals*); *P* (*Prolegomena to Any Future Metaphysics*); *R* (*Religion within the Limits of Bare Reason*)—and by the page numbers of the appropriate volume of *Kant's gesammelte Schriften, herausgegeben von der Deutschen* [formerly *Königlich Preußischen*] *Akademie der Wissenschaften* (Berlin: de Gruyter [and predecessors], 1902–). Citations of Kant's other writings give, without abbreviated title, the volume and page from this edition. Translations are my own.



And what about this? Is it not apparent that while in the case of the just and the honorable many would choose to do, possess, and think the things that are opined to be so, even if they are not really so; yet when it comes to the good nobody is content to have the things that are opined to be, but rather each seeks the things that really are, and here everyone straightaway disdains mere opinion?

Quite so, he said.

Every soul pursues this and does everything for the sake of it, divining that it is something, but is perplexed and cannot adequately grasp what it is or acquire a stable belief such as it has about other things, and because of this misses even the benefit those other things might provide. Will we say that even the best people in the city, into whose hands we place all things, must be thus in the dark about a thing of this kind and importance?

Least of all, he said.

I suppose, at any rate, said I, that the just and the honorable, when it isn't known in what way they are good, will not have secured a guardian of much worth in someone who is ignorant of this, and I divine that until this is known, no one will understand them adequately.

Plato, *Republic* 505d–506a



I



Introduction

WHEN Socrates propounded his famous paradox, that he knew only that he did not know, he spoke for all of philosophy. Or so we might surmise when we consider that consensus is a touchstone of knowledge. For philosophers commonly confess that if they agree on anything in matters of philosophy, it is that agreement is not to be had. However far they carry their investigations into the principles of knowledge and action, they encounter incomprehension, doubt, and opposition when they attempt to communicate their ideas. Nor do these miscarriages always indicate a deficiency in intelligence or education, or a greater relish for controversy than for the fruits of study. The predicament is as old as philosophy and as familiar as it is inescapable. It is recorded in Plato's depiction of the bewilderment, disbelief, and even hostility and disgust elicited by Socrates' perplexing arguments and assertions. And it is acknowledged by Hume when he likens his abstract speculations on the principles of human nature to the anatomist's hideous portraiture of the hidden organs and tissues of the human body, and when he contrasts such representations with the painter's pleasing and familiar images of its visible outward form. When ordinary life and experience are carved up with the philosopher's instruments of reflection and analysis, what do we ever find but strange systems of cold, lifeless shapes that offend the eye?

Yet Hume's comparison, even if well suited as a characterization of his own theory, does not capture the specific feature of the philosopher's predicament that sets it apart from the situation generally faced in the theoretical sciences. All such sciences begin with familiar, apparent things encountered in experience and seek to make them more intelligible. In

doing so, they commonly bring to light further things, initially unnoticed and often different in kind—perhaps smaller objects hidden within the apparent ones, as in Hume’s example—and investigate the laws determining the behavior and configurations of those other items. But the philosopher is entangled in a peculiar perplexity in that the activities of human thought and action of concern in philosophy are understood from the start to be such that we must already, just in being engaged in them, be somehow acquainted with their principles. If an account of such principles strikes us as strange or contrived, this fact can itself be surprising, even disconcerting; nor will any philosophical explication of them be found satisfactory unless we are able, on reflection, to recognize it as having been in some way familiar to us all along. It would appear, then, that if there is a remedy for this quandary, it must begin in the recognition of this distinctive relation philosophy bears to the principles with which it concerns itself. That recognition must be made to serve as a rule not only for philosophical reflection itself, but also for its communication and interpretation, so that pronouncements in philosophy are both given and received as expressions of the very same common understanding that philosophical reflection presupposes as the truth it aspires to articulate.

1. There are few philosophers, if any, whose writings more strikingly exemplify this distinctive predicament than Kant’s. The primary aim of his philosophy is to expound and to secure the fundamental principle of morality, which he says has its source in reason alone, and which, as the supreme law of our existence, presents itself in our practical lives as a categorical imperative, or unconditional obligation. Kant holds this principle to be one with which we are all intimately familiar and goes so far as to maintain that common human reason “always actually has it before its eyes and employs it as the standard of its judgment” (*G* 403). Yet the famous canonical formulation of it that he sets out for use in moral judgment—his so-called formula of universal law: “act only according to that maxim through which you can at the same time will that it become a universal law”—can hardly be said to have enjoyed the warm general reception by his readership that we might expect would naturally follow upon the discovery of an implicitly familiar standard we tacitly employ in our ordinary moral reflection. On the contrary, it has been the source of much puzzlement and the object of many criticisms. Indeed, even those readers who are most receptive to Kant’s ethics usually look with considerably more favor on his other two famous formulas, which at least have

some power to stir and to inspire the mind. Attention has been drawn especially to the second, the injunction to employ humanity, in one's own person and in the person of every other, always as an end, never merely as a means. For this formula has also seemed more intelligible and more plausibly representable as an articulation of a basic moral principle, whatever the prospects might be of tracing it to a source in reason. Thus, while the formula of humanity has been of central importance in many of the best recent philosophical treatments of Kant's ethics, and while the idea of autonomy, introduced in Kant's third and final formula, has been generally recognized as having a fundamental significance both in his account and as an ethical concept in its own right, the formula of universal law is often set aside as a defective or less adequate expression of his principle.

Much of the criticism of the formula of universal law has stemmed from dissatisfaction with Kant's attempt to illustrate how more familiar duties arise from it when it is applied in particular cases. The difficulties encountered in this connection have contributed in turn to the doubts often raised regarding Kant's assertion that the several formulas all express the very same law, and especially his suggestion that the formula of universal law is equivalent to the seemingly more plausible formula of humanity. But the suspicion under which the formula of universal law so often falls stems most fundamentally from doubts raised against Kant's claim that it is based in practical reason. Here the criticisms commonly take one of two forms. To many, this imperative has appeared to be, in Hegel's words, an "empty formalism," a high-flown and possibly confused expression of the idea that if an action or maxim is right or reasonable for one person, then it is right or reasonable for any other in a relevantly similar situation. Such critics may grant that this formula is an unexceptionable principle of reason, but they hold that by itself it places no significant restriction on our conduct, and that Kant must therefore, in his effort to connect it with familiar duties of morality, supplement it with tacit appeals to other considerations, such as the good or bad consequences that would ensue were the maxim universally followed.¹ Others may allow that this imperative bears some resemblance

¹ This opinion seems to be expressed, for example, in J. S. Mill's remark that "To give any meaning to Kant's principle, the sense put upon it must be, that we ought to shape our conduct by a rule which all rational beings might adopt *with benefit to their collective interest*" (*Utilitarianism*, chap. 5; see also chap. 1). See also Allen W. Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999), chap. 3.

to certain moral precepts, but they see it as reflecting an overly idealized and legalistic vision of human beings as autonomous legislators united as fellow citizens in a cosmopolitan realm of ends. These critics would not dispute that some may find this principle inspiring, but they maintain that it has no discernible basis in reason and hence no bearing on our will except through being in some way imposed on it from without, whether by education, social sanctions, or other familiar means of influencing or shaping moral opinion.² These criticisms reflect different sides of the same difficulty, that of understanding how one and the same principle can be both a formal principle of reason and also a substantive principle of morality.

Yet it is also true, on the other hand, that many have thought it possible to hear in the formula of universal law at least a faint ring of reason and truth, something reminiscent of such commonplace maxims and expressions of ordinary morality as the Golden Rule, or the familiar question "What if everyone did that?" So the formula has continued to hold the attention of philosophers and has inspired repeated attempts to reach a better understanding of the idea it expresses. In view of the difficulty just mentioned, what especially needs to be explained is how this principle can be based in practical reason yet also substantive in its implications.

If such an explanation can be supplied, it may also enable us to address other familiar problems associated with the formula, including the doubts already noted relating to its application and its supposed equivalence with the other two formulas. These additional problems may stem largely from a common source, a misunderstanding of what the formula of universal law asserts, of what it means to will a maxim as a universal law. And we may reasonably presume that if this principle does indeed have the status Kant ascribes to it, then a proper understanding of what it says will not be reached until we can see it as a pure yet morally significant expression of practical reason. Thus, if light can be thrown on the main problem, that of explaining how the formula of universal law is based in practical reason, we may hope it will also yield a reflected il-

²Bernard Williams, for example, asks, "Why should I think of myself as a legislator and—since there is no distinction—at the same time a citizen of a republic governed by these notional laws?" *Ethics and the Limits of Philosophy* (Cambridge, Mass.: Harvard University Press, 1985), 63.

lumination that will help us see both how this principle applies and how the several formulas are related.

Such an explanation is certainly part of what Kant himself undertakes to provide in his exposition of the categorical imperative in the second section of the *Groundwork for the Metaphysics of Morals*. Beginning with a characterization of a rational being as a being with the capacity to act according to principles, or representations of laws, he proceeds to the idea of the categorical imperative as the expression in an imperfectly rational being of the single unconditional principle of reason, the moral law, and then argues that the bare concept of a categorical imperative already entails that such an injunction is precisely captured in the formula of universal law. Like so many of Kant's arguments, however, this line of thought is presented in a highly terse and cryptic fashion. And while various attempts have been made to elucidate this reasoning, it has not been found by many readers, not even among those who are sympathetic to Kant's project, to be perspicuous or compelling.

The chief difficulty has usually been thought to lie in the final step, where Kant reaches his formula on the basis of the claim that the thought of what a categorical imperative of reason requires is just the thought of the necessary conformity of a rational being's maxim to "the universality of a law in general" (G 420–421). Kant appears to slide *from* the merely formal and undisputed principle that a rational being should conform its will to any universal law it can recognize as valid for it as such (where this recognition is informed by its knowledge that what any such law prescribes as rational for one such being it must prescribe as rational for any other comparably situated) *to* the substantive and questionable principle that such a being should act only on maxims it can will as laws holding universally, for all rational beings.³ It is one thing to say that a fully rational being would will to do what it recognizes it

³ As intimated earlier, the idea that Kant's argument faces this sort of difficulty seems to be implicit in Mill's criticism in *Utilitarianism*, chaps. 1, 5. More recent discussion of the problem can be found in Bruce Aune, *Kant's Theory of Morals* (Princeton, N.J.: Princeton University Press, 1979), 28–34, 86–90; Thomas E. Hill, Jr., "Kant's Argument for the Rationality of Moral Conduct," *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca, N.Y.: Cornell University Press, 1992), 121–122; Allen W. Wood, *Hegel's Ethical Thought* (Cambridge: Cambridge University Press, 1990), 163–167, and *Kant's Ethical Thought*, 78–82; and Henry E. Allison, "On a Presumed Gap in the Derivation of the Categorical Imperative," *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy* (Cambridge: Cambridge University Press, 1996).

should, as a rational being, do; it is quite another to say that such a being would also will that all other rational beings do likewise. To highlight the difference between these principles it is sometimes noted that a rational egoist would accept the first, but reject the second. The egoist would allow that a fully rational being would *recognize* that every rational being *ought* to do what it can recognize it should, as a rational being, do, but would not agree that such a being would *will* that every rational being *do* what it can recognize it should, as a rational being, do. In short, Kant's argument trades on an ambiguity in its talk of a maxim's conformity to "the universality of a law in general," confusing willing in accordance with a law with willing a law. If we read these words in a way that preserves a connection with the bare idea of a rational being, we arrive at a principle that seems to fall short of a substantive moral requirement; but if we understand them in a way that fits with the formula of universal law, we are left with a principle whose connection with practical reason seems open to question.

2. The fact that this objection has often been raised does not show, of course, that Kant's derivation does not deserve further consideration. It suggests, however, that it would be more worth our while to step back from the details to examine some of the fundamental ideas that guide his thinking, ideas on which his argument may tacitly rely. Obviously, the notion of practical reason is of particular interest in this connection. Even if never fully articulated by Kant himself, this idea is basic to his entire practical philosophy, and it provides the natural starting point from which to consider his claim that the categorical imperative expresses a fundamental principle of practical reason.

In seeking to recover an understanding of practical reason, however, it will be important, especially now that we have mentioned the egoist, to keep the following points in view. Although the objection just described may sharpen our appreciation of the difficulties confronting any attempt to explain how the formula of universal law is based in practical reason, we are not bound from the outset to accept its implicit assumption that to be successful such an explanation must yield an argument that could move any rational but ostensibly amoral agent to embrace morality and take up the pursuit of virtue. We need not suppose that the problem must be one of linking the formula to an attenuated conception of practical reason, as a capacity to find means to given

ends, say, or to specify actions that best serve one's interests, a capacity that would be unproblematically attributable to the egoist or the amoralist. It may be sufficient to show that this formula is connected with a conception that accommodates the possibility of full-fledged objectivity in practical thought, provided that such a conception is one in which we can see our practical thinking to be rooted.

*Cognition →
Knowledge*

The account to be developed here begins with the idea that morality is a type of cognition, a type of knowledge. As Kant observes, this knowledge is rational, not empirical, and practical, not theoretical. As rational, it lies in judgments made by reason, the cognitive capacity, rather than in feelings or in the workings of the senses. As practical, it has a certain efficacy in respect of its object, representing the latter as "what ought to be" rather than as "what is." On account of this efficacy, Kant is able to identify practical reason, the capacity for such cognition, with the will (G 412).

That Kant takes morality to consist in such cognition—practical rational cognition—is a fact whose evidence pervades his writings on the subject, though it has received relatively little attention from his expositors and critics. Indeed, it is already apparent in the opening paragraphs of the Preface to the *Groundwork*, where he explicitly marks out the position morality occupies in a taxonomy of rational cognition. Considered in respect of its genus, the rational cognition in which morality consists is "material" in that, like theoretical knowledge of nature, it concerns itself with objects and the laws to which they are subject. In this respect it differs from logic, which, being concerned only with the universal rules of thinking in general, is a merely formal science, having no proper subject matter. (Thus, it is no part of Kant's conception of morality that its rationality implies that it is an "empty formalism" in the sense in which logic might be said to be.) Specifically, morality belongs to the practical rather than the theoretical branch of material rational knowledge. In contrast to theoretical cognition, which determines laws "according to which everything happens," morality determines laws "according to which everything ought to happen." The objects of practical rational cognition are later identified in the *Critique of Practical Reason*, where we are told that the sole objects of practical reason are those of the good and the bad. The practical laws to which these objects are subject, though often called laws of freedom, are accordingly describable as laws "of the good" (G 414).

If the substantive (material) exercise of reason, the cognitive faculty in general, lies in the judgments that make up our cognition of objects, then we can say, similarly but more particularly, that the exercise of practical reason, the faculty of practical knowledge, lies in the practical judgments that make up our knowledge of the good, or “what ought to be.” Consideration of these practical judgments will reveal that they share a certain form on account of being exercises of the faculty of practical knowledge, a form that finds expression in the formula of universal law. In this way, the formula can be traced to practical reason. The idea of practical knowledge will provide the middle term, so to speak, enabling us to see the connection.

The elaboration of this line of thought in the pages to follow will contain no argument for the strong thesis that the capacity to recognize oneself to be bound by the categorical imperative, as expressed in the formula of universal law, must belong to any conceivable rational being. Kant himself, as we shall see, regards this proposition as beyond proof. It will, however, be argued that this capacity belongs to any rational being who has the capacity to know the good, or to know what it would be good to do, where the concept of the good that figures in this knowledge, being practical cognition’s formal representation of its own object, is a nonrelational concept that applies not merely to means, but also, and in the first instance, to ends. Though this argument will not show there to be a connection between morality and the most minimal conception of rationality we might frame for ourselves, a connection that would provide a means of convincing any conceivable moral skeptic amenable to argument, it will nevertheless bring to light a connection between the categorical imperative, as expressed in the formula of universal law, and practical reason, conceived as a faculty of practical knowledge. Such an elucidation will furnish a sufficient explanation of how morality is based in practical reason, provided that we can already recognize practical reason, so conceived, to be operative in our practical life through the judgments we make employing the concept of the good, and provided further that morality, or at least some system of duties we can regard as articulating its rational content, can be seen as based in the principle expressed in the formula of universal law. (These two provisos will not be explored in depth in this study, but a few points relevant to the first will be noted in the discussion of practical judgment (§II.7), and the second will receive support from the examination of how the formula of universal law is re-

two
provisos

lated to the formulas of humanity and autonomy (§VI.2, 6–7) and from the illustrations of how it applies in some particular cases (§VII.)

3. By developing the idea that practical reason is the capacity to know the good, the account to be presented here should offer a further significant advantage. In addition to showing how the principle expressed in the formula of universal law is connected with practical reason, it will explain how this imperative is related to the good. The importance of such an explanation is not difficult to appreciate. Since Kant's primary objective in the *Groundwork* is to show morality's source in reason, this work does not much concern itself with articulating morality's relation to the good. Yet Kant's way of pursuing that primary objective has given many the impression that he sees no connection at all between morality and the good. He prominently contrasts morality and happiness, at times even setting them in opposition to one another, and he emphasizes that the motive figuring in action done from duty does not depend on the consideration of any good that may result from conduct conforming to moral requirements. It can thus easily appear that Kant denies all connection between morality and the good, or, worse still, maintains that they are in fundamental opposition.⁴ Such an appearance can in turn bolster the impression that Kant's moral philosophy offers very little, to say the least, in the way of encouragement and support for the cause of morality and virtue, despite Kant's own acknowledgment that a primary aim of practical philosophy is to help strengthen the moral motive, or, as he puts it, "to secure acceptance and durability" for morality's principle (*G* 405). If we agree with Plato that while many people prefer the reputation for justice and morality over the genuine possession of these qualities, everyone agrees in prizing reality over appearance when it comes to the good itself, then Kant's moral philosophy can seem to be at best a dubious ally of morality,

⁴This is not to say that the significance of the concept of the good (or the closely related concept of an end) within Kant's moral philosophy has been altogether overlooked. Numerous interpreters have emphasized its importance. Two prominent recent examples are Christine M. Korsgaard and Barbara Herman; see in particular Korsgaard's "Two Distinctions in Goodness," *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), and Herman's "Leaving Deontology Behind," *The Practice of Moral Judgment* (Cambridge, Mass.: Harvard University Press, 1993). This interpretive approach is not without critics, however; see for instance J. B. Schneewind, "Kant and Stoic Ethics," in S. Engstrom and J. Whiting, eds., *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty* (Cambridge: Cambridge University Press, 1996).

bound to thwart its own avowed purpose. Or to put the thought in the terms of Hume's comparison, if Kant's anatomy of morality is accurate, the painter's task would seem to be impossible.

If the account to be developed here is successful, however, it will reveal that our interest in knowing what the good is and where it lies already contains within itself an implicit recognition that the principle expressed in the formula of universal law is the form of all knowledge of the good. And if we can show that this principle stands in such a relation to our knowledge of the good, a relation that reveals in turn the way in which it is related to the good itself, we can demonstrate how practical philosophy can, in at least one significant way, further the aim Kant attributes to it of securing "acceptance and durability" for this principle. For by bringing this relation to the good into view, we can remove the appearance that, as a purely formal principle of reason, it must be unconnected with the good or even opposed to it and hence lacking in authority and power to determine our will, yet do so in a way that does not conflict with Kant's well-known insistence that neither our recognition of the normative authority of the moral principle nor the virtuous person's motive in observing it can depend on the consideration of any good that may result from such observance.

Obviously an account developed along the lines just indicated will have to include, as a central part, a treatment of practical judgment. And because practical judgment consists in the employment of the concept of the good, much of the discussion to follow will concern this concept and the relation in which the formula of universal law stands to it. Indeed, we can find a convenient and fruitful way into our topic if we begin by noting the intimate connection Kant sees between morality and the idea of action to which goodness or badness is *intrinsic*. Though this connection is often overlooked in treatments of Kant's ethics and in discussions of the formula of universal law in particular, an appreciation of it can throw considerable light on questions that will engage us in what follows; it will accordingly provide one of the guiding threads of the interpretation to be outlined below. Bearing in mind the connection Kant sees between morality and intrinsic goodness and badness—a connection we can succinctly express by saying that morality consists in the knowledge of intrinsic goodness—will enable us to provide an account of the formula of universal law that brings its purpose more clearly into view and allows us to avoid problems readers commonly face when at-

tempting to understand how it applies in particular cases and how the “test” it contains is to be carried out.

4. Before turning to the connection just described, however, we need to address a concern the foregoing remarks may already have brought to mind. The worry springs from the following considerations. One of the main attractions of Kant’s ethics is that his conception of morality as originating in the autonomy of the will makes him seem, at least in this one important respect, more modern in spirit than do the other great moral philosophers from his day and before.⁵ The idea that we are the legislators of our own principles not only suggests an appealing image of ourselves as enlightened authors of our own destiny, but also frees us from the difficulties we would face if we followed the lead of other traditional moral philosophers and supposed that our true ends and directions for living are to be taken from some external source or authority—from nature, say, or the cosmos, or divine revelation. In pursuing the latter course we would run afoul of Kant’s own compelling arguments that practical guidance cannot be derived from such sources. We would leave ourselves confronted by the familiar difficult questions of how these directions could be understood to have their footing in any such external setting, given the scientific theoretical knowledge we now have of nature and the physical universe, and of how we could ever know them if they were somehow there outside us, written into things. Yet the proposal just broached—that our reading of Kant should be guided by the thought that practical reason is the capacity to know the good, and by the idea that morality is knowledge of intrinsic goodness—threatens to return us to just such an antiquated traditional conception. To read Kant through the lens of these suppositions would be to impose on him the view that there is a teleological order in nature that we can somehow rationally apprehend and that provides us with directions for living—a view we today can look upon only with suspicion.⁶

While detailed examination of these metaphysical issues lies beyond the scope of this study, it will suffice for our purposes here to note

⁵Not everyone, of course, has looked with favor on Kant’s ostensibly modern notion of the autonomy of the will. Dissenters include G. E. M. Anscombe (see “Modern Moral Philosophy” and “Authority in Morals,” reprinted in her *Ethics, Politics, and Religion* [Minneapolis: University of Minnesota Press, 1981]) and Iris Murdoch (see *The Sovereignty of Good* [New York: Schocken Books, 1970], 79–81).

⁶Cf. J. L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin Books, 1977), 38ff.

the following points in response. Any suggestion that goodness or badness can inhere in things—even if the things in question are persons, the activities they enjoy, and the actions they perform—will likely seem strange if it is imagined that these qualities would be originally objects of theoretical rather than practical knowledge. And it will seem especially strange if it is assumed that they would have to be like sensible qualities, yet apprehensible neither by the senses nor by our capacity to feel pleasure and pain, but only by reason—not to mention that such an assumption tacitly regards reason as a passive, receptive capacity and thereby introduces further puzzlement as to how the rational apprehension of such qualities could itself move us to act. It was from a picture of this sort that Hume was understandably recoiling when he denied that such qualities are truly in things, likening them instead to colors with which we gild or stain the latter. But no such view of these qualities is entailed by the interpretation here proposed.

Kant takes the concepts of good and bad to be *a priori* concepts of practical reason (specifically, they are the pure categories of reality and negation, practically employed). To say that these concepts are *a priori* is, in part, to say that they are entirely formal in character (as we shall see, they are based, respectively, in the reflectively employed concepts of agreement and conflict). It is accordingly to say that, unlike our representations of sensible qualities, such as colors and sounds, or our feelings of pleasure and pain, they are not themselves derived from the actual observation or experience of already existing things, even though they do of course apply to things it is possible for us to experience, and even though there is nothing else to which we *could* apply them, nor any way for them to figure in determinate knowledge of the good or the bad without empirically given representations that supply material for these *a priori* forms. To say that these concepts are concepts of practical reason is, in part, to say that they are practical representations, that in their use they have an efficacy whereby they work to bring their objects into existence or to maintain them in existence. To the extent that these concepts have the characteristics just indicated, there is nothing about them that prevents us from employing them in judgments whereby we have knowledge—indeed, efficacious knowledge—of the good, where among the things we can thereby know to be good are existing things, and where the goodness itself, being a form that is brought into existence or kept in existence through the concept's efficacy, is in those existing

things themselves, as their formal constitution. Usefulness, for example, which is a type of goodness, exists through our knowledge of it. An artisan's tool has a specific mode of usefulness, a specific function, present in it as its essential form, and the technical-practical concept, or knowledge, of this form not only governs the tool's initial production but also maintains the form in existence through the care and the skill with which the artisan uses the tool and keeps it in good repair.

It should be clear enough from these points that **the above suggestion that we can have practical knowledge of the good does not imply that the good must be independent from us in a way that leaves us with no readily comprehensible route of cognitive access to it.** Everything we can know to be good, even what we can know to be intrinsically good, stands in an internal relation—a relation of essential knowability—to practical reason, our capacity to know it. This relation is implied by the idea that the objects of practical reason necessarily conform to the cognition gained through its exercise, an idea that in turn is the practical application of Kant's famous "Copernican" proposition that objects, or things we can know, must conform to our cognition.⁷ Though this proposition may at first seem paradoxical as a statement concerning theoretical knowledge, it can hardly occasion a comparable puzzlement where the knowledge in question is understood to be practical. For the objects of such cognition are conceived as things whose very existence depends on the knowledge of them.

The peculiar puzzlement that can arise in connection with practical knowledge concerns, not how the object can conform to the cognition of it, but how the cognition can bring its object into existence, how knowledge can be practical, or efficacious. Kant addresses this latter question in the final section of the *Groundwork*, where he asks how a categorical imperative is possible, or how pure reason can be practical, and he revisits the issue in the *Critique of Practical Reason*. A full examination of this question lies outside our concern here (cf. §III.7). All that will be required for our present purposes is that we conceive of reason in a way that does not gratuitously close off the possibility of its practicality. We can do

⁷Thus the earlier characterization of practical reason as the capacity to know the good should not be understood as implying that rationality is secondary to goodness or value. To respect the proper order of concepts, we must say, not that the rational is good, but that the good is rational.

this by explicating this cognitive power along the lines already articulated in Kant's theoretical philosophy, namely through the appreciation, available in cognition's own self-consciousness, of the spontaneity characteristic of cognition in general. For the one great impediment to understanding how knowledge can be practical is the assumption that reason, the cognitive capacity itself, is receptive in nature and hence passive in operation—the assumption, that is, that reason's combination of representations in cognitive judgments depends on its being somehow, directly or indirectly, determined from without, by conditions external to its own conscious activity, to combine them in the way that it does. Such an assumption is part and parcel of an exclusively empiricist conception of cognition, and so long as it is in place, it will seem natural, as it did to Hume, to say that "reason is perfectly inert," that only something separate from it can ever move us to act. Our account of practical judgment and cognition will accordingly be developed from a conception of reason as spontaneity. Reliance on this conception constitutes no special assumption, since the spontaneity of reason is already implicated in the reflective self-awareness characteristic of cognition in general.

In the pages to follow, then, the concept of the good will be identified with the concept of the object of practical knowledge. "Good" will be taken to signify what can be practically cognized. For it to be good to do something is just for it to be practically knowable that the thing in question is to be done. Just as "the true" and "the real" can be used to pick out the object of theoretical knowledge, so "the good" refers to the object of practical knowledge. And in keeping with the "Copernican" understanding of the connection between knowledge and its object, we shall understand the concept of the good through the concept of an object of practical knowledge and say nothing about the good in general beyond what can be seen to follow from the consideration that we can have practical knowledge of it. We can see, for instance, that goods must be compatible, since knowledge is consistent with knowledge. Similarly, the conceptual connection between the capacity to recognize a certain action to be something it would be good to do and the capacity to be moved by that recognition to do it (or perhaps to help another to do it) reflects the efficacy of practical knowledge, whereby it works to bring its object into existence or to maintain it in existence.

the good

5. It is readily apparent that Kant regards a good will, and the action expressing a good will, as intrinsically good. In the famous argument with which he begins the *Groundwork*, he sets the good will apart from all other ostensibly good things by noting that while the latter are good only under certain conditions (since they can be bad if improperly used), the good will has an “inner, unconditioned worth” (G 394). And in the ensuing, much discussed account of the “moral worth” distinctive of action in which a good will is expressed, such worth too is said to be “inner” and “unconditioned” (G 397, 400). The closely connected ways in which Kant employs these terms reflect his recognition that inner goodness is unconditioned. Something has *inner* goodness if it is good *in itself*. That is to say, it has such goodness if the consideration of it by itself alone, and so without regard to its circumstances and effects, suffices for the recognition that it is good. Anything that can be known to be good from such a consideration, however, cannot be merely good in certain respects (even something bad can be good in certain respects), nor can it be merely good where circumstances permit or where other things are equal; it must be good in all respects, that is, unconditionally, or absolutely, and hence in all possible circumstances (cf. A324–326/B380–382). For the same reason, what has inner goodness must also be distinguished from things that are good on account of their effects. It is “not by its fitness to attain some proposed end” that a good will is good (G 394). This last contrast, as applied to action, reappears later in Kant’s well-known distinction among imperatives: an imperative is hypothetical if it represents an action as “good merely as a means *to something else*,” but categorical if the action is “represented as *in itself* good” (G 414; cf. *KpV* 62).

Though less often remarked upon, it is equally true—and no surprise, given his claim that a good will’s morally worthy action is intrinsically good—that Kant regards moral wrongdoing as intrinsically bad. This view is not explicitly asserted in the *Groundwork*,⁸ but there are many indications in that work that he holds it. At one point, for example, he says in effect that if an action is intrinsically bad, then the imperative that prohibits it is categorical, a requirement of morality (G 419). And there is good reason to think he would affirm the converse

⁸ It is fully explicit in *Religion within the Limits of Bare Reason*, however: “Only what is contrary to the moral law is in itself bad [*an sich selbst böse*]” (R 58).

as well. For it seems evident that if something is categorically prohibited by reason, then it must be bad regardless of circumstances. Indeed, Kant comes close to asserting this when at another point he says, “if I deviate from the principle of duty, this is most certainly bad” (*G* 402). Another sign is that he takes the capacity to distinguish what is morally right from what is morally wrong to be shared by all persons and hence to be independent of any special fund of knowledge about the world and any special talent or skill in calculation or deliberation (*G* 403, *KpV* 36–37). For it would be difficult to see how this capacity could have such independence if one supposed that distinguishing whether an action is morally right or wrong depends on considering how it stands to facts or circumstances that are related to it externally, or contingently.

A further indication that Kant supposes moral wrongdoing to be intrinsically bad is that he takes it to involve an internal inconsistency. This point is of particular interest for our purpose in this study, since it promises to throw light on the idea of intrinsic badness and should help bring the formal character of the concept of badness more fully into view. Imperatives of each of the two sorts Kant distinguishes—hypothetical and categorical—appear to be, in some sense, demands of consistency placed on willing, so that willing an action that conflicts with an imperative always involves a contradiction. But whereas the contradiction that arises in willing an action that conflicts with a hypothetical imperative lies *between* what one wills to do in willing the action and what, in the light of the circumstances and the limits on one’s productive powers, one recognizes it to be necessary to do in order to realize something else one happens to will, willing an action that conflicts with a categorical imperative involves a contradiction that lies somehow *within* the willing of that very action itself, within the maxim, or principle of willing, that the action expresses.

6. It might initially seem doubtful whether an account of moral wrongdoing that traces it to an inner contradiction provides a helpful elucidation of the idea that such wrongdoing is intrinsically bad. Speaking, as we were just now, of a contradiction within the willing of a certain action might appear to be tantamount to speaking of a contradiction within the *content* of the will or the maxim, an inconsistency in what the agent wills to do. But it is hardly credible that in the case of impermissible action the content of the will conflicts with itself. Making a false

promise seems obviously not to involve any inconsistency in what one wills just in willing such action, nor does there seem to be any incoherence just in refusing to help others in need. Indeed, if the contradiction is supposed to lie in the content, then Kant's account of impermissible action turns out, absurdly, to entail such action's absolute impossibility. The implication would be that such action is "unthinkable"—not just in the sense that the thought of performing it would never occur to a virtuous person, but in the sense that there is nothing there for anyone to think at all.

This criticism suffers from a glaring omission, of course, which might at first glance seem to leave open an obvious response. For the contradiction Kant has in view is supposed to result through the application of a certain test or procedure of universalization carried out in thought or in the will.⁹ His suggestion is not that the contradiction lies in the rule of action making up a particular maxim's content, but that it lies in the willing of the rule as a universal law. "I could indeed will a lie," Kant says, "but by no means a universal law of lying" (*G* 403).

This reply is not entirely satisfactory, however, for it invites an obvious rejoinder. If Kant means to elucidate the idea that moral wrongdoing is intrinsically bad by taking such wrongdoing to involve an inner contradiction, then his reliance on a universalization test would seem to be seriously misguided. A contradiction resulting from the application of this test would evidently thwart the attempt to will the maxim as a universal law, not the attempt to will it just by itself. The introduction of such a test amounts to an admission that impermissible action is not really in conflict with itself and therefore—on the supposition that wrongness lies in inconsistency—not intrinsically wrong.

⁹It will be convenient in this study to follow the common practice of speaking in this connection of a "test" or "procedure" of "universalization," in order to have a shorthand designation for the reflective exercise of the will whereby it is ascertained whether a maxim can be willed as a universal law. Though such expressions do not exactly correspond to Kant's own terminology, various things he says make them seem suitable. In his example of a man contemplating suicide, for instance, he says, "Now he tests [*Nun versucht er*] whether the maxim of his action could become a universal law" (*G* 422), and later he indicates that in using the formula of universal law one "proceeds by the strict method [*nach der strengen Methode verfährt*]" (*G* 436). If used carelessly or indiscriminately, however, these expressions can impede understanding of how the categorical imperative functions in moral reflection (see §VII.9, §VIII.4). Perhaps the most serious misunderstanding would be to suppose that the imperative is meant to furnish moral deliberation with a mechanical decision procedure.

7. But there is another possibility. The universalization test might be intended to serve the expository function of making the intrinsic wrongness of impermissible action manifest. To play this role, the test would have to serve merely to bring clearly into view something already somehow present in the willing—in the maxim—so that the contradiction could be recognized as located within the maxim or willing itself rather than as the result of the maxim's having a logical operation externally imposed on it.

This way of understanding the role of universalization has two distinct advantages. First, it does not invite us to ask questions such as the following: Why should it matter to us whether a maxim we follow results in a contradiction if universalized? Why should we suppose that an externally imposed logical operation is invested with the authority to constrain our will?¹⁰ If universalization merely makes manifest a contradiction that is already implicit in impermissible action, then it is possible to understand Kant as holding that impermissible willing is in conflict with itself, and that so far as it becomes explicitly self-conscious, it will involve a clear recognition of this inner inconsistency.

Second, if the universalization test is intended to make explicit what is already present in the maxim, then it becomes possible to hold on to Kant's idea that morally impermissible willing is internally inconsistent while avoiding the untenable idea that the inconsistency must trace to a contradiction in the maxim's content. For what the universalization test would make explicit—universality—cannot be thought of as already belonging to that content, since the content of a maxim contains precisely what makes it possible to distinguish one maxim from another, whereas the test is intended to be applicable across the board. As something all maxims have in common, the universality would therefore belong to their form, for if maxims differ on account of differences in their contents, they are all alike maxims in virtue of their common form. And in fact Kant does say that a maxim's form "consists

¹⁰In Fichte's words, "why should I adopt maxims for myself in a certain sphere only on the condition that they can be thought of as universally valid? Kant does not answer this question at all" (Letter to Reinhold, 29 August 1795). In Daniel Breazeale, ed., *Fichte: Early Philosophical Writings* (Ithaca, N.Y.: Cornell University Press, 1988), 407. In a similar spirit, Williams asks, "why must I prescribe any rule?" *Ethics and the Limits of Philosophy*, 62.

in universality" (*G* 436).¹¹ The contradiction that the universalization test is supposed to bring to light in the case of an impermissible maxim would thus not lie in the maxim's content. Nor, of course, could it lie in the form, for then all maxims would be equally contradictory and hence equally wrong. It would therefore have to lie in the relation between the two, so that whether an action is morally right or morally wrong would be a matter of whether the maxim's content agrees or conflicts with its form.

8. This appeal to the idea of the form of a maxim raises questions of its own. Some of them concern the idea of form itself and its relation to content. Others relate to Kant's statement that a maxim's form lies in universality. With regard to the idea of form, we need to ask what it means to say of a maxim that its content "agrees" with its form. If the content is what is willed, what could it be in agreement or conflict with if not something else that is willed? Is the form of willing then *itself* in some sense willed? If so, why does it not belong to the content? If not, how can it belong to the will?

One way of understanding Kant's idea of the form of a maxim would be to take the form to lie in what willing generally presupposes. Presuppositions of willing would not themselves belong to the content of the will, but insofar as they are characteristic of willing in general, they fit with the idea of form. If we can spell out such presuppositions, we can perhaps comprehend how the will's content can be in agreement or conflict with them. This way of conceiving of a maxim's form would yield an understanding of practical self-contradiction that would avoid the difficulties, mentioned earlier, that are sometimes thought to afflict Kant's suggestion

¹¹Two comments: (i) The fact that Kant characterizes maxims as subjective principles of volition, contrasting them with objective principles, which are valid for all rational beings (*G* 400n, 420n), might be taken to suggest that when he says a maxim's form consists in universality he is just thinking of generality across occasions of conduct in an individual's life rather than the sort of universality he has in mind in the universalization test—a universality that extends across persons as well as occasions. But the context of the remark at *G* 436 suggests that Kant sees a very close connection between a maxim's form and the universalization test, and in the argument leading up to his initial statement of the moral principle in terms of the idea of universal law, he proceeds *from* the idea of the maxim of an action done from duty *to* the idea of law *by way of* the notion of the will's "a priori principle, which is formal"—"the formal principle of volition in general" (*G* 400). (ii) Sometimes Kant speaks of the form of a maxim in a sense different from (though not unrelated to) the one of interest to us here; in this other sense, different maxims can have different forms (*KpV* 27, *R* 36).

that in morally impermissible action the will contradicts itself. It would thereby free the way toward clarifying how the badness, or wrongness, of such action could be intrinsic, and it would at the same time contribute to a fuller understanding of the good will's intrinsic goodness.

Though the idea of a presupposition of willing promises to help us understand Kant's talk of a maxim's form, it does not by itself enable us to answer the further questions that are raised by Kant's statement that this form lies in universality. If we conceive of the universalization test as intended to serve an expository function, then we need to think of the universality constituting a maxim's form as in some sense willed in the maxim itself. Yet how are we to make sense of this idea? Even if we leave aside attenuated conceptions of practical reason that would be unproblematically applicable in the case of the egoist mentioned earlier, we seem to face a difficulty here. Is it really possible to conceive of a maxim—even a permissible one—as involving, in some formal way, the willing of itself as a universal law? Maxims seem to be intimately related to choice and are often described as intentions, or intentions of a certain type, but a universal law hardly seems to be the sort of thing one could intend or choose.¹²

To answer these further questions, we will need to bring into focus a conception of the will on which willing does not boil down to bare intending. We can do this by developing the practical-cognitivist account of the will described earlier. A subsequent elucidation of practical knowledge as a species of rational cognition will then enable us to trace universality to the form of practical cognition, and this will put us in a position to make sense of how universality can be the form presupposed in willing.

To articulate the presuppositions of willing, we shall start with Kant's idea that the will, as practical reason, is the capacity for practical knowledge. It will be proposed that willing, as the exercise of this capacity, consists in a certain sort of judgment (§§II–III); that such judgment involves presuppositions that relate it to the form of practical knowledge and turn out to provide the sort of self-constraint on willing implied by the idea of a categorical imperative (§§IV–V); and finally that this constraint is expressed in Kant's formula of universal law and in his other

¹² Here again is Williams: "Practical thought is radically first-personal. It must ask and answer the question 'what shall I do?'" (*Ethics and the Limits of Philosophy*, 21). Cf. W. D. Ross, *Kant's Ethical Theory* (Oxford: Oxford University Press, 1954), 29–30.

two formulas as well (§§VI–VII). Developing this account will be essentially a matter of identifying and explicating the specific concept of will that is of interest to us here. This will lead to further concepts, including those of desire, intention, choice, wish, judgment, knowledge, and law, in connection with which various subsidiary questions will arise. But we will be guided throughout by the primary aim, which is to investigate, at the most basic level, the relation between morality and reason, and in particular to explain that relation by showing that the imperative of morality expresses formal presuppositions involved in willing conceived as the exercise of practical reason, the capacity for practical knowledge, and that these presuppositions contain the form of such knowledge and thereby also the form of its object, the good.

9. It will also be necessary to consider the application of this imperative in some particular cases. This is especially important in view of the considerable difficulty interpreters have faced in attempting to understand the examples Kant himself provides to illustrate how specific duties can be derived from the formula of universal law. In the two centuries that have passed since the *Groundwork* was published, these examples have occasioned a long train of well-known objections raised by many distinguished critics. In response, many interpreters have supplemented the universalization test by introducing further conditions, such as benefit to society, or the rational constraints provided by hypothetical imperatives, and many have discounted or abandoned the formula of universal law in favor of the formula of humanity or the formula of autonomy. The interpretive issues that arise in this connection—how the different formulas compare with one another, whether any one of them is the primary or the most complete expression of morality’s principle, and whether they are equivalent, as Kant suggests—will be addressed (in §VI) so far as is necessary for the main purpose.¹³

¹³For some recent discussion of such issues, see Thomas E. Hill, Jr., “The Kingdom of Ends,” *Dignity and Practical Reason*; Christine Korsgaard, “The Right to Lie: Kant on Dealing with Evil,” *Creating the Kingdom of Ends*, 143–144, 151–154; Onora O’Neill, “Universal Laws and Ends-In-Themselves,” *Constructions of Reason: Explorations in Kant’s Practical Philosophy* (Cambridge: Cambridge University Press, 1989); Thomas W. Pogge, “The Categorical Imperative,” in *Grundlegung zur Metaphysik der Sitten: Ein kooperativer Kommentar*, ed. Otfried Höffe (Frankfurt am Main: Vittorio Klostermann, 1989); Barbara Herman, “Leaving Deontology Behind,” *The Practice of Moral Judgment*, 224–230; Paul Guyer, “The Possibility of the Categorical Imperative,” *Philosophical Review* 104 (1995): 353–385; Allen Wood, *Kant’s Ethical Thought*, chaps. 3–5.

Within the limited scope of this investigation, we shall be concerned with the imperative's application only in its guise as the formula of universal law, which Kant identifies as the canon for use in moral judgment. And we shall concentrate almost entirely on the most primitive cases involving duties to others. The focus on primitive cases reflects the rational procedure followed in the *Groundwork*. That procedure requires an initial abstraction from empirical knowledge, including the empirical concept of human nature, which is not contained in the original consciousness of obligation from which Kant develops his account of the categorical imperative. The empirical concept of human nature (not to be confused with the a priori idea of humanity, on which it depends) is introduced subsequently, in the development of the full doctrine of duties in the *Metaphysics of Morals* (cf. *G* 389, *MS* 216–217). It will therefore be necessary to abstract from such empirical acquaintance with human nature both in our general reflections on the will and practical cognition, and also in our consideration of the primitive duties to others that belong to that part of the basic framework of morality (our knowledge of intrinsic goodness) that concerns our relations to one another and includes primarily the duties of right (or natural justice) and benevolence. At the conclusion, we briefly examine two issues relating to Kant's divisions among types of duty: how to understand the distinction, much discussed in the literature on the universalization test, between what have come to be called "contradictions in conception" and "contradictions in the will"; and what, if anything, universality has to do with duties to oneself.



WILLING AS PRACTICAL KNOWING

II



The Will and Practical Judgment

1. Two Ways of Approaching the Will

Stationed at the junction of reason and desire, knowledge and life, the will can be regarded from two vantage points, which afford contrasting ways of understanding it. In identifying the will with practical reason, Kant reverses the order of concepts in the traditional Scholastic conception of the will as rational desire, reconceiving the will as desiderative reason.¹ This reconception is not, however, a simple rejection of earlier accounts. Kant accepts the traditional practice of identifying the will with the higher faculty of desire, so long as the distinction between the higher (rational) and lower (sensible) faculties of desire is properly drawn.

¹ Through the German school philosophy of his day Kant had ready access to the Scholastic conception of the will as *appetitus rationalis*, a conception whose roots extend back through Aquinas to the ancients, in particular to Aristotle's conception of *boulēsis* ("wish") as rational desire and to his related notion of *prohairesis* ("choice") as "desiderative reason or ratiocinative desire" (*De Anima* 414b2, 432b5; *Nicomachean Ethics* 1139b4–5). Aristotle says wish and choice differ in that the former relates to the end and the latter to the means (1111b26–27), and he explicitly distinguishes the notion of choice from the broader notion of the voluntary (1111b6–9). As we shall see below, Kant draws a similar distinction between wish and choice, locating both under the will. Although in his disjunctive characterization of choice Aristotle moves from "desiderative reason" to "ratiocinative desire" with apparent indifference, the latter expression becomes ambiguous and liable to mislead when not tethered to the former, since standing alone it may suggest the quite different idea of desire *in conformity with* reason rather than *of* reason: desire that, in Hume's words, "arises not from reason, but is only directed by it" (*A Treatise of Human Nature* II.iii.3). Kant avoids the potential for such confusion by setting aside the ambiguous *appetitus rationalis*, consistently portraying the will as practical reason. (Whether Aristotle has a concept of the will has been debated; for discussion, see T. H. Irwin, "Who Discovered the Will?" *Philosophical Perspectives*, Vol. 6: *Ethics*, ed. James E. Tomberlin [Atascadero, Calif.: Ridgeview, 1992].)

Noting Kant's relation to this tradition, we shall first approach his conception of the will from below, by specifying its location under the heading of desire. This will enable us to distinguish the will, as Kant conceives of it, not only from sensible desire but also from mere intention. Marking the contrast between willing and intending will be particularly important for our purposes here, as it will bring into sharp relief the ideas of practical knowledge and practical judgment, putting us in a position to appreciate the significance of Kant's reconception of the will as practical reason. Later (§IV), with the aim of better appreciating that significance, we shall approach his conception of the will once again, but from the other side, by situating it under the heading of reason.

2. *The Faculty of Desire*

In one of the senses in which philosophers have used the term "will," to say of an action that its agent wills to do it, or that it is done willingly, is to say no more than that the action is voluntary. Of course, since "will" and "voluntary" are cognate terms, it is always possible to assert this equivalence as a simple tautology, as Hobbes appears to do when he says, "a *Voluntary Act* is that, which proceedeth from the *Will*, and no other."² But there is a familiar sense of "voluntary," tracing back to Aristotle,³ which provides a natural starting point from which to approach Kant's conception of the will. In this sense, action is voluntary so far as it proceeds, or springs, from desire (in the broad sense of "desire"). Thus, in a typical case, running is voluntary, but not stumbling; so are eating and reading, but not shivering from cold or trembling from fear; nor do we count as voluntary what is done in ignorance, like the things Oedipus did.

This sense of "voluntary" is too broad, of course, to capture the meaning "will" (*Wille*) has in Kant's usage. Since Kant identifies the will with practical reason, he takes the act of the will to be an exercise of reason. But even action springing directly from sensible desire or instinct can count as voluntary in the broad sense just noted. **Young children and the beasts have traditionally been thought capable of voluntary action even though they lack reason.**

²*Leviathan*, chap. 6. Compare Aquinas: "Voluntary is what proceeds from the will" (*Summa theologica*, Ia IIae, q. 6, a. 3).

³*Nicomachean Ethics* III.1.

For animal movement to be voluntary in this broad sense is for it to spring from what Kant calls the “faculty of desire” (*Begehrungsvermögen*). Here of course “desire” does not mean just *sensible* desire, or inclination. Kant emphasizes that the faculty of desire should not be defined in a way that would build into its very concept the supposition that a feeling of pleasure must lie at the basis of its determination, so that desiring something always depends in the end on this capacity’s being or having been affected by some pleasing experience. To do so would be in effect to rule out from the start, by definitional fiat, the possibility of conceiving of the will as practical reason rather than merely as rational desire. It would be to build into the very concept of desire the Humean supposition that desire cannot arise from reason, even though it may be directed by it. A thoroughgoing empiricism would be introduced, by stipulation, into practical philosophy. Instead, the definition should be confined to what can be understood just from the general concept of desire. It should be based on no more than the recognition that desiring something consists in a representing that is also a producing—a making actual, or causing to exist—of its object. Kant accordingly defines the faculty of desire solely through a characterization of the form of operation peculiar to it as a distinct type of representational power, leaving unspecified how or by what its operation might be determined: it is, he says, the capacity a human being or other animal has “to be through its representations the cause of the actuality of the objects of those representations” (*KpV* 9n; cf. *MS* 211). Notwithstanding the long-established tendency, reflected in such expressions as “want” and “passion,” to associate desire with need or passivity, the primitive idea, articulated in this definition, is free of any such suggestion. While dependency may be characteristic of one of its species, desire, according to its original conception, is simply the operation of a certain type of representational power. And since this operation is a representation (or representing) through which an effect is produced, in that through the representation the object represented is made actual, we can characterize desire in general as *efficacious representation*. Kant describes this power as an expression of “lifepower” (*KpV* 23), indicating the distinctive and essential place it occupies in the existence of animate beings.

3. The Faculty of Desire in Accordance with Concepts: Intention

Although the sense of “voluntary” just identified is too broad, a more specific notion, closer to Kant’s conception of the will, is readily available. For an action to be voluntary in this sense is for it to be deliberate, or for it to spring from a desire of a certain type, namely an *intention*.⁴ We can connect this richer notion with the capacity Kant seems to have in mind at one point when, in making a transition from a consideration of the faculty of desire in general to a discussion of the will, he speaks of a faculty that lies between these two powers, in that it presupposes the former, yet is presupposed by the latter. He calls this capacity “the faculty of desire in accordance with concepts” and says that “so far as it is combined with the consciousness of the capacity of its action for the production of the object” it is called the “power of choice” (*Willkür*) (*MS* 213). Where such consciousness is absent, he adds, its act is called wish. Like his definition of the faculty of desire in general, Kant’s notion of the faculty of desire in accordance with concepts leaves unspecified how the act of this capacity is or might be determined. The capacity is conceived merely in terms of the form of act peculiar to it as a distinct type of desiderative power, though Kant says of this faculty that insofar as it is determined to action by a ground lying in itself rather than in its object, it is called “a capacity *to do or to refrain as one pleases*.”⁵

In characterizing this faculty of desire as “in accordance with concepts,” Kant indicates that the desires belonging to it are conceptual representations. And since concepts, as the representations employed in thinking, belong not to sensibility but to the understanding, or the capacity to think, the exercise of the faculty of desire in accordance with concepts lies in a type of thinking and so must be distinguished from sensible desire, or inclination. Yet Kant’s entitlement to speak of the faculty

⁴ A similar distinction between two senses of “voluntary” is drawn by Aquinas: *Summa theologiae*, Ia IIae, q. 6, a. 2. Hobbes’s characterization of will as “the last Appetite in Deliberating” suggests that his conception is at least close to the richer sense of “voluntary.”

⁵ This last capacity is not further explained, but presumably “the faculty of desire in accordance with concepts” counts as “a capacity to do or to refrain as one pleases” to the extent that its exercise does not lie in choices that are constrained by the conditions and circumstances of action, including those resulting from coercion by others. Possession of this capacity is evidently a requisite condition of happiness, where “everything goes according to wish and will” (*KpV* 124).

of desire in accordance with concepts—his entitlement, that is, to identify the capacity he has in view as a specific type of faculty of desire—cannot lie simply in the consideration that the desires belonging to it are conceptual representations. The feature that marks a representation out as belonging to the faculty of desire in general is its efficacy, for it is in the efficacy, strictly and properly speaking, that the desiring lies.⁶ Distinctions between forms of desire and between the specific faculties of desire corresponding to them must therefore rest, not on differences in the ways desires *represent* their objects, but on differences in the ways they are *efficacious*. Here, and here alone, can we find different forms of *desire*. What is distinctive about efficacious thinking as a form of desire must therefore lie in the specific form of its efficacy. As we shall see, this type of efficacy belongs solely to efficacious representation of *action*, so we may, in keeping with Kant’s traditional Aristotelian usage, call it *practicality* and call the thinking that has it *practical thinking*.

functional specification in terms of efficacy

To identify practical thinking’s specific form of efficacy, we must turn to the distinction on which Kant implicitly relies in speaking of the faculty of desire in accordance with concepts—that between understanding and sensibility—and consider how it bears on the efficacy of representations. As is well known, Kant bases this distinction in the contrast between spontaneity and receptivity, or between a cognizing subject’s capacity to be self-consciously active and its capacity to be aware of the affections it undergoes.⁷ It is to this contrast, then, that we should look to find the differentiae. While practical thinking and inclination both fall under the broad heading of desire that corresponds to Kant’s notion of the faculty of desire, in that each is a representation with an efficacy that lies precisely in its working to bring the object it represents into existence, they differ in that this efficacy is related to the subject through spontaneity in the one and through receptivity in the other. As spontaneously efficacious, practical thinking is a type of desiderative representation in which the self-consciousness distinctive of conceptual

spontaneity and receptivity

⁶ “As *striving (nisus)* to be by means of one’s representations a *cause*, a desire is . . . always causality” (MS 356).

⁷ Cf. A51/B75. Ultimately, this distinction lies in a distinction between two forms of consciousness: consciousness that, being aware of itself as arising from itself, necessarily recognizes itself as the source of its object and thereby understands the latter as its product, and consciousness the subject recognizes as arising from a source outside itself and so regards as an affection of itself by something acting upon it.

representation belongs to the representation's very efficacy, to the striving constituting it as desire. This amounts to saying that the efficacy of practical thinking is integral to the thinking itself, so that the self-consciousness essential to thinking in general also pertains, in the case of practical thinking, to that thinking's efficacy, its productive power, and hence that the efficacy depends on the consciousness of it. Thus **practical thinking can make its object actual through and only through its consciousness that it can do so**. Its practicality, then, lies in this distinctive self-conscious and hence self-constituting "can," through which the object of such thought is always, correspondingly, represented as possible (actualizable) and thereby constituted as such. **Sensible desire, on the other hand, is representation with an efficacy of which the subject is aware only through being affected by that very efficacy**. This form of awareness falls squarely under the heading of feeling—pleasure or displeasure—rather than thinking. As Kant characterizes it, feeling is both the effect of a representation on the subject (*MS* 212n) and a consciousness of that representation's causality in respect of the subject's state of mind (*KU* 220). **What is distinctive of inclination, or sensible desire, then, is that the efficacy by which it works to bring the object it represents into existence is one the awareness of which lies in the feeling that accompanies it, salient modes of which are pleasure in the presence of the object (the enjoyment of an apple, for instance) and displeasure or pain where that object's realization is blocked or hindered. The difference between the two ways in which desire can be efficacious, then, is that while in the case of inclination the awareness of the efficacy depends on the efficacy, in practical thought the efficacy depends on the awareness**. The efficacy either has an effect *on* or is an effect *of* the subject's consciousness. **In the one case we are aware of the efficacy from its effect, in the other from its cause**.

practical
thinking vs.
inclination/
sensible desire

This abstract sketch can be elucidated by considering the act of intention, which has the self-conscious efficacy just described and so belongs to the capacity for practical thought (specifically, to this faculty operating as the power of choice). **It is part of the idea of intention that the efficacy whereby an intention can bring about the performance of the action it represents lies in the intention's understanding itself to have such efficacy**. When I mean to cross the street, my intention includes the implicit understanding that I *therein can* cross it. This understanding is not a further thought, distinct from the intention, such as

might be reached by an inference from an introspective awareness that I intend to cross the street to the conclusion that I can cross it, but the intention's self-conscious awareness of its own efficacy. That an intention's efficacy lies in its understanding itself to be sufficiently efficacious is reflected in the often-noted connection between intention and expectation. Independently of any confidence or diffidence regarding one's abilities that one may have acquired from experiencing previous successes or failures in execution, whether one's own or others', one's intention immediately engenders expectation, its natural and necessary effect on theoretical thought in the latter's ordinary practically related use. To intend to do something is also, thereby, to expect to do it, and thus successful achievement is never wholly surprising.⁸ Since this connection between intention and expectation expresses the recognition, in the case of practical thought, of the general connection between the basic concepts of power and act, according to which "can" implies "will," it also reflects the self-conscious efficacy of intention. Another indication that intention's efficacy is self-conscious—one that will be considered more closely below—is that intention is subject to a certain condition of rationality, which functions to prevent us from intending to do what we can see (in theoretical judgment) to be impossible.

On account of its distinctive self-conscious efficacy—what we might also call its practical self-consciousness—practical thought contains within itself the conception of the *practical subject*, the self of this self-consciousness. No thought of a practical subject need be contained in bare inclination, where such self-consciousness is not involved. But the very idea of intention as a practical conception implies that intention—that is, the act of intending—includes, as part of its form, an at least implicit

⁸ Some argue that intention does not imply belief (e.g., Donald Davidson in "Intending," in his *Essays on Actions and Events* [Oxford: Oxford University Press, 1980], 91–96), and to the extent that surprise at an outcome is a mark of the antecedent presence of an opposing belief, the fact that we sometimes find our own success surprising would seem to indicate that there is a sense in which this is true. But even where one is surprised by one's own success (say, in passing a difficult examination), this surprise differs in character from, and falls short of, the surprise that would result were the same outcome to occur despite one's not having even attempted to achieve it. By the same token, a primitive element of surprise is characteristically contained in the experience of failure, as an essential constituent of disappointment. The expectation tied to intention is not merely an incidental concomitant, nor is it a prediction based on prior experience of intentions and their results, though such experience can greatly affect its degree (one's level of confidence); its root, its foundation, is nothing but the intention's understanding itself to be sufficiently efficacious.

awareness not only of the subject of the intention but also of that subject as being, through that very intention, agent of the action represented in the intention. It implies, in other words, that to intend to do something is to be aware of oneself, the subject of the intention, both as distinct from the act of intention and also as necessarily identical with the agent of the intended action.

This implication involves two points, one resting on the consideration that practical thinking is *thinking*, the other relating to its *practicality*. **As practical *thought*, intention must be at least implicitly conscious of itself.** This self-consciousness, however, includes not only the self-awareness of a *specific* act of thinking (the awareness, in intending to do a certain thing, of that very intending itself), but also, present in that same specific act, the self-awareness of a *general* act of thinking. And the latter self-awareness, in being aware of its own generality, is conscious of itself as capable of being present in other possible specific acts of thinking distinct from the present one. When someone exercises for the sake of health, for instance, diverse specific intentions are involved, and any one of them (an intention to bend a certain limb, say) contains within it the *same* general intention to exercise that is present also in the *others* (to straighten, to bend (again), etc.); moreover, that general intention itself contains the yet more general intention to preserve one's health, which is also present in various *other* intentions carried out in the diverse actions undertaken for the sake of that aim (regulating one's diet, avoiding danger, treating illnesses, and so forth). The general self-awareness must therefore be distinct from, even though present in, any such specific act of thinking. It thus constitutes the *thinking subject* on which the present specific thought, along with every other in which this same general self-awareness is present, depends, and to which they all, therefore, in virtue of this dependence, belong, as specifications of it.

Second, since this self-consciousness belongs to a *practical* conception, that is, to a conception whose efficacy, being proper to it as a type of *thought*, lies precisely in the conception's *understanding* itself to be efficacious, it is a consciousness of oneself as a *practical* subject, or agent—a subject that engages in practical thought and thereby in action, bringing what it thinks into existence. Hence, as was noted, such a subject understands that the agent of the intended action is necessarily identical with itself, the subject of the act of intention.

Drawing on these points, we can characterize intention in terms of its practical subject. To intend to do something is to specify, in an act of practical thought, what one means to do, and this is just to specify one's conception of what one, as a practical subject, means to do, which in turn is to specify one's conception of oneself as an agent, or what we may call one's practical self-conception. And since this self-conception is efficacious, it constitutes the agent, the practical subject, itself. It follows that, even if a practical subject depends entirely on sensible desires for the basic representations of the objects it means to realize, a representation of an object of sensible desire can be included in intention only insofar as the subject has, in deciding what to do, taken it up into its practical thought and hence into itself, thereby appropriating (or "incorporating") it.⁹ It follows too that, as the agent's *own* efficacious specification of what it means to do, intention can also be characterized as the *self-specification* of this practical self-conception, an act that might also be described as a practical subject's attachment of a practical predicate to itself.

the intention to do something

incorporation

For an action to be voluntary in the more robust sense just outlined, then, is for it to spring from an intention, or an exercise of practical thought in which one attaches to oneself a conception of some action—a practical predicate—and therein efficaciously specifies what one means to do. The efficacy of this specification lies in its understanding itself to be (at least in the actual conditions) able, or sufficient, to effect the action it represents and is such that, to the extent that conditions allow, it does result in the carrying out of the specified action. Intention is initiated by one's (perhaps tacit) deciding what to do.

4. Means–End Rationality

In contrast to the broad notion of the voluntary with which we began, the more specific conception we have just been considering is connected with a certain form of rationality. This form, often called "means–end"

⁹This taking up of a representation of an object of sensible desire is similar to the taking up—or "incorporating," to use the word some translators employ—of an inclination into one's maxim that Kant speaks of in describing the freedom of the power of choice (*R* 24). It is not the same, however, for as will be explained below, bare intention does not amount to an exercise of that power. Kant's "incorporation thesis" is discussed by Henry Allison in *Kant's Theory of Freedom* (New Haven, Conn.: Yale University Press, 1990), 5–6, 39–40, and *passim*.

or “instrumental” rationality, has long been recognized, under various names, as a distinct type of intelligence, and though not inherently opposed to the form of reason specific to ethical thought, it has long been distinguished from it. Aristotle, for example, notes the difference between “cleverness” and practical wisdom, characterizing the former as “the power to do the things that tend toward the mark we have set before ourselves, and to hit it.”¹⁰ And Kant marks a similar contrast in dividing the imperatives of reason into hypothetical and categorical.

Because practical thought is a type of desire and operates under this form of rationality, it can, with a certain justice, be described as rational desire. The capacity exercised in such thought—“the faculty of desire in accordance with concepts”—is thus far more liable to be confused with the will in Kant’s sense (as practical reason) than is the faculty of desire in general. No one supposes that willing in Kant’s sense is simply a matter of desiring, but interpreters have not infrequently sought to assimilate it to intending, in one way or another reading his account of practical reason through the lens of means–end rationality. Such assimilations can in turn exacerbate the difficulty of explaining how morality’s categorical imperative is based in practical reason. For they can help hold in place the common assumption—an assumption we set to the side at the outset—that to be successful such an explanation must yield an argument that could win over a moral outlier such as a rational egoist, to whom only means–end rationality or the like would be unproblematically attributable. It is particularly important for our purposes here, therefore, not to confuse these capacities. To put ourselves in a favorable position to appreciate what there is to willing in Kant’s sense beyond mere intending, as well as what willing and intending share in common, it will be useful to bring into sharper focus the sort of rationality of which bare intending is capable. This form of rationality will prove in the end (§V.5) to be one of two essentially different forms, one of which is tied to the relation the faculty of desire in accordance with concepts bears to the subject’s capacity for theoretical knowledge, the other to the relation it bears, in a human subject, to the capacity for practical knowledge.

¹⁰ *Nicomachean Ethics* 1144a24–26. In Kant’s terms, the capacity is *Klugheit* (*G* 416). Strictly speaking, Aristotle’s cleverness and Kant’s *Klugheit* include, in addition to the formal perfection of means–end rationality, the perfected talents of judgment and wit, but these latter lie outside our concern here.

“Rational,” in the sense it takes in “rational desire,” indicates a relation to reason, a relation of conformity or agreement. Reason, as Kant characterizes it, is “the faculty of principles,” or our capacity for “knowledge from principles,” by which we know the particular in the universal (A299–300/B356–357; cf. *G* 412). Since the principles spoken of here are so called because they are primary in reason’s cognition, knowledge from such principles is a priori knowledge: “rational knowledge and a priori knowledge are the same” (*KpV* 12). It is by reason, then, that a cognizing subject is able to reach judgments about particular matters from the universal principles on which it relies in its cognitive activity. We will examine this conception of reason more closely below (§IV); for now it will suffice to note that the principles to which Kant here refers are, in the first instance, the principles of cognition in general, principles on which cognition depends for its constitutive consistency and systematic unity, and which, on account of the self-consciousness characteristic of cognition, must themselves, upon reflection, be recognizable as such by the cognizing subject. **The notion of rationality therefore has no footing in actions or activities that do not depend on the exercise of the cognitive capacity, no application to the life-functions and instinctive behavior of bare animal existence.** Nor does it gain purchase in activities where this capacity is engaged but not in a way that depends on an awareness of any principle of cognition governing the activity, as happens, Kant argues, in the appreciation of natural beauty. But where a particular instance of an activity of the sort that does depend on a consciousness of governing principles of cognition is in nonaccidental agreement with those principles—expressing them, so to speak—it is rational.

Now practical thought just as such is not any type of cognition or judgment about anything. Being merely the efficacious specification of what one means to do, it can be arbitrary in a way that judgment and cognition never are. It includes, however, a certain self-understanding that serves as a principle for it and relates it to principles of cognition, enabling it to be regarded as a rational activity (and even to count as knowledge of a sort—“knowledge of what one is doing”—in a subject possessing the requisite know-how or skill). Since practical thought distinguishes itself from other types of thought—principally from theoretical thought—through being efficacious, or productive of what it thinks, this principle is at bottom just practical thought’s own understanding of itself as efficacious thought. As the practical self-consciousness that

constitutes the practical subject, it is one and the same in all of the subject's practical thinking and hence universal with regard to it. And though it is precisely what distinguishes practical thought from theoretical, it at the same time establishes practical thought as standing in an essential relation to theoretical cognition and its principles. For to understand one's practical thought as efficacious is to understand it as *capable* of making what one therein thinks *actual*, yet "actual" signifies nothing if not what one can in principle know theoretically through experience. Indeed, such knowledge of the actuality of the action represented in practical thought is the true aim of practical thought; it is only in reaching this knowledge (or at least what the subject supposes to be such) that the efficacy—the pursuing, the trying—brings itself to rest. In the paradigm case this knowledge is gained by witnessing the outcome, but it can be acquired in other ways as well. A man who means to bring down a house might, after undermining its foundation, turn his attention to other matters without waiting for the event, knowing already that the building will fall. In short, intending to \emptyset essentially rests on the practical understanding that one *can* \emptyset and naturally culminates in the theoretical cognition that one *does* \emptyset or *has* \emptyset -ed. In thus regarding the intended action as possible, the subject presupposes that no antecedently attainable theoretical knowledge entails its impossibility, its impracticability. Practical thinking, then, is rational to the extent that it is in agreement with its self-understanding and thereby compatible with theoretical knowledge.

As it stands, however, this statement is abstract and one-sided. Since it leaves out of consideration the conditions in which practical thought arises, it does not display such thought's distinctive form of rationality in its most familiar guise, as a form relating end and means. The presupposition just noted—that the action is not impossible—simply expresses the formal condition to which practical thought is subject through relating itself to theoretical cognition. In a human being, however, practical thought also depends on sensible desire as its activating material condition. Indeed, we might even say, borrowing the hoary but still useful terminology of the Aristotelians, that while human practical thought looks to theoretical knowledge for its form and final cause, it finds in sensible desire its matter and efficient cause. To reach a more determinate formulation, then, we need to consider practical thought not only with respect to theoretical knowledge, but also in its relation to sensible desire, recognizing that the latter both sets it in operation and also supplies its materials,

the objects such thought represents itself as producing through the action it effects. In doing so, we shall be contemplating human practical thought in the setting in which it implicitly understands itself to be operative, as practical thought in a rational animal, a being in whom such thought, as rational desire, can complement sensible desire by framing a representation of action that can bring about the object the latter represents. We have, then, to consider what the rationality of practical thought comes to insofar as such thought understands itself to be both sufficient to effect the action it represents, yet dependent on sensible desire for the materials—possible objects to be produced—on which it relies in framing its representation of that action. Since we are concerned at present with the rationality of practical thought just as such, we shall here leave unspecified whether, in being affected by this material condition, the human capacity for practical thought is also *determined* by it, or whether it is free, or determinable by pure reason (cf. *MS* 213).

By its nature as sensible, sensible desire arises from a pleasing representation that comes to be through an object's somehow affecting the senses. This representation must therefore be either itself sensation (the tastes and smells figuring in the eating of an apple, for instance) or dependent on sensation, typically the experience of success in something one has undertaken to do (the recognition that one has steered safely between two rocks, say, or helped a friend out of difficulty, or managed to get ahead in life).¹¹ In Kant's terminology, any representation that comes to be in this way can be characterized as a representation of the *existence* of an object, to indicate that it is of a type suited to be or to figure in theoretical cognition, cognition of what *is*. Such representation is thus to be distinguished from the efficacious representation in which desire consists, which is not *of* its object's existence, but *productive* of it, in that through its efficacy it makes its object actual. By its nature as sensible, then, sensible desire arises from a pleasing representation of the *existence* of some object. And by its nature as an expression of lifepower in the operation of the faculty of desire, sensible desire reflects the self-productive character of that power in that it is essentially reproductive (and therefore habitual), having as its object the very thing that brought it about. But since what brings about sensible desire is not simply the existence of the object the representation of whose existence is found to

¹¹ See *KpV* 21–25, *KU* §§2–3.

be pleasing, but the *pleasing representation* of that object's existence that arises through the object's affecting the senses, the power of the object thus to give rise to such pleasing representation belongs to its form as an object of sensible desire, a form Kant marks by designating that object as "the agreeable" (*das Angenehm*), and the *operation* of that power accordingly belongs to the object's existence. Thus, **the efficacy of sensible desire works, not to bring an object having such power into existence absolutely (someplace in space and time, it matters not where or when), but rather to bring into existence such an object in a certain relation to the subject, namely the relation constituted by the representation of the object's existence, which in the primitive case is just the pleasing sensation or experience of the object—in a word, the enjoyment of it.**¹²

Notwithstanding its efficacy, however, sensible desire properly and strictly so called includes no practical representation whatsoever, no representation, that is, of its agreeable object as to be produced. For as we have noted, such desire's efficacy is recognized only through feeling; it is not of the sort that depends on the consciousness of it, the consciousness that constitutes the original concept of action. The efficacy of sensible desire is therefore always subjective, never objective: rather than immediately issuing in action that, conditions allowing, produces its object, sensible desire works on the subject, and so far as the resulting action is deliberate at all, sensible desire leads to it only in virtue of being an affection of the subject's capacity for practical thought, the faculty of desire in accordance with concepts.

As a modification of the receptivity of the practical subject's faculty of desire, sensible desire arouses the capacity for practical thought, stimulating (but not necessarily determining) it to activity as the practical subject frames a problematically practical representation of itself as producing the object of sensible desire. This representation is "problematically practical" not in the sense that it is practical in a way that poses some difficulty, but in the sense that it is practical *in potentia*, pending judgment as to the action's practicability. **It is not itself intention, but rather the primitive thought of action out of which intention arises, a thought that precedes and indeed is presupposed in any specification of how the object is to be produced and any judgment of whether the action is possible, that is, practicable.** It differs from intention, then, in being

¹² Sensible desire and its object will be considered further in §III.2 below.

the starting point of deliberation, not the end point. Practical thought must begin with a merely problematically practical representation because the bare concept of an object of sensible desire, being only the representation of an object the subject finds agreeable, does not itself originally contain the thought that this object is something the subject is able to bring about. That concept is merely of something that can *produce* an effect *on* the subject, not of something that can *be produced* as an effect *by* the subject. Because the object of sensible desire is always one whose existence is to be empirically represented, however, the object the subject problematically practically represents itself as producing is always one standing to the subject in the relation described above, the relation constituted by the sensation or experience in which the subject's enjoyment of the object is found. Now such experience or sensation must be, or belong to, a possible theoretical cognition, for an experience is itself such a cognition, and a sensation is just what can figure as sensible material in the theoretical cognition of the actuality of its object. But since, as we have noted, the subject of practical thought must always understand itself to be as well a subject of theoretical cognition, it also has an implicit understanding of theoretical cognition and its constitutive principles, including the recognition that every bit of theoretical knowledge must be in agreement with every other bit. So in order for practical thinking to be rational, or in agreement with its own constitutive understanding of itself as sufficient to effect the action it represents, the represented action must be one whose actuality is compatible with what the subject theoretically knows about the circumstances in which it acts and about its own productive powers—including not only what it knows about the limits of the latter but also what it knows, on the basis of its awareness of what it already intends to do, about how these powers are, or will be, exercised in the pursuit of other objects of sensible desire.¹³ Practical thought is thus rational only if the subject's theoretical

¹³ Because this knowledge is partly based on the subject's *practical* self-conception as specified through the various practical predicates it has actually attached to itself, it is not theoretical knowledge in the narrow sense with which Kant is concerned in the *Critique of Pure Reason*. It is nevertheless theoretical in the sense that it is of *what is*, even though it depends on practical thought of *what is to be*. Because it has this special character, this sort of theoretical knowledge falls within the subject's control in a way that other theoretical knowledge does not, to the extent that the subject has the capacity to decide to abandon or restrict its pursuit of one object of sensible desire to make room for the pursuit of another.

knowledge does not already entail that such action—the action represented in the full-fledged practical thought arising from the initial, problematically practical representation of the production of the object of sensible desire—is not possible.

This condition is merely negative, however. To have in full the efficacy it understands itself to have, practical thought must also agree positively with its understanding of itself as sufficient to effect its action. Beyond its representation of the end, practical thought must also contain a representation of the means. It must include more than just the representation of its action that figured in the initial, problematic thought of the subject as (somehow) bringing about the enjoyment of the object of sensible desire. It must also contain a positive representation of how the object is to be produced, that is, a representation of the action that comprises solely representations of actions the subject can, by exercising its theoretical cognitive capacities, judge to be not only possible (either immediately or conditionally upon the performance of other actions the subject can similarly judge to be possible) but jointly possible, given its theoretical knowledge of its powers, their limits, and the circumstances. The successful exercise of the capacity to hit upon such a representation—the capacity that, in its perfected form, Aristotle calls “cleverness”—thus establishes a *specified* practical representation in positive agreement with its own formal understanding of itself as sufficient to effect its action. But if this is to happen (if, in other words, deliberation is to reach a happy conclusion), the specifically represented action must be one the subject can theoretically recognize to be not only *possible*, but *sufficient* to produce the object it originally problematically represented itself as producing (or, if that object is deemed unattainable, some practicable substitute decided on through that same deliberation).

In positioning itself, in the manner just described, as an intermediary of sorts between sensible desire and theoretical cognition, practical thinking elaborates its intrinsic understanding of itself as efficacious thinking into an understanding of itself as subject to two conditions: a condition of possibility, or practicability, on the side of its relation to the subject, the cause; and a condition of sufficiency on the side of its relation to the object, the effect. Together they constitute a twofold condition of rationality, which we can express in the following proposition: (i) the problematically represented action of realizing an object of sensible

desire—the end—is to be intended only insofar as it is possible, or practicable, and (ii) what is specifically practically represented in accordance with the subject's theoretical judgments of practicability—the means—must be enough, or sufficient, to produce the object. This two-sided condition—that the end be possible and the means be sufficient—expresses the relation binding end and means together in a single action. For since we determine the possibility of an end by reference to the means at our disposal and the sufficiency of means by reference to the end, it merely lays down that end and means are each to agree with the other.

Although Kant does not himself elaborate this condition of rationality for bare practical thought, we can see it expressed in his definition of the “power of choice” as the faculty of desire in accordance with concepts “so far as it is combined with the consciousness of the capacity of its action for the production of the object” (*MS* 213). For we may reasonably take this consciousness to be practical consciousness of the action's sufficiency to produce the object (clause ii), which includes consciousness of the possibility, or practicability, of the object's production (clause i). The same condition of rationality is also clearly reflected in what Kant says about the will, or practical reason. The first of the two clauses is explicitly registered in his statement that “If the object is assumed to be the determining ground of our faculty of desire, then the *physical possibility* of it through the free use of our powers must precede the judgment whether it is an object of practical reason or not” (*KpV* 57). And the second clause, recast as the statement that the means must include everything the subject needs to do,¹⁴ finds direct expression in the famous analytic proposition Kant puts forward to explain the possibility of hypothetical imperatives: “Who wills the end wills (so far as reason has decisive influence on his actions) also the indispensably necessary means to it that are within his power” (*G* 417).¹⁵

¹⁴ At first sight this statement may seem not to be equivalent to the second clause, even if entailed by it (cf. Onora O'Neill, “Consistency in Action,” *Constructions of Reason: Explorations in Kant's Practical Philosophy* [Cambridge: Cambridge University Press, 1989], 91). But it is necessarily true that doing *everything necessary* to achieve some object or end is identical with doing *something sufficient*. For to do something sufficient is obviously to do everything necessary, and until one has done something sufficient there is more that needs to be done.

¹⁵ Two comments: First, Kant adds the qualification “that are within his power” not because one can *not* but because one *can* rationally will what lies beyond one's power (e.g., a

That the condition of rationality for intending should be reflected in Kant's discussion of rational requirements on willing does not mean, however, that rationality for bare intending can be articulated through his idea of an *imperative* of reason. If we assume that in a given subject the faculty of desire in accordance with concepts is, in respect of its representation of ends, determined by sensible desire, or by the effects that objects found to be agreeable have on the subject's faculty of desire (an assumption the bare notion of intending does not enable us to rule out), then the condition of rationality we have just considered will never manifest itself in an imperative in Kant's sense. For as a command of *practical* reason, an imperative expresses the *efficacy* of reason (G 413), but in such a subject the efficacy essential to practical thought as a form of desire will always spring from the object, through a sensible desire for the latter, rather than from an efficacious principle of reason. Though practical thought will rely on its self-understood relation to *theoretical* cognition for the specific *content* of its representations, its efficacy, even though self-conscious, will in the end always stem from sensible desire. In such a subject, practical thought is, and can only be, the slave of inclination: it "arises not from reason, but is only directed by it."¹⁶ This is as much as to say that, in the sense of "practical" we have been using here, the rationality of bare intention is not a form of *practical* rationality, though it could be described as the rationality of practicality.

If a rational animal of the sort just described is not subject to imperatives—if, in other words, its understanding of the sort of rationality of which it is capable is not itself rational *desire*—then neither is there any possibility of its displaying in its practical thought the full-fledged *irrationality* of thinking in opposition to that understanding. For the only

universal law, the highest good): the qualification serves merely to restrict attention to the bearing the more basic unqualified truth has on one's willing of things within one's power. Second, Kant classifies the hypothetical imperatives standing under his proposition into two types, imperatives of skill and imperatives of prudence (G 417–419). Since our present concern is merely to outline the rationality of intention, we need not attempt a comparable division here, though we can anticipate that the distinction would rest on the difference between bringing the rationality condition just considered to bear on specific acts of practical thought taken individually and in isolation from others and bringing it to bear on a global act of practical thought in which the subject means to pursue, as a single object, the totality of the objects of sensible desire (the agreeable in general). We shall consider the latter case below, in connection with the will (§III.4–6, §V.5), but see also §III.7.

¹⁶ *A Treatise of Human Nature* II.iii.3.

thing a desire can oppose is another desire. Nor is it even conceivable that a rationality condition that lies merely in the form thinking understands it must take if it is to make any contribution at all to the production of the object of sensible desire could ever in any way conflict with such desire. As Hume observes, "The moment we perceive the falsehood of any supposition, or the insufficiency of any means, our passions yield to our reason without any opposition."¹⁷

Yet nothing in the notion of such a subject entails that it is not liable to practical inconsistencies in cases where opposition arises among its sensible desires. Stretched between its self-understood relation to theoretical cognition on the one side and the forces exerted upon it by a multitude of sensible desires on the other, the practical thought of a finite subject, in whom practical attention can be scattered among various objects, may fall into inconsistency. Not that the subject thinks a contradiction. But owing to limitations in its capacity for practical thought, it fails to ensure that its various intentions are collectively integrable into its theoretical cognition. The fabric (though not the form) of its practical self-conception is rent asunder by the opposing pulls from the competing objects of its attention. Everyone knows of the pressures and tensions that can arise from sensible desires in human life, and bare practical thought is no less beset by such forces. One aspiration contends with another, and all of them compete with the enjoyments of ease and indolence; the mirror and scale wage war against the table and pantry; ambition disrupts the pleasures of friendship and society. Pursuing too much, the subject may end up with little or nothing. Discrepancies of this sort, made possible by the distractions to which this subject's practical thought is liable, can be regarded as a kind of irrationality. In fact, however, they are merely manifestations in practical thought of opposition between sensible desires (relative to the subject's theoretical cognition), displays of the strains practical thought undergoes in serving multiple masters according to the varying levels of relative strength they individually exert when, rather than merely providing the materials to which practical thought applies itself in pursuing its primary object (the agreeable in general), they also influence it, impairing its operation.

¹⁷ *Ibid.* The recognition that the possibility of irrationality is ruled out may lead us to ask whether such a subject can, after all, be properly characterized as rational, or whether rationality entails the real possibility of irrationality. We shall reserve questions such as these for later consideration, in connection with the will (§IV.4).

The foregoing considerations confirm that the notion of intention, unlike that of the voluntary in general, is connected with a form of rationality. Yet this idea clearly does not exhaust all that is implied by Kant's conception of *Wille* as practical reason or by the Scholastic philosophers' cognate definition of will as rational appetite. One indication of this is that the conception of rationality as lying merely in the agreement of practical thought with its understanding of itself as the exercise of a productive capacity does not depend on any idea of a rational capacity to know the good (unless "good" is defined simply as the necessary means to some object aimed at in practical thought), whereas such an idea is plainly involved in the notions of practical reason and rational appetite employed by Kant and the Schoolmen.

According to Kant, practical reason has as its sole objects those of the good and the bad; it is by reason, he holds, that good and bad are judged (*KpV* 58). And he says of the Scholastic formula *nihil appetimus nisi sub ratione boni, nihil aversamus nisi sub ratione mali* (we desire nothing except under the idea of the good, we are averse to nothing except under the idea of the bad) that it is "indubitably certain" if it is rendered as "we will nothing according to the direction of reason except insofar as we hold it to be good or bad" (*KpV* 59–60). But from the supposition that a practical subject, by maintaining coherence among its intentions in accordance with its theoretical knowledge, exercises its practical thought in a manner that agrees with this thought's own understanding of itself as efficacious it by no means follows that the general act of practical thought that constitutes that subject and to which all of the latter's specific intentions belong involves any understanding of itself as rational cognition of the good. Nor does it follow that this act or any depending on it is itself good or has anything good as its object, nor even that any of these thoughts involve the concept of the good at all.

5. The (Free) Power of Choice

We can locate a sense of "will" closer to Kant's conception of *Wille* by considering the expression "weakness of will" in its familiar use as a term connoting incontinence, or *akrasia*. Weakness of will is manifested whenever the will is thwarted or hindered by an opposing sensible desire; and to the extent that the latter prevails in the conflict,

action contrary to what is willed may ensue. But while most would agree that such action may itself be intentional,¹⁸ the conflict never boils down to a conflict among mere intentions and so never amounts to the sort of irrationality described above. Action reflecting weakness of will is “against one’s better judgment”: to act against one’s will is to act, not just against an intention, but against a judgment as to what one should do or how one should act, or in spite of the fact that one “knows better,” where this knowledge of the better occupies a more or less settled position in one’s character.¹⁹ These points reveal that willing in this sense bears an intimate kinship to intending but does not reduce to it. In particular, they indicate, on the one hand, that, like intending, willing is self-consciously efficacious action specification. For if in willing no action were specified, one’s deeds could neither agree nor conflict with one’s will; if willing had no efficacy, the will could have neither strength nor weakness; and if the efficacy were not self-conscious, willing could not lie in any sort of thinking and hence could neither be nor be based in judgment or knowledge. But they also show, on the other hand, that willing is unlike mere intending in that it amounts to, or at least contains at its basis, a recognition or a judgment that the specified action is good, so that wittingly acting against one’s will involves the thought or awareness that one is acting badly or wrongly.

Moving from the bare notion of intention to this concept of will does not quite carry us all the way to Kant’s idea of *Wille*, but it does seem to bring us to his closely related notion of the “choice” (*Wahl*) that lies in the exercise of the “free power of choice” (*freie Willkür*), the power of

¹⁸ As J. L. Austin remarked, succumbing to temptation—in his well-known example it was helping himself to more than his share of dessert at high table—need not imply that one has lost control of oneself (“A Plea for Excuses,” in *Philosophical Papers* [Oxford: Clarendon Press, 1961], 146).

¹⁹ Contemporary discussions commonly depart from traditional treatments of weakness of will in dropping the condition that the judgment be an act of knowledge. See, e.g., Donald Davidson, “How is Weakness of Will Possible?” *Essays on Actions and Events*, 21. This is a noteworthy modification, in that it eliminates a critical element of the traditional puzzle and a principal factor on which its philosophical interest depends. If morality lies in a type of knowledge, this condition must be retained, at least for the fundamental case, so far as weakness of will is a topic for ethics.

choice insofar as it can be determined by pure reason (*MS 213*).²⁰ For Kant identifies this power as the source of maxims (*MS 226*), which seem, from his characterizations of them, to amount to willing in the sense just described: On the one hand, maxims must be self-consciously efficacious action specifications if, as Kant says, a maxim is “the principle on which the subject *acts*” (*G 420n*); and on the other hand, they must be judgments of a certain sort if, as he holds, they are “volitional opinions” (*Willensmeinungen*) in which the concept of the good is employed (*KpV 66*), claims or presumptions that a certain way of acting is good, albeit under a condition that may hold only for the subject in question (*KpV 19*). Moreover, Kant conceives of maxims as the basic determinants of character (*KpV 152*), and it is in action that is “out of character” that weakness of will as traditionally understood is revealed.²¹

In employing the word “choice” to refer to the exercise of *freie Willkür*, we are assigning to it a sense that differs in some respects from ordinary usage. In one respect this sense is narrower. For in cases of action betraying weakness of will, the notion of choice just introduced is clearly not applicable to the practical thought or desire that issues in the action, yet there seems plainly to be a sense in which such action can, at least in some instances, be described as chosen. In other respects it is broader. “Choice” is commonly applied only where the action is

²⁰ Kant thus distinguishes between the power of choice (*Willkür*) in the general sense noted earlier and the *free* power of choice (*freie Willkür*). But he also uses *Willkür* and *Wahl* even when speaking of action that is voluntary only in the broad sense: he distinguishes the power of choice in a human being (*menschliche Willkür*), which is exercised in accordance with concepts and so requires understanding, from the power of choice in a beast (*tierische Willkür, arbitrium brutum*), for which the capacity to represent objects even without their being present (i.e., the imagination) is required, but not the understanding (see, e.g., *MS 213*). The different senses of “will” that we have so far noted might thus be marked by the expressions “animal choice,” “rational choice,” and “free choice.” They signify, respectively, the desiderative or practical employment of imagination (in inclination), of understanding (in intention), and of pure reason (in maxims), and they correspond to the three elements—(animal) life, rationality, and accountability—that figure in Kant’s stepwise exposition of the human being’s naturally determined end as comprising animality, humanity (rational animality), and personality (accountable rationality). (See *R 26–28* and §III.7 below.) Since only the last of these three senses will fall squarely within the scope of our concern in what follows, “choice” will henceforth be used, except where otherwise indicated, as shorthand for “free choice.”

²¹ See *Nicomachean Ethics* VII.8. In commenting on the frailty of human nature (*R 29*), Kant briefly discusses the failure to carry out the action one wills in accordance with one’s maxim.

within our control, or where we have a confident expectation of success. If you found you had lost your keys while running errands about town, for instance, we might say that you could choose to retrace your steps in search of them, but not that you could choose to find them. The exercise of *freie Willkür*, on the other hand, can include the election of any action we think we may be able to carry out, any action we do not regard as lying beyond our power (*MS* 213). An additional but closely related difference is that choice is often thought to be at least primarily of means, while the exercise of *freie Willkür* also includes the determination of ends for action (*MS* 384–385). Indeed, a maxim issuing from this power, to the extent that it is a true principle, or starting point, of action, is precisely the act of determining such an end (*KpV* 19, *MS* 385).

This last difference, however, though significant, is not as great as it might seem. For **in the first place, and as will be explained more fully below, Kant does not hold that for an object to be a freely held end at all is for it to be an object of choice (that is, of *freie Willkür*); such an end can be an object of mere wish, and it can even—as in the case of the highest good—be something an individual could not possibly have as an object of choice.** Second, as is reflected in the fact that he views maxims as both principles of action and acts that determine ends, Kant does not take the choice of an end to be anything distinct from the choice of an action—the action of realizing that end—nor to have as its object anything thought by the subject to lie beyond its power as an agent. Indeed, this identification of choice of end with choice of action follows straightforwardly from the consideration that the free power of choice, though distinctive in that, as free, it can be determined by pure reason, is nevertheless a power of choice (*Willkür*) and therefore falls under the definition of the latter as the faculty of desire in accordance with concepts “so far as it is combined with the consciousness of the capacity of its action for the production of the object” (*MS* 213). For this definition implies that any object of choice and hence any end so far as it is chosen is practically represented, in the act of choice, as the effect to be produced through the subject’s action. In other words, the determination of an end in an act of choice is the specification of a practical conception of an *effect* to be produced, a conception that already contains the practical conception of the relation of causal dependence in which the object, the effect, is to stand to oneself, the subject and agent, and this latter conception is just the

practical conception of the *action* whereby that object is to be produced.²² So while the exercise of *freie Willkür* is not confined to the choice of means (that is, to determining how to carry out some action), it does always lie in the choice of action, and since choice is confined to objects deemed to lie within the agent's power, it is always limited, in accordance with judgments of practicability, to actions for which the specification of means is thought possible.

It is also worth noting that where the ends determined through the exercise of the free power of choice are true ends rather than mere means to further ends, the actions therein chosen differ in form from the type of action whose aim is to bring about some change and whose end therefore comes to be after it in time—the type of action, that is, whose bringing about of its end is equally its bringing itself to an end (in time). For—to anticipate a point that will be considered in greater detail below (§III.2)—true ends are so constituted that, even though insofar as they are actual they will come to an end if conditions on which they depend cease to obtain, they never bring themselves to an end, and it follows from this that the action chosen in the very act of determining such an end likewise cannot be the type of action that brings itself to an end. The practical representations of such actions must, therefore, rather be of ongoing activities to be engaged in and so must be “practical principles,” or “general determinations of the will”: they must, in other words, be maxims, for maxims are just such principles so far as they are also determinations of the free power of choice (*KpV* 19, *MS* 226). The actions specified in maxims can accordingly be called practices, or modes of conduct, to distinguish them from actions that unfold in time and whose ends come to be after them in time. In sum, then, choice can be of ends (e.g., happiness so far as it is thought to be within one's power) and of the practices that directly serve them (e.g., the practice of prudence), as well as of the means that are practically represented through the specification, in accordance with the subject's theoretical awareness of existing circumstances and conditions, of how this or that mode of conduct is to be carried out, and which can include actions whose ends follow them in time (e.g., to visit a physician for treatment of an ailment).

²² Cf. A205/B250: “Action already signifies the relation of the subject of the causality to the effect.” Because this conception of the relation of dependence is practical, or efficacious, it constitutes that relation of causal dependence itself.

These caveats having been noted, it will be convenient here to follow the translators' common practice of using "choice" to designate the exercise of *freie Willkür*, reserving "will" for the translation of *Wille* (and *Wollen*, which signifies the exercise of *Wille*). Choice, then, implies intention and even counts as a kind of intention. But it is distinctive in that the efficacy constituting it as intention is the consequence, the natural result, of a judgment determining what it would be good to do. To understand the special character of choice, therefore, we must turn our attention to this judgment lying at its basis.

6. Practical Judgment and Practical Knowledge

It is clear that the action deemed good in this judgment must be the very action chosen in the act of choice resulting from it. As we have already observed, willing in the sense we have just been considering—the sense fixed by the familiar use of the expression "weakness of will"—involves the awareness that to act in accordance with one's will is good and to act against it is bad. So the content of the judgment, what is deemed good, and the content of the choice, what is chosen, are the same.²³ It is also apparent that this judgment is tantamount to the judgment that the action is *choiceworthy*. For in determining that to do a certain thing would be good, and hence that not to do it would be bad, the judgment determines that the former is worthy, and the latter unworthy, of choice, thereby bringing it about, through the efficacy it has in respect of the power of choice, that the action deemed worthy of choice is chosen. This means that the judgment must deem the action good, not just in this or that respect, but on the whole, or on balance, since to judge

²³ Provided, that is, that we think of the intention in which choice consists along the lines set forth above, as constituting a complete practical thought. If, on the other hand, we conceive of intention (or choice) as the practical predicate that figures in such a thought, then of course its content will not be identical with that of the practical judgment; for the practical judgment includes, in addition, thought of the subject of the action. When "intention" is understood in the latter way, we can say that King Francis I of France and Emperor Charles V had the same intention—to rule Milan—even though the practical judgments in which these two brothers determined what they should do were different, indeed were in conflict (cf. *KpV* 28). But since intention is essentially efficacious, and since its efficacy belongs, not to the practical predicate by itself (a mere concept of a possible action), but to the act of practical thought in which it is attached to the subject, intention strictly and properly so called lies in such an act. "Maxim" is similarly ambiguous.

merely that an action is good in some respect is not to imply that its opposite is unworthy of choice, or bad.²⁴ It follows that this judgment amounts to a judgment that one has sufficient reason to do the action, for when we take an action to be good, or choiceworthy, we suppose not only that we have reason—a reason—to perform it (a supposition that would not entitle us to say more than that the action is good in some respect), but in addition that we are not in need of any further reason. The judgment does not simply express the presence of considerations that count or weigh in favor of the action, considerations on which one might rely in practical reflection or deliberation about what it would be good to do; it is rather the result of this reflection, its conclusion. In such a judgment, therefore, the action is implicitly regarded as rationally supported, as falling under the heading “what ought to be.”

Now a judgment that determines what it would be good on the whole to do and through this determination results, wherever possible, in the efficacy of its specification is the sort of judgment in which what Kant calls “practical knowledge”—efficacious knowledge, knowledge of the good, or of “what ought to be”—consists (cf. Bix-x, A633/B661, *KpV* 46). Such a judgment may therefore be called a *practical judgment*.²⁵ As practical, it has the same self-conscious efficacy that belongs to practical thought, the efficacy that lies in the subject’s act of attaching a practical predicate to itself, except that in this case the attachment is an act of cognition of the good, where the consciousness of efficacy, and therein the efficacy itself, originates in the consciousness of goodness.

²⁴In general, no judgment deeming an action good or bad in one respect ever stands in a relation of logical agreement, conflict, or entailment with one deeming it or its opposite good or bad on the whole or in some other unrelated respect; such relations are possible only where both judgments regard the action in the same way. To say that an action is good *on the whole* is not, of course, to say that it is *unconditionally* good.

²⁵This is not an expression Kant frequently employs, but when he does use it, he shows no hesitation in doing so (see, e.g., *G* 403–404, *KpV* 57–58, *KU* 280, 20:246), and other terms that often figure in his discussions of practical cognition, such as *Satz* (“proposition”) and *Grundsatz* (“fundamental proposition,” “principle”) (see, e.g., *KpV* 19, 31; *G* 419–420), signify, in his usage, either judgment (*Urteil*) or judgment of a certain type. Kant states at various points during his critical period that a *Satz* is an assertoric (as opposed to problematic) judgment (24:934; cf. 8:193–194n, *L* §30n3), though as Günther Patzig has noted, in the first *Critique* he seems to use *Satz* as an equivalent to *Urteil* (“Die logischen Formen praktischer Sätze in Kants Ethik,” in G. Prauss, ed., *Kant: Zur Deutung seiner Theorie von Erkennen und Handeln* [Köln: Kiepenheuer & Witsch, 1973], 219–220). Practical judgment and practical knowledge will be examined further in §IV.

Indeed, it is not difficult to see that it is in virtue of its possessing such efficacy that the practical judgment grounding choice must, as we just noted, represent the action as good on the whole, not just in this or that respect. For as self-conscious, this efficacy must lie in the judgment's *understanding* itself to be sufficient to determine the subject to act accordingly—sufficient, in other words, to determine the subject to choose to do the action deemed good in the judgment. But because a judgment asserting no more than that an action is good *in a certain respect* includes an awareness of a diversity of possible respects in which the action may be regarded and leaves open the possibility that the action might be recognized to be bad in some other respect, it cannot understand itself to be sufficient to determine the subject to act. For were the knowledge that the action is in one respect good a ground sufficient to determine the subject to choose to do it, then a recognition that it is in some other respect bad would likewise be a ground sufficient to determine that subject to choose not to do it; but it is inconceivable that two actual opposing grounds should both be sufficient. As practical, or self-consciously efficacious, therefore, the judgment must be that the action is good on the whole.

This argument raises a question, however, about the way in which practical judgments are efficacious. It was implicitly suggested in the final step that it is possible for a subject's choice to be determined not only by positive judgments but also by negative ones, in which actions are deemed bad, or wrong, and that when this happens the subject is determined to choose not to do the action rejected in the judgment. Now it does seem plainly correct that negative judgments have an efficacy that can determine choice; judging an action to be bad is not simply refraining from judging it to be good. The earlier characterization of a practical judgment's efficacy, however, identified the content of a practical judgment with the content of the choice it can determine; that account was tailored to the positive case and may seem not to fit where the judgment is negative. For if the content of a positive judgment is what it deems good, then it would seem that the content of a negative judgment is what it deems bad, yet this is by no means the content of the choice the judgment can determine. Such a judgment's efficacy can never result in the choice of the action deemed bad in the judgment; it can only issue in the determination *not* to do the thing in question. My recognition that it would be wrong to take an extra portion of ice cream could never ground

or explain my choosing to do so, but only my choosing to refrain (and thereby also my not choosing the action). To clarify the specific way in which such judgments are practical, further consideration of the constitution of practical judgments is required.

It belongs to the idea of a practical judgment as a judgment involving the notion of sufficient reason—what Kant calls a determining judgment—that such a judgment includes a (possibly tacit) *comparison* between an action (to \emptyset) and the omission (not to \emptyset) that stands opposed to it as its logical complement. This comparison presupposes in turn a division in thought of the space of possible content for practical cognition into two complementary, or logically opposed, contents. Because it includes this comparison, the judgment (the affirmation or negation) regarding the one member of the division is *in relation* to the other and so is *identical* with a judgment of opposite quality (negation or affirmation) regarding the latter. In other words, the act of judgment in practical cognition presupposes an understanding of a relation of logical interdependence among the opposing possible contents, a relation on account of which the actuality of the one in affirmation—that is, its actuality as a content of practical cognition—is identical with the other's being excluded from such actuality and hence is also identical with the latter content's loss of the possibility it originally had of being a content of practical cognition. The excluded content accordingly stands to the actualized one as patient to agent: the advance from possibility to actuality in the one is the loss of possibility in the other. Thus the act of affirmation establishing the one content as a content of practical cognition is equally the negation in which the opposing content is excluded from the content of such cognition. An act of practical judgment therefore always has two sides, which make up, so to speak, the different faces it bears to the complementary contents. But since, notwithstanding the logical interdependence of these contents, the positive content (to ϕ) is prior in thought to the negative (not to \emptyset), the face of the judgment that contemplates the positive content is likewise primary and stands originally in the foreground of practical consciousness, while the other, opposite to the first in respect of both content and quality, is secondary. Though practical judgments always have this double-sided character, we can nevertheless classify them as positive or negative in quality by reference to the quality of the primary face, the one considering the positive content: according as this side is positive or negative,

determining
judgment &
sufficient
reason

so is the judgment itself. Thus a positive practical judgment determines that it would be good to act a certain way, yet is also secondarily a negative judgment, that it would be bad not to act that way; and similarly a negative practical judgment determines that it would be bad to act a certain way, but is secondarily a positive judgment, that it would be good not to act that way. The efficacy of a practical judgment, however, always belongs to its positive, or affirmative, side, the one employing the concept of the good. So what we call a negative practical judgment is one that deems a certain way of acting to be bad, but whose efficacy operates through its positive guise, working to ensure that the subject does not act that way, despite any opposing inclination to do so that may be present and that may indeed be what first prompted consideration of the way of acting subsequently deemed bad in the practical judgment.

Although the comparison between opposing contents that figures essentially in a practical judgment ensures that the concepts of good and bad are coeval, positive practical judgments, in the sense just indicated, are nevertheless basic. Their primacy is evident from the function of negative judgment. As a judgment that deems a certain way of acting to be bad, a negative judgment is the aversive act of the capacity for practical cognition in response to the consideration of a way of acting that, at the prompting of sensible desire, is put forward in practical reflection as a possibly worthy candidate for choice. Such a judgment thus functions to ward off bad or wrongful practical judging and is therefore possible only on the supposition that the capacity for practical knowledge can be exercised wrongly or incorrectly, a supposition that—as will be explained below (§IV.4)—does not belong to the original conception of such a capacity. Indeed, strictly and fundamentally speaking, the object of practical cognition comprises solely contents of positive practical judgments. So while there is a sense in which one can say that the content of a negative practical judgment is what it deems bad, we must also say that so far as this judgment is viewed as practical knowledge, its content is rather the complementary content deemed good in that same judgment. Though Kant identifies the “sole objects” of practical reason as the good and the bad (*KpV* 58), he obviously recognizes that the bad counts as such an object only in a negative and secondary sense, as an object of aversion. He repeatedly characterizes practical knowledge as knowledge of what ought to

be, never as knowledge of what ought not to be, says it works to make its object actual, not to prevent it from becoming actual, and holds the “unconditioned totality” of pure practical reason’s object to be, not the complete good along with the complete bad (an absurdity to be sure), but simply the complete good (the *summum bonum*) (*KpV* 108, 110). (The idea that the good is the true object of practical knowledge is also, of course, central to the interpretation here being proposed, which traces badness to practical incognizability.) In what follows, attention will be focused on the primitive case, where the judgment is positive: as an exercise of the capacity for practical knowledge, a practical judgment is to be conceived, in the first instance, as knowledge of the good.

7. *Some Distinctions*

The ideas of practical knowledge and practical judgment just introduced will lie at the center of our concern in the pages ahead and will be further elaborated as the discussion progresses. But in view of their importance, and because it is possible for them to be mistaken for certain other concepts investigated in practical philosophy, a few comments are in order here to mark the differences before proceeding further.

First, the expression “practical knowledge” is ambiguous. There are two other familiar senses from which the one of interest to us here should be distinguished. One of these uses is long-standing and noted by Kant himself (*L* 86–87, *KpV* 26n). In the sense introduced above, “practical knowledge” means knowledge of what ought to be and stands over against “theoretical knowledge” broadly understood, as signifying knowledge of what is. But as Kant points out, the expression is also employed in a different sense, to mean theoretical knowledge so far as it has an actual or possible use in relation to practical knowledge in the first sense. Knowledge that corn is nutritious, that hammers drive nails, or that wet surfaces are often slippery is in this sense practical, for we may rely on it in determining what we should do. In this second sense, “practical knowledge” covers all the theoretical knowledge that belongs to the productive, technical, or applied sciences, in contrast to theoretical knowledge that, having no such use, counts as speculative, “merely

theoretical.”²⁶ This second sense is kindred, of course, to the sense “practice” takes when set over against “theory” to mark a contrast between abstract study and concrete experience or between the ideal and the practicable.

The other contrasting sense of “practical knowledge” has gained currency more recently. In the sense introduced above, “practical knowledge” signifies the knowledge one has insofar as one “knows what to do”—that is, knows what it would be good to do. In recent years, however, it has often been used to refer to the knowledge one has insofar as one “knows what one is doing,” where this knowledge is related to what one is doing not by observation but by being the self-understanding that constitutes the efficacy of practical thought in general—that is, by being such thought’s understanding itself to be sufficient to effect the action—and where this understanding lies in the exercise of a knowing *how* to do what one is doing.²⁷ At root, the two senses differ in that one picks out the exercise of a capacity for productive cognition (“desiderative reason”), the other the exercise of a capacity for cognizant production (“ratiocinative desire”). On the first way of understanding the expression, practical knowledge is always a judgment of the sort just described, on which choice can be based; on the second, it is always an instance of the practical consciousness distinctive of intention. Insofar as the possibility of intention does not depend on the capacity to know the good, these two senses are distinct. Hume’s sensible knave and the clever but unwise scoundrels Aristotle describes as “capable of anything” have practical knowledge in the second sense, though the choices of such individuals would not be grounded in practical knowledge in the first sense, knowledge of the good.²⁸ But practical

²⁶ It may be worth mentioning that Kant’s three postulates of pure practical reason do not fall under the heading of practical knowledge in either of the two senses just noted. Since these postulates, which assert the freedom of the will, the immortality of the soul, and the existence of God, are propositions about “what is” rather than about “what ought to be,” they are *theoretical propositions*, yet they are not *theoretical knowledge*, since their objects cannot be given in experience. Hence, in contrast to knowledge that is practical in the second sense, they provide no independent positive support for practical knowledge in the first sense, but are rather themselves supported by their relation to the latter, amounting to practically rational belief, or faith (cf. *KpV* 122, 126).

²⁷ This seems to be the sense G. E. M. Anscombe has in view in her discussion of practical knowledge in *Intention* (Ithaca, N.Y.: Cornell University Press, 1957), §§28–32, 45–48.

²⁸ See Hume, *An Enquiry Concerning the Principles of Morals* § IX, Pt. II, and *Nicomachean Ethics* VI.xii; cf. also *R* 26n.

knowledge in the first sense does require, as a condition of its very possibility as knowledge that is practical, the real *possibility* of successful execution, and therefore insofar as it has perfection in its practicality—that is, perfection in respect of efficacy and hence in its capacity for execution—it is also knowledge of *how* to do what it would be good to do and thereby, in the exercise of the latter, practical knowledge in the second sense as well. Thus *efficacious knowledge* (practical knowledge in the first sense) can perfect itself in respect of its efficacy through developing its implicit understanding of itself as efficacious into *knowledge of its efficacy* (practical knowledge in the second sense). Since what the second of the two senses distinguished picks out is better described if we speak of it, not as practical knowledge, but as practicality's self-cognizance, and since such cognizance belongs, or can belong, to practical thought generally and therefore to bare practical thought as well as to practical knowledge in the first sense, we shall in what follows reserve the expression "practical knowledge" for the first of the two meanings distinguished. It is practical knowledge in this sense that Kant has in mind and that is our concern here.²⁹

Second, while practical judgments can be described as judgments in which certain actions are deemed good or bad, they should be distinguished from the judgments of appraisal that figure in our approval or disapproval of particular actions and conduct. Often discussed under the heading "moral judgments," such appraisals are of central concern in the empiricist tradition in moral philosophy that traces back to Hutcheson and Hume, who interpreted them as deliverances of a moral sense and made them the starting point for their philosophical reflections. Since in judgments of appraisal actions are regarded as expressions of choice and character, these judgments bear a certain resemblance and even an intimate relation to practical judgments. But they are also inherently different from them. As judgments with an efficacy suiting them to be determinants of choice, practical judgments concern things a person is regarded as at least in some sense *capable* of doing, whereas appraisals of action concern the *actual* things persons do. The former pertain to "doables," or things a person can do, could do, or could have done; the latter to "deeds," things that are, have been, or will be done.

²⁹ We shall, however, revisit the notion of practicality's self-cognizance when we take up the idea of practical self-sufficiency (§III.5).

Just as actuality presupposes possibility, so judgments of appraisal presuppose practical judgments. Indeed, they presuppose them in two ways: not only in that the actions that are the objects of appraisal themselves issue from practical judgments made by the agents who perform them, but also in that practical judgments are involved in the appraisals themselves and are even the source of the efficacy these appraisals have whereby they can find expression in praise or blame. When we learn of the exemplary conduct of some individual, even a complete stranger far removed from us, such as the man Kant describes from the court of Henry VIII who, though threatened by his prince with death, refused to calumniate the innocent Anne Boleyn, feelings of approval and admiration are awakened, whose intensity only grows the more closely and attentively we contemplate the case.³⁰ In this approval, however, we presuppose the independent actuality of the action (hence the possibility of observing it in experience) and regard it as expressing a practical judgment. And we deem the action good by comparing it with, and finding it to be subsumable under, the predicate that figures in our own practical judgment specifying what the man should do, what it would be (or have been) good and right for him to do, a judgment that accordingly provides the basis of our approval and the source of the latter's efficacy, through which it can be expressed in our praise of the action. A practical judgment, in contrast, does not presuppose the independent actuality of the action it has as its object, but rather necessarily precedes that actuality, containing the subject's understanding of itself as the cause that, through that very judgment, works so far as it can to bring the action about, to make it actual. Thus whereas it is in judgments of approval that we *experience* the good, it is through practical judgments that we *produce* it.

³⁰ *KpV* 155–156. Although judgments of appraisal are based in practical judgments and hence in the exercise of the capacity for practical knowledge, it does not follow that they are themselves cognitive judgments. While Kant criticizes accounts that trace judgments of approval and disapproval to a moral sense (cf. *MS* 400), he agrees with the empiricists to the extent that he takes the approval or disapproval to lie, not in thinking, but in a feeling of pleasure or displeasure. A judgment of this sort thus counts as an aesthetic judgment in his technical sense, though on account of its distinctive dependence on practical judgment he characterizes it more specifically as an “aesthetic-practical judgment” (20:230–231). In fact, this feeling depends on both practical and theoretical cognitive judgments, being just the inward manifestation of their mutual furtherance or conflict. An account of aesthetic judgments of the good is outlined in §§4–5 of the *Critique of Judgment*.

Though judgments of approval do constitute the culmination of practical judgments, the consummation of their efficacy, it follows from what has just been said that judgments of appraisal are secondary to practical judgments in the sense that they depend on the latter as well as on theoretical judgments for the terms that figure in the comparisons they involve. We could even say that a judgment of approval stands to its corresponding practical judgment as effect stands to cause, since it is in such judgment that the action issuing from the practical judgment naturally terminates. It belongs to the very idea of practical cognition's efficacy, however, that in the first instance the effect is understood through the cause, not the cause through the effect. Our concern in what follows will accordingly be confined to practical judgments, which constitute the fundamental use of the concept of the good. We shall therefore leave to one side all employment of the concepts of good and bad, or of right and wrong, that presupposes the actuality of the object brought under these concepts, including not only moral judgments as usually understood and the assessments of character they support, but also a wide range of reactive attitudes, such as gratitude and resentment.

Third, in view of another countervailing current of empiricism in modern and contemporary moral philosophy, it bears emphasizing that a practical judgment is not a relative judgment in which an action or end is deemed good *according to* the person making the judgment. If it were a judgment of this sort, we would have to say, with Hobbes, that the words "good" and "bad" "are ever used with relation to the person that useth them: There being nothing simply and absolutely so."³¹ There would be no communicability, no possibility of different subjects' agreeing in their judgments, nor therefore even of their disagreeing, even though the possibilities of harmony and conflict in the actions that might result from such judgments—possibilities traced with keen discernment by Hobbes himself—would of course necessarily remain. This impossibility of agreement and disagreement would stem, not simply from an absence of a common subject matter, but solely from the relativity of the judgments themselves. The judgments might concern the very same action—your practicing on your trumpet, for instance—yet

³¹ *Leviathan*, chap. 6.

still no agreement or disagreement would be possible: your judgment would specify only whether it would be good according to you, mine only whether it would be good according to me. But agreement in judgment must be possible if knowledge is to be possible, and therefore a practical judgment, as the sort of judgment in which practical knowledge consists, must be one with which different subjects can agree or disagree. Similarly, the concept of the good, as the concept of the object of such knowledge, carries an implicit recognition that agreement about the good must be possible, a recognition of its own communicability that distinguishes it from the concept of the agreeable (*KpV* 58).

Nor do agreements and disagreements among practically judging subjects involve judgments that, though not relative in themselves, or in respect of their content, are relative in their efficacy, or in respect of their capacity to move their subjects to act. Such judgments would in effect employ a relational concept of the good—a concept containing reference to an indeterminate *relatum* (some individual, for instance, or type of individual, or group)—and be efficacious in the judging subject only if the specification of the *relatum* that enables the concept to be applied fixes it to be that subject. If we express this concept by “good for *s*,” where “for *s*” marks the concept’s reference to some yet to be specified practical subject, its indeterminate *relatum*, we can say that a judgment employing it would have efficacy in respect of the judging subject only where *s* is specified as being that very subject. We might imagine a land populated by practical subjects (perhaps in a Hobbesian “natural condition”) who employ this concept, sometimes agreeing, sometimes disagreeing about how, for given specifications of *s*, it is to be applied. If one of them were to agree with a second that it would be good for the latter were the latter to develop a certain talent, then this judgment would be efficacious in relation to the second but practically inert in relation to the first. It was such relative efficacy that was supposed to belong to the judgment of the rational egoist we briefly considered at the outset in connection with the common objection raised against Kant’s derivation of the formula of universal law (§I.1). Put in terms of the relational concept of the good, what the egoist was prepared to allow was that all rational beings could agree that for any rational being *s*, *s*’s acting rationally would be good for *s*. Thus, for each specification of *s*, the resulting judgment instantiating this proposition could move the rational being to which it refers to act accordingly, but could not move

any other to lift a finger to support or to further that individual's rational action. Such agreements, however, are not genuinely practical. A judgment that can be shared by different subjects yet have efficacy in one of them only insofar as it lacks it in others cannot count as practical knowledge. It is excluded by the very idea of such cognition. As knowledge, practical knowledge must be shareable, and as practical, it must be *itself* efficacious and hence efficacious in everyone who possesses it (even if this efficacy in individual subjects is not always direct and even if the actual effects in different subjects vary in degree); its efficacy, therefore, is likewise essentially shareable. Nor, consequently, would these nonpractical agreements and disagreements involving the relational concept generate the distinctive familiar feelings that are connected with agreements and disagreements involving the concept of the good and that reflect its presence in our practical thinking. With respect to questions about the applicability of this concept, disagreement can arouse sentiments of opposition and indignation, distrust and enmity, and agreement can inspire feelings of goodwill and affiliation.

Of course, practical disagreements often seem difficult to resolve. This is especially true in the case of disputes that are proper to practical judging rather than merely incidental to it. Incidental disagreements can arise where practical judgments depend on empirical theoretical judgments of practicability. In such cases, conflict among the former can depend on conflict among the latter. A disagreement over whether it is good to smoke after meals, for example, might derive from conflicting theoretical judgments about the effects such action has on our health; one person may think it harmful to the lungs and the heart, while another supposes it to be an aid to digestion (as it was once advertised to be). In such instances, the disagreement does not depend specifically on misemployment of the capacity to know the good, but derives instead from error that is already present in the conditions in which the exercise of this capacity takes place and on which it in part depends, conditions in the form of the subject's theoretical judgments pertaining to the circumstances and effects of action. Strictly speaking, disagreement of this sort lies in theoretical judgment and for this reason falls outside the central focus of our investigation (cf. *KU* 172). The type of disagreement that can seem particularly intractable lies originally between practical judgments, as in the case of the two brothers who want "the same thing"—Milan (*KpV* 28). Such disputes do not reflect error in some other sort of judgment and so cannot be traced to the condi-

tions in which the capacity to know the good is exercised, though they may reflect the influence of certain inclinations belonging to those conditions. They spring directly from practical judging itself.

The apparent intractability that often marks disputes of this latter sort may lead some to doubt whether rational resolution is possible, or even to deny, as Hobbes did, that the disagreements are genuine. But their deep-seated *presumptive* reality is beyond dispute, and as long as it is, the decision to adopt a skeptical attitude toward them should by rights defer to attempts to understand them as what they plainly purport to be, genuine practical disagreements about the good. It will accordingly be taken for granted in this study, as it is in Kant's moral philosophy, that practical agreements and disagreements are real relations, constituting and shaping our practical lives.

8. Intention, Choice, and Practical Judgment Compared and Related

Let us return to the path we had been following. We have seen that choice, in the sense indicated earlier, consists in an intention that is also a practical judgment: in choosing, one specifies what one means to do by determining what it would be good to do. Since not every intention is a practical judgment, nor every practical judgment an intention, choice cannot be identified with either; it lies rather at their intersection.

Thus, choice is not simply identical with intention. When I am indifferent between two ways of carrying out something I have chosen to do and arbitrarily settle on one of them, I specify what I mean to do, but not by a practical judgment.³² For here there is no question of better or worse, yet this question always has a bearing in practical judging concerning the actually achievable good. Unlike intention, then, choice cannot be arbitrary in this sense. In choice, practical self-specification is also self-determination.

³²This is not to deny that my indifferently settled-on specification is effected through practical judgment. In choosing something as an end, I deem it good, practically representing myself as having to do something sufficient to realize it. So given that one of my two equally viable alternatives must be adopted if I am to realize the end, practical reason demands that I pick one, and my intention to pursue *this* option, not the other, is thus the result of my practical recognition that I must do something sufficient. Here there is rational efficacy without rational determination: practical reason brings about the act of specification even though it does not determine the latter's content.

Nor is choice simply identical with practical judgment. I might judge that I should have saved more of my earnings when young, or that it would be good for you to develop a certain talent. But neither of these judgments constitutes a choice, since in neither do I specify what I mean to do.³³ So choice is related to practical judgment as species to genus.

Practical judgments are in fact the basis of a variety of practical attitudes, which arise in the mind when these judgments are accompanied by certain theoretical judgments concerning their objects. One might, for instance, wish that one had not squandered one's resources early on in life, be glad that one's children are making more sensible choices, and hope that their children will do the same. Each of these attitudes has a practical judgment lying at its basis, but none of them amounts to choice, and they differ from one another and from choice on account of the different modalities in which their objects are represented, modalities that underlie in turn the different modes of feeling they involve. While it is not necessary for our purposes here to arrange these attitudes in a systematic classification, the ones just mentioned might be related along the following lines. If a (positive) practical judgment concerns something thought to be actual, it is the source of gladness, or pleasure (for instance, approval of another's action); otherwise it is the source of what we might call wanting (in a sense that implies lack as well as desire). If such wanting concerns what is thought to be impossible, it is mere wish; if it pertains to what is not impossible, then it is hope so far as it concerns what does not lie within the person's control, and choice

³³ Judgments by which one can specify what one means to do must determine actions that are not thought to lie beyond the judging subject's power of agency and must have as their subject either oneself or an agent of which one is a part. This study, however, will for the most part not be concerned with agents that have agents as parts. (Any agent constituted through an act wherein different agents join together as one is dependent on those agents and is therefore at least in that sense secondary in relation to them. Our main concern here is with the primary case and hence with singular rather than plural practical judgments.) It is possible, of course, for a practical judgment to be efficacious indirectly, or in a manner other than the immediate way, characteristic of choice, that we have been focusing on here. My judgment that it would be good for you to develop a certain talent might prompt me to encourage you to do so or lead me to help provide you with an opportunity. (Here of course "good for you" does not express the relational concept of the good described earlier.) And as was noted earlier, practical judgments are the source of the efficacy of judgments of appraisal, through which they can find expression in praise and blame.

so far as it concerns what does not lie beyond the person's power.³⁴ Obviously the different modalities that distinguish these practical attitudes are not internal to the contents deemed good in the practical judgments. If they were, then what one chooses to do could never be the same as what one is glad or pleased to be doing, mere wishing would involve the judgment that the impossibility of what is wished for is good, and so forth. As we noted, practical judgments always concern "doables," things the person in question can be regarded as capable of doing, at least in a problematic way (that is, prior to any determination of how or whether, in the actual conditions of action, they are practicable). The attitudes just described are distinguished according as these doables are thought, in accompanying theoretical judgments, in one or another of the different modalities just distinguished, as possible or impossible, actual or not actual, and so on. One and the same content is both deemed good and also thought in one or another of these ways.

Practical judgments that concern the past or the actions of others, however, are all secondary to judgments of self-determination, that is, to the judgments constituting choice (and also wish, understood in a sense to be explained below). For the end of practical judgment is the good it determines, and for this only judgments of the latter type are absolutely indispensable. It would be in principle possible for the good to be correctly determined and effectively pursued even were the capacity for practical judgment to be exercised in choice alone, without also being employed in judgments about what it would have been good to do and about what it would be good for others to do. But it would not be possible for the good to be pursued to any extent at all, much less to be realized, if the capacity for such judgment were never exercised in choice but only in judgments concerning other persons and the past. This asymmetry does not imply, however, that this capacity's employment in judgments of the latter sort is a merely arbitrary or gratuitous extension of its employment in judgments of self-determination. Nor is it to be inferred that secondary judgments are merely needed for remedial purposes, as helps and supports by which ill-considered choices might

³⁴ Since practical judgments are double-sided, these attitudes have counterpart modes of feeling supported by the complementary (negative) judgments. Regret, for instance, is complementary to mere wish ("I regret having frittered it all away; if only I hadn't"), and fear accompanies hope ("I'm afraid you may fail, but hope you can bring it off").

be corrected and improved, or by which persons might be assisted in carrying out their chosen actions. On the contrary, as we have already noted in effect and shall later see in greater detail, the capacity to make judgments outside choice is integral to the capacity for practical judgment, in that such judgment, as the type of judgment in which practical knowledge consists, presupposes the possibility of agreement in judgment among diverse practically judging subjects, in which different judgments second and confirm one another. The asymmetry does nevertheless indicate that choice—along with wish—is the original type of practical judgment and hence the primary species; when we turn, therefore, to an examination of fundamental practical judgments in the following section, the other types will for the most part be left to the side.

In sum, intention is action specification that is efficacious, practical judgment is action specification that is rationally determined in practical cognition, and choice is action specification that is efficacious through being determined in such cognition. Choice can accordingly be considered from two sides: viewed from above, or heard as it were in the active voice, it is a rationally determined specification whose determination issues in its efficacy (“desiderative reason”); regarded from below, or in the passive voice, it is an efficacious specification whose efficacy arises from its rational determination (“ratiocinative desire”).

9. The Will

To arrive at Kant’s conception of the will (*Wille*) we need only turn to the conception of practical judgment on which the concept of choice depends. In determining what it would be good to do, practical judgment represents the action it specifies as rationally supported, or as one for which there is sufficient reason, and it results, wherever possible, in the efficacy of its specification. It thus consists in the exercise of practical reason, the capacity for practical knowledge. Kant identifies the will with practical reason (*G* 412, *MS* 213), so practical judgment is just the exercise of the will in his sense. The content of the will—what is willed—is the action or end that in the exercise of the will is judged to be good, or (what comes to the same) is regarded as an object of practical knowledge.

Since choice—free choice—is related to practical judgment as species to genus, the free power of choice stands in the same relation to the will; in Kant's words, it is "contained under the will" (*MS* 213). So all exercise of the free power of choice is also exercise of the will. But as the capacity for practical knowledge, the will has a wider purview than does the power of choice, extending to objects whose attainment lies beyond any person's power; it reaches even to the highest good, the totality of all objects of practical knowledge. Kant can accordingly characterize the will as "a capacity either to produce objects corresponding to representations or at least to determine itself, that is, its causality, to the effecting of them (be the physical capacity sufficient or not)" (*KpV* 15). On account of this broader scope, the possibility is not ruled out that the will of one person might directly contradict that of another ("I should" / "No you shouldn't"), whereas different persons' choices can never stand in such a relation, not even where the actions they represent are actually incompatible or even essentially opposed, as in cases of direct competition ("I'm going to defeat you" / "No, I'm going to defeat you"). And by the same token, it is possible for the wills of different persons to be united in perfect agreement, sharing "one and the same object" (*KpV* 28) in an *identity* of practical thought that can never be found in their individual choices, however harmonious the ensuing actions may be.

III



Fundamental Practical Judgments: The Wish for Happiness

1. Choice and Wish

Practical judgments can depend in various ways on other practical judgments. Choices, for instance, when considered against the backdrop of the theoretical cognition they presuppose, characteristically stand in relations of interdependence with other choices. For the object of choice is always either some end or the means recognized to be necessary for its attainment, and in practical cognition the choice of one of these objects is always in tandem with the choice of the other, as is reflected in Kant's proposition that "who wills the end wills (so far as reason has decisive influence on his actions) also the indispensably necessary means to it that are within his power" (*G* 417). This relation is not wholly symmetrical, of course, since in the one direction the dependence is in respect of practicality, while in the other it is in respect of practicability; yet neither choice can stand without the other, as they are bound together according to the twofold condition of rationality for intending, which requires, as we have seen, that the end be possible and the means be sufficient (§II.4).

But such reciprocal relations *between* choices are not the only ones in which the practical judgments that constitute choices are involved. As judgments determining what is choiceworthy, or what it would be good on the whole, or on balance, to do, choices also always depend nonreciprocally on practical judgments of another *type*, in addition to depending (at least where they are not also determinations of what is unconditionally good) on a certain sort of theoretical judgment. They depend on the

former insofar as they are determinations of the *good*, and on the latter insofar as they are also no more than determinations of what is good *on the whole*.

As a judgment about what it would be good *on the whole* to do, choice depends on the subject's judgments of practicability. For the qualification "on the whole" indicates that the judgment is made in the light of certain discriminations between what is practicable, or feasible, and what is not, discriminations reflecting the subject's awareness of limits on its capacity as an agent. Judgments of practicability are not themselves practical (efficacious), but are rather empirical theoretical judgments that specify what one is able to do (if one wishes), marking the extent of one's powers of agency.¹ These judgments make choice possible and limit it in accordance with the rational requirement of practicability, which applies, as we noted earlier, to all intention and hence to choice in particular, as a kind of intention. This requirement, however, obviously falls far short of being a sufficient criterion for determining an action to be good on the whole, or choiceworthy. There are always many different ways of acting that are consistent with the conditions of practicability, but not all of them are good on the whole. Indeed, the notion of practicability, taken by itself, involves no thought of anything good at all.

The judgments in which choices consist, therefore, as judgments that determine what it would be *good* on the whole to do, also depend on judgments of another sort, in which certain objects are deemed good, not on the whole, or on balance, but simply, or immediately, prior to consideration of whether they are actually attainable and such as can be

¹ Here and throughout in this study, "theoretical" is used in Kant's usual sense, to mark the distinction between representation of what is and representation of what ought to be. Theoretical judgments accordingly rely on sensation and perception, not feeling and desire, in determining the concept of reality. (This contrast will be considered more fully below in §IV.6.) Strictly speaking, the empirical theoretical judgments of practicability on which choice immediately depends are not *purely* theoretical judgments, but rather applications of the latter so far as they bear on the exercise of choice. For since they specify one's limited powers of agency, they are based on the concept of *oneself* as a subject with such powers—a concept that first arises in relation to one's *practical* self-conception, in that choice, as a form of intention, immediately engenders expectation, which presupposes in turn the theoretical judgment in which one deems *oneself*—the very subject that constitutes itself as practical in practical self-consciousness—as capable of the intended action. On account of this relation to choice and action, judgments of practicability are more suitably characterized as beliefs than as theoretical cognition in the strict sense.

included among the jointly practicable goods. For such judgments determining what is simply good must already be made in order for it to be possible to make the comparisons between such goods—comparisons to determine which among them are better, which more easily attained, and so forth—that become necessary, once the empirically specified limits of practicability are taken into account, in order to judge what it would be good on the whole to do. We have already characterized judgment determining what it would be good on the whole to do as *choice*; we can now characterize judgment concerning what is simply good as *wish*, and thereby succinctly express the way in which such judgment differs from yet is related to judgment of the former sort.²

These judgments concerning what is simply good are also practical, at least *in potentia*, and their practicality shows itself in two respects. First, they are always efficacious in that a specification of what is simply good *becomes*, when that good is seen to be practicable, a specification of what it would be good on the whole to do. The underlying presence of such judgments and this aspect of their potential practicality is vividly brought to light in the story of the ring of Gyges, recounted in Plato's *Republic*.³ When the shepherd who found the ring discovered that it

²Though wish (*Wunsch*), like choice, is said by Kant to be “contained under the will,” it differs from choice in that it is “not connected” with “the consciousness of the capacity of [one’s] action to produce the object” (*MS* 213). This characterization lends itself to wider and narrower interpretations. Sometimes Kant uses *Wunsch* in a way that calls for a narrow reading, one on which the object of wish is taken to lie beyond one’s power; we might speak of wishes in this sense as “mere wishes” (cf. *KU* 177n) to indicate that they are wishes and nothing more (i.e., not also choices), or as “empty wishes” in cases where they *cannot* be anything more since their object is necessarily beyond one’s power: “If only Jupiter would restore to me the past years” (*KU* 178n). But on other occasions—e.g., when happiness is said to lie in everything’s going “according to wish and will” (*KpV* 124; cf. *G* 418, *MS* 387, 480), or when a wish for happiness is attributed to all humans (*G* 418, *MS* 387)—Kant’s usage suggests a wide reading, on which in determining whether something is an object of wish we leave aside the question whether one has the capacity to produce it. It is in this latter sense that practical judgment concerning what is simply good amounts to wish. In this wide sense, a wish can also be a choice; indeed, being always at least potentially practical, it *will* be, wherever practicability allows. Expressed in other terms Kant uses, the difference between wish in this broad sense and choice is just the difference between the inner and the outer employment of the free power of choice (*MS* 214).

It is possible, of course, to apply “wish” and “choice” more broadly than will normally be done in the present study, where for the most part they will be used to pick out types of practical judgment. In a broader usage, they can be applied also to bare practical thought, in accordance with the distinction drawn earlier between problematic and full-fledged practical thought (§II.4).

³ 359b–360b.

enabled him to make himself invisible, the presence within him of a (supposedly natural) wish to get the better of others was immediately revealed; though rendered largely dormant and invisible in ordinary circumstances by the laws of society, which forced him to treat others justly and equally, this wish converted itself into a plan of action as soon as the ring's extraordinary power was noticed. Judgments concerning what is simply good also show themselves to be efficacious in a second way, in that even when their objects are not known to be practicable, such judgments naturally tend, at least when those objects are not deemed *impracticable*, to generate deliberation, or even action, directed to the discovery of a way in which they may be practicable. And even where the objects are thought to be unattainable, the same efficacy can show itself in a tendency to turn one's attention to those objects and to dwell on them in thought.

Like all other practical judgments, these judgments involve the idea of sufficient reason. Holding something to be simply good, or worthy of wish, involves supposing that, provided the thing is practicable, it is choiceworthy, or good on the whole; it involves supposing, therefore, that, if practicable, no further reason is needed to pursue it—no further reason, that is, beyond the basis one has for deeming it simply good in the first place. And to have a sufficient basis for deeming the object in question to be simply good, no more is required than that, in addition to relying, as it must, on certain material conditions, which lie in the awareness of the object as agreeable, as something one finds gratifying or enjoyable, the act of practical judgment also satisfy the formal conditions of practical knowledge, by being in accordance with the general form of knowledge of the good. We shall undertake to articulate these formal conditions in due course (§§IV–V), though we may certainly presume that they would not be satisfied if the judgment's object were the one figuring in the above example, that of getting the better of others, or maintaining the upper hand over them, an object characteristic of the passions for power, honor, wealth, and the like. For the present it will suffice simply to say, in view of the points just noted, that these judgments determine the good insofar as they are in accordance with the following general practical principle: *To the extent that it can be an object of practical knowledge, what one finds agreeable is simply good.*

Now as will be explained more fully in the discussion to follow (§III.4), the practical judgments in which a person brings specific agreeable

objects under the concept of the simply good all depend on a primitive act of practical judgment, in which what one finds agreeable *in general* is made into an object of one's will—whether absolutely or only in such a way and to such an extent as is compatible with this judgment's having the form of practical knowledge—through being brought under this concept of the simply good. And like any other judgment, this general judgment amounts to practical knowledge—knowledge of the good—only insofar as it satisfies such cognition's formal conditions. It amounts to practical knowledge, in other words, just insofar as it is identical with the general practical principle just stated, or rather with its instantiation in one's own case.

2. *The Agreeable and the Good*

Since Kant is known for the importance he places upon strictly distinguishing between the agreeable, as the object of inclination, or sensible desire, and the good, as the object of the will, or practical reason, any suggestion that these are so intimately related as is implied by the principle above may seem to represent a substantial departure from his doctrine. The good, he maintains, being an object of reason, is judged “through concepts, which allow of being universally communicated,” whereas the agreeable is judged “through mere sensation, which restricts itself to individual subjects and their receptivity” (*KpV* 58).⁴ We can appreciate the contrast Kant has in view here if, bearing in mind the points noted earlier concerning the universal communicability characteristic of practical judgments determining the good (§II.7), we briefly consider judgments of the agreeable, where such communicability is altogether lacking.

To judge an object to be agreeable is to characterize it in respect of the pleasure one feels in one's representation of its existence, a pleasure that lies in sensible (receptive) awareness of the object's affecting one's faculty of desire in affecting one's senses. On account of this basis in feeling, judgments of the agreeable are relative. By such judgments, no two

⁴ Cf. *G* 413. Here “sensation” translates *Empfindung*, which Kant commonly uses to refer to impressions of the outer senses. He points out, however, that this expression is ambiguous (*KU* §3), and in some contexts he employs it in its other sense, to refer to the feeling of pleasure (or displeasure). In the passage quoted, he is thinking specifically of gratification, the type of feeling connected with sensible desire.

persons can either agree or disagree about whether some object is agreeable, though they can, of course, frame a relational concept along the lines discussed earlier and by using it concur on what is agreeable to the one and what is agreeable to the other (*KU* §7). The relativity of judgments of the agreeable is not merely the outcome of an antecedent failure on the part of persons to agree on which objects are agreeable, a failure leading them to withdraw from initial absolute judgments, like exhausted disputants who, settling for a truce in lieu of a resolution, retreat from “is” to “seems.” (Nor is the relativity the result of a failure to agree on which objects are good: the agreeable is not the apparent good; as will be explained below, it is the material condition of the simply good.) Rather, the relativity is intrinsic to the agreeable, belonging to its very form. Even if every person found the same things agreeable, and even if these things were the same in a sense strict enough to exclude objects the sharing of which would imply opposition (such as ruling Milan and getting the better of others), the relativity would not be removed. Even if we all loved gardening, sunshine, and peace and quiet, the agreeable for you would in a crucial sense remain completely different from the agreeable for me; for the ground of your pleasure would still be a representation in you, while the ground of mine would be a representation in me. And since sensible desire has its own cause for its object, this difference would be reflected in our inclinations as well: yours would find satisfaction only in your experience of the object, mine only in mine. For such pleasures and desires, a common ground can never be found. They are rooted in the very conditions of receptivity under which each individual person constitutes itself as distinct from all others. Whether I am enjoying an apple or making a judgment about the agreeable in general, I am thinking of what is agreeable *to me*, or of what I *myself* find enjoyable, and it is in my own enjoyment of the object that my desire terminates. So what Hobbes said of “good” and “bad” Kant sees as holding rather of “agreeable” (*angenehm*) and “disagreeable” (*unangenehm*): they “are ever used with relation to the person that useth them: There being nothing simply and absolutely so.”

The good and the agreeable are essentially different, then, in that judgments of the former alone have the universal communicability characteristic of cognition. In view of this difference, further comment on the principle stated above is needed, both to confirm its implicit identification of the agreeable as the material condition of the simply good and

to indicate how it harmonizes with Kant's understanding of the way in which the agreeable and the good are related.⁵

It may be noted, first of all, that neither this principle nor the primitive general act of practical judgment in which a particular person makes the agreeable, the object of sensible desire, into an object of its will through bringing it under the concept of the good, are derived from experience. Though it is only through experience that we learn which things are agreeable, it is not an empirical discovery that agreeable objects, whatever they in all their diversity may turn out to be, are simply good to the extent that they can be practically cognized, while disagreeable things are excluded across the board. Indeed, it is in this a priori principle that we can find the element of pure philosophical truth in the claim of hedonists such as Epicurus, Hobbes, and Hume that the good lies in what pleases. As will be explained below, the agreeable is suited by virtue of its form—its agreeableness—to be deemed simply good, where good at all.

Kant states that the subsumption of the agreeable under the concept of the good takes place “through the concept of an end.”⁶ No doubt part of what this statement is meant to convey is the thought that in the judgment through which the agreeable is brought under the concept of the good the agreeable is regarded as the *effect* of that act of practical cognition (a thought that is of course an implication of Kant's view that the concept of the good represents the object of practical cognition). For Kant standardly characterizes ends as the objects of representations in which they are regarded as effects of those very representations: “an end is the object of a concept, so far as the latter is regarded as the cause of the former (the real ground of its possibility)” (*KU* 220; cf. *MS* 384). Thus, through an act of subsumption under the concept of an end, what is initially conceived, under the heading of the agreeable, as the external *cause* or occasioning condition of some pleasing representation in one-

⁵A more extended treatment of Kant's view of the relation is provided in “Kant on the Agreeable and the Good,” in Sergio Tenenbaum, ed., *Moral Psychology* (Amsterdam: Rodopi, 2007). In keeping with our decision to focus on positive practical judgments (§II.6), we shall concentrate on sensible desire and its object, the agreeable, leaving to the side aversion and the disagreeable.

⁶“The agreeable, which as such represents the object solely in relation to the senses, must first be brought under principles of reason through the concept of an end in order to be, as object of the will, called good” (*KU* 208).

self (some sensation or experience) comes to be conceived *practically*, under the heading of the good, as an *effect* of oneself, an effect one can produce through one's action.⁷ This is not to say that the initial conception of the object as an external cause is simply *replaced* by the practical conception of it as an effect of oneself. Rather, the latter conception includes the former and hence is logically dependent on it: the former is, so to speak, *aufgehoben*, in that the affection by that cause on which one's pleasing sensation or experience is in the first instance conceived as dependent comes itself to be practically regarded as something one can secure through one's action. To draw on one of Kant's examples: if I am taken by the agreeable verdure of a meadow I chance upon while walking through the country, I wish to linger in the enjoyment of it and may choose to do so if I have the time, or I may decide to visit it again or perhaps even to cultivate a garden at home as a sort of surrogate. A transition is thus made, with respect to my enjoyment of the object in question, from an awareness of bare dependency on something outside myself to a (practical) conception of self-sufficiency.

But while this interpretation of Kant's statement is correct as far as it goes, it does not seem to capture all of his meaning. We could have understood him along similar lines even had he said instead that the agreeable must be brought under the concept of the good "through the concept of the *means* to an end," for to practically cognize something as the means to an end is equally to regard it as the effect of that act of practical cognition. Kant's characterization of an end as the object of a concept in which it is represented as the latter's own effect permits a very broad interpretation of the term, according to which any effect of practical knowledge, even a mere means, can be called an end. But there is no reason to suppose that Kant thinks the agreeable is to be brought under the concept of the good through the concept of means. So it would seem that in specifying that the agreeable is to be brought under the concept of the good through the concept of an end he means to convey more than the general point that in being brought under the concept of the good the agreeable comes to be regarded as the effect of practical cognition. Further consideration of the concept of an end can therefore be expected to help

⁷Thus, Kant says, "the good or bad is properly referred to actions, not to the state of sensation of the person" (*KpV* 60). It is the agreeable (or disagreeable) that is properly related to a person's state of sensation: "The agreeable . . . as such represents the object solely in relation to the senses" (*KU* 208).

clarify the relation between the agreeable and the good. We should first consider how this concept is related to that of the good, and then ask how and under what conditions the agreeable can be subsumed under it.

One of the observations Kant makes in contrasting the good and the agreeable is that while it makes no sense to speak of something as *mediately* agreeable, we recognize a difference between what is mediately and what is immediately good, as when we distinguish between what is useful and what is good in itself, or between what is good as a means and what is good as an end (*KU* 208; cf. *G* 414). The concepts of end and means are accordingly complementary concepts that divide the concept of the good, yielding two species of goodness, or two ways in which a thing might be good. When thus situated under the concept of the good, the former signifies what is represented in practical cognition as being for its own sake and the latter what is represented in such cognition as being for the sake of something other than itself. But the relation of being-for-the-sake-of is just the relation of efficacy practically represented. So for something to be deemed good as a means, or for the sake of something else, is for it to be represented in a practical judgment as furthering something else or somehow contributing to its furtherance.⁸ And for something to be deemed good as an end, or for its own sake, is for it to be

⁸The concept of means may be applied either to action itself (*G* 414) or to what is used in action (what is practically represented as a material condition of the possibility of the action) (*G* 427). If riding a bicycle is a means of traveling to work, then in another sense the same is true of the bicycle itself. In the former case we might speak of the means as formal, in the latter as material. Where the means are material, the efficacy too can be described as material rather than formal, or in other words as the efficacy to be found among the objects represented in practical knowledge as to be used rather than the efficacy proper to practical knowledge itself and by which it makes its object actual. (Practical knowledge is not itself, just as knowledge, *for the sake of* its object, though its practicality, or the action whereby it produces the object, may be said to be.) In the first instance, material efficacy is just the enabling of an action, but often, especially where material means are themselves products of practical knowledge, contrived to enable certain types of action, they can be conceived as efficient causes (“Guns kill”), even though the effects are produced only through their use (“Guns don’t kill, people kill”). Though formal and material efficacy differ essentially, they are closely related. For, on the one hand, material efficacy depends on and is indeed constituted by formal efficacy: it is only because of the formal efficacy of practical knowledge—i.e., only because of the *action* issuing from such knowledge—that the material relations of efficacy represented in that cognition have any reality. We say our coats keep us warm, but they can do this only because we can maintain and wear them—only, that is, because we can use them to keep ourselves warm. And on the other hand, such action would obviously not be possible were there not antecedently given (available for use) an object suitably constituted to receive the material efficacy with which the action endows it. There is a parallel interdependency in the case of ends.

represented in a practical judgment as furthering itself. It therefore belongs to the concept of an end that an end is in necessary agreement with itself, that all the elements that may belong to it are in systematic harmony with one another, so that, all these elements thus furthering all and therein also being furthered by all (or, what comes to the same, being related as both means and end to the whole of which they are parts), there is nothing internal to the end that could bring it to an end or even impede it in any way at all.⁹ So to the extent that external conditions allow, an end always *sustains itself*.

Now as we have noted, the agreeable is an object the representation of whose existence is found pleasing. This representation can be considered in two relations. As a representation of *existence* it stands in a certain relation to its object; as *pleasing* it bears a certain relation to the subject. Though these relations may at first glance seem merely to point in opposite directions, closer inspection reveals them to be elements of a self-sustaining nexus between the subject and the object.

With regard to the first, it will suffice to observe that, in the sense of the phrase that interests us here, a representation of the existence of an object involves an implicit consciousness of the dependence of its own actuality on that of its object through the effect the latter produces, directly or indirectly, on the subject's senses. Representation of this sort may lie in sensation, as it does in the enjoyment of a meadow's verdure, or it may consist in experience (empirical knowledge), as in the shepherd Gyges' gratifying cognizance of his getting the better of others. But in either case

⁹ Cf. A832–833/B860–861, *KU* §64. The elements in question here are to be understood as related to one another as parts, or members, not as form and matter. Thus in the case of the highest good (the ultimate end for persons), which according to Kant consists in universal happiness collectively consequent upon universal virtue, the elements that in this end stand to one another in the relation of mutual furtherance just described are its members, namely persons. But insofar as the highest good is viewed as composed of virtue and the resulting achievement of happiness, related as form and matter, the furtherance among the components is not reciprocal: virtue produces happiness, not happiness virtue. This asymmetry is obviously due to the fact that virtue is the *form* of the highest good, the form that constitutes it as an end: it is because virtue is shared by all the members that the latter, through their actions, mutually further one another's existence and happiness. (Cf. A812/B839: "Morality in itself constitutes a system, but happiness does not, except insofar as it is distributed in exact proportion to morality.") And because virtue is the form of this end, it is not a *means* of achieving happiness, despite the fact that it produces it. Being itself an end, virtue furthers itself; and lying in practical (efficacious) knowledge, it also produces its object, the highest good, of which it is itself the form.

it involves a consciousness of the dependence of its own actuality on the subject's being somehow affected by the object whose existence it thereby represents. Such dependence is, of course, the characteristic empirical element in theoretical knowledge, its reliance on sensation.

As for the other relation, this representation, as pleasing, has a certain effect on the lifepower of the subject to whom it belongs. The pleasure that accompanies it is of a specific type, which Kant calls gratification (*Vergnügen*) or enjoyment (*Genuß*). This type is distinguished from other forms of pleasure, such as the one figuring in judgments of approval, discussed earlier, in that it arises directly through a representation of an object's existence, independently of any act of practical judgment. At the same time, of course, it also shares with the other forms a certain general character, common to them all as forms of pleasure. Kant articulates this common character in his definition of pleasure in general as the sensible awareness of the efficacy of a representation to sustain the state of mind to which it belongs (*KU* 220). Such efficacy is possible, obviously, only insofar as the representation in question is itself capable, in the state of mind to which it belongs, of sustaining *itself*. In the case of the enjoyment of the agreeable, however, the self-sustaining representation is one that, as we just noted, depends for its own actuality on the existence, or actuality, of the agreeable object. Thus the efficacy whereby this representation of the existence of the agreeable sustains itself and its own state of mind must equally be a determination of the subject's capacity to produce effects outside itself by bringing into existence objects it represents, that is, a determination of the subject's faculty of desire, specifically a determination of it through which the subject acquires a desire that has that same agreeable thing as its object. And accordingly the pleasure, as the feeling, or sensible awareness, of that self-sustaining efficacy, must equally be a feeling of that determination of the faculty of desire. Desire that arises in this way, through the feeling of gratification, is sensible desire, and its object—what it works to bring into existence—is the very thing from the gratifying representation of whose existence it itself arises. The effect is the same as the cause.

The concept of the agreeable, then, includes not only the concept of a causality, manifest in sensation, through which the agreeable object produces a pleasing representation in the subject, but also, in the understanding it implicitly contains of the pleasure that accompanies

this representation, the thought of a felt determination of the subject's causal power through which the agreeable object itself can in turn be brought into existence. Putting these two points together, we can say that the concept of the agreeable includes the idea of a sensibly manifest self-sustaining relation between subject and object, in which the object produces a representation through which the subject is in turn moved to bring into existence that very object in that same productive relation to itself. Thus, *pleasure in the agreeable* serves as a sensible criterion for the application of the concept of an end, and the capacity to feel such pleasure accordingly constitutes the receptivity of the capacity for practical knowledge. Though the concept of an end, as the practical idea of systematic agreement, is originally an a priori concept of practical reason derived from the form of practical knowledge, it can be used empirically insofar as it can be determined, or rationally specified, through empirically given representations, in accordance with this criterion. Pleasure in the agreeable is therefore also the sensible criterion for the application of the concept of the simply good, since the latter is just the concept of a practically cognizable end, but in its primitive, problematic mode, in which the object is considered independently of any determination of how or whether it is practicable in the actual conditions of action. Accordingly, provided that the practical judgment in which the concept of the simply good is applied does not conflict with any presuppositions it may involve (formal conditions of practical knowledge), the application of this concept to the agreeable is valid, and hence that agreeable object, as represented in that judgment, is indeed simply good.

It is important, however, not to lose sight of the point noted at the outset, that in Kant's standard characterization of it, an end is an object of a concept in which it is regarded, or represented, as the effect of that very concept. It is not simply qua agreeable that an object falls under the concept of an end (or the simply good). What is brought under this concept—or rather, to speak more precisely, what is reached through its determination—is a *practical* representation of the agreeable object. This representation is indeterminate (at least initially) in that it does not include any specification of how or whether the realization of the agreeable object is practicable in the actual conditions of action. But as will be explained below, it must nevertheless be in accordance with a certain condition, based in the representation of a common end to which all of a

person's diverse ends belong—the condition, namely, that one's pursuit of the agreeable object be in systematic agreement with one's pursuit of every other object one undertakes to pursue. Obviously the sensible criterion for the application of the concept of an end does not by itself guarantee the satisfaction of this condition. For it is not in themselves that the agreeable objects belonging to a person's end are related as members of a system,¹⁰ but only indirectly, through the systematic relation that the practically represented pursuit of each such object bears to the practically represented pursuit of every other element, a relation of nonaccidental mutual furtherance that is not possible except insofar as the pursuit of each is carried out in accordance with the single practical representation of a common end to which they belong.¹¹ So while it is true that pleasure found in an agreeable object reveals, independently of any practical representation of the latter, that such an object meets *one* of the requisite conditions that must be satisfied by anything that can enter into an empirically determined end as one of its elements, it is also true that the act of practical representation by which this object is included as such an element and thereby brought under the concept of an end—an end for the very person who brings it under this concept in that act—must be in accordance with the further condition that this object be pursued as a member in a system of elements belonging to a common end. This end will be considered more closely below (§III.4).

¹⁰ Where the enjoyments of these agreeable objects are based in instinct we can discover through experience that they have a natural limit whereby they stand to one another in relations of mutual furtherance as elements of a natural system of human animal life (cf. *KU* 430). But not all enjoyments have this natural systematic connection, as is revealed by addictions, the appetite for luxuries, and especially by the passions (*Leidenschaften*), whose objects are always intrinsically bad and so cannot be included as elements in any system at all. Obviously an appreciation of this natural system and the capacity to distinguish between what does and what does not belong to it are among the chief requisites of the prudent exercise of the power of practical judgment (cf. *G* 395–396).

¹¹ As "pursuit" is being used here, the pursuit of an agreeable object lies primarily in the enjoyment of it, but it also includes, secondarily, the actions undertaken to make such enjoyment possible. Thus the furtherance of the pursuit of an object can be of two sorts: the furtherance of the enjoyment itself, and the furtherance of the actions. But if something furthers the actions while at the same time preventing the enjoyment, it does not further the pursuit, except in a derivative sense. A man who feeds hay to his mule to give it renewed strength to pursue the carrot he dangles before it does not, strictly speaking, further the mule's pursuit of the carrot.

3. Types of Practical Judgment

Although judgments concerning what is simply good are like other practical judgments in being practical and in involving the idea of sufficient reason, they are distinctive in that they do not determine what it would be good on the whole to do and so do not depend on empirical theoretical judgments of practicability. We may therefore call judgments of this type *purely practical* judgments to indicate their lack of dependence upon such theoretical judgments and thereby to mark them off from judgments concerning what is good on the whole.¹² Alternatively, as was noted above, we may characterize such judgment as constituting *wish*, in contrast to choice. Although these purely practical judgments on which choice depends are not intentions, they must still, as practical judgments, together belong to a self-conception that is practical *in potentia* and that, to the extent that it satisfies the formal conditions of practical knowledge, amounts to an *ideal* practical self-conception, representing what Kant calls “the highest good in a person” (*KpV* 110). Thus both wish and choice constitute practical self-conceptions, one problematically, the other assertorically, practical; and both the idea of what is good simply, the object of wish, and the idea of what is good on the whole, the object of choice, are connected with action. The difference is that only the self-conception constituted by choice is specified subject to the requirements provided by empirical judgments of practicability, and only the idea of what is good on the whole implies that those requirements are satisfied.

Simple goodness should not be confused with the unconditioned goodness that Kant ascribes to the good will and to its expression in morally worthy action.¹³ If something is unconditionally good, it is a fortiori

¹²The fact that purely practical judgments do not depend on theoretical judgments of practicability does not imply that they must be “pure” judgments in Kant’s technical sense (i.e., free of all empirical ingredients); as we have just noted, a judgment in which a specific agreeable object is deemed simply good depends on both sensation and the feeling of gratification. Insofar as a purely practical judgment is valid, however, its object will be “morally possible” and hence an “object of pure practical reason” (cf. *KpV* 57–58). The “unconditioned totality” of the latter is said to be the highest good (*KpV* 108).

¹³The difference in meaning that the terms “simple” and “unconditioned” are meant to convey here corresponds to the distinction Kant points out between two senses of “absolute” (A324–326/B380–382). It is similar to the distinction Christine Korsgaard has drawn between what is good as an end (valued for its own sake) and what is intrinsically good (“Two Distinctions in Goodness,” in *Creating the Kingdom of Ends* [Cambridge: Cambridge University Press, 1996]).

both good on the whole and simply good. But simple goodness entails neither unconditioned goodness nor even goodness on the whole. What is simply good, or worthy of wish, is not good on the whole, or worthy of choice, unless it can be included among the jointly practicable goods. For something to be simply good is for it to be eligible, so far as it is considered merely in itself, for inclusion among the things good on the whole. But possession of this intrinsic eligibility is not by itself enough to warrant inclusion among the things deemed jointly practicable and good on the whole; determining whether an object should be included—whether it is choiceworthy—requires comparison with one’s powers of agency and therefore also, given that such powers are limited, with other goods.¹⁴ Obviously an action that is good on the whole need not be simply good; for the things deemed simply good as well as good on the whole are ends, but actions may be chosen as mere means. In sum, in the judgment of practical reason, what is good simply is the object of wish, what one *would* do or pursue; what is good on the whole is the object of choice, what one *will* do or pursue; and (as will emerge below) what is unconditionally good is the object of pure will, what one absolutely *must* do or pursue. The agreeable, or the pleasant, falls under the first, the prudent, or the useful, under the second, and the morally worthy, or the obligatory, under the third. We can thus locate under these three categories of goodness traditional moral philosophy’s

¹⁴ Because judgments concerning what it would be good on the whole to do are made in the light of considerations of practicability, they always depend upon some at least implicit comparison of the things deemed good in purely practical judgments. There would be no judgments concerning what is good on the whole at all if there were no judgments of practicability (i.e., no acquaintance with any limits on one’s powers of agency), and there would be no judgments of practicability without some experience of a lack of some of the things one deems good, and an ensuing recognition that which of the things deemed good will actually be achieved, and to what extent, depends on how one acts. Purely practical judgments, in contrast, are presupposed in all comparison among things deemed good and hence in all judgments estimating one such thing to be better or worse than another, and so do not themselves involve or depend on any such comparison. It is only once we take into account the limits on our powers of agency that we need to compare goods and to ask which among them are better. Let us fancifully suppose, as Hume at one point invites his readers to do, “that nature has bestowed on the human race such profuse *abundance* of all *external* conveniences, that, without any uncertainty in the event, without any care or industry on our part, every individual finds himself fully provided with whatever his most voracious appetites can want, or luxurious imagination wish or desire”: then not only would there be no virtue of justice, as Hume points out (*An Enquiry Concerning the Principles of Morals* § III, Pt. I), but in addition there would be no distinction between choice and wish, nor any practical thought of one object as better than another.

classification of the three objects of will: the pleasant, the advantageous, and the noble.¹⁵

We have now surveyed several types of practical judgment. In addition to the special case of choice, we have distinguished other empirical practical judgments that determine what it would be good on the whole to do (for instance, my judgment that it would be good for you to develop a certain talent) and also wishes for what is simply good. As was stated earlier, however, we shall direct our attention in what follows chiefly to the first-person case, which is primary. Maxims, the principles of action that Kant describes the moral law as governing, have their source in the free power of choice (*MS 226*), so it is with the regulation of this power that Kant himself is principally concerned in his account of the categorical imperative. But because the free power of choice depends on the capacity to wish and even contains the latter within itself, in that choice is merely the expression of wish under the conditions provided by the subject's judgments of practicability, it will be necessary to give further consideration to this more basic form of practical judgment as well. Indeed, strictly speaking it is wish, the root from which choice springs, that the moral law regulates. Choice is the body and the letter of the act of the free power of choice, wish its soul and spirit; only by determining the latter can the moral law extend its inward influence to the former. And if there is in turn a fundamental act of wish, it will be in this that the moral law first applies.

4. *The Wish for Happiness*

The account of the presuppositions of practical judgment to be presented below will not rely on any assumptions about what particular human persons judge to be good. It is possible, however, without making any such assumption, to see that of the various judgments making up a particular person's exercise of the free power of choice, there is one that we can identify and characterize in advance, even if only in general terms. For one judgment is implicated in the very concept of a person, or a subject with the capacity for practical knowledge. It is implicated as the act in which an individual person first constitutes itself as such, being as much a judgment that makes a person as a judgment a person

¹⁵ See *Nicomachean Ethics* 1104b30–32, *Summa theologiae*, Ia, q. 5, a. 6.

makes.¹⁶ Though fundamental to a person's other practical judgments, this judgment is distinct from and posterior to the formal conditions of practical knowledge to be discussed below. For like the other judgments, it depends on the receptivity of the subject's capacity for practical knowledge—it depends, that is, on the capacity to feel representations of existence determining the faculty of desire—and is therefore an individuating act, whereas the formal conditions are the same in every person and presupposed in all practical judgments, even the one in which an individual person first constitutes itself as such. This fundamental judgment, to which we now turn our attention, lies in the wish for happiness.

Picking up the thread of our earlier discussion (§III.1–2), we may return to our observation that the judgments in which one brings particular objects one finds agreeable under the concept of the simply good all depend on a primitive act of practical judgment, in which one subsumes what one finds agreeable *in general* under that same concept, even if only in such a way and to such an extent as is compatible with its being an object of practical knowledge. The primitive judgment differs from the others in that it alone involves a certain necessity. That this or that object is something one finds agreeable is discovered in experience, but that the agreeableness of an object enables it to be known to be simply good so far as it can be known to be good at all is something every person must already implicitly understand, pleasure in the agreeable being the very material of the simply good. The primitive act of practical judgment just mentioned thus engages an a priori recognition of the essential relation agreeableness has to goodness, a recognition that remains untouched, however contingent one may suppose it to be that it is these objects, not those others, that one finds agreeable. To be a person at all, then, is to have a conception of the agreeable in general, a general understanding of the character shared by any and every object one finds agreeable and therein comes to desire, namely that the representation of such objects' existence is pleasing to one.

Since pleasure is a consciousness of the efficacy whereby a representation sustains itself and the state of mind to which it belongs, it excludes, on account of its very form, any thought of term or limitation. The original concept of the agreeable accordingly includes the notion of

¹⁶ Judgments, in the sense used here, are activities, not episodes or events; cf. §IV.2.

ongoing, or enduring, enjoyment. Subsequently, of course, we may attach limitations and restrictions to this thought. For in the first place, we can see from the start that the pleasure is contingent, both in that its relatedness to the representation it accompanies is not knowable a priori, and in that the latter, as a representation of existence, depends on an object affecting the subject. In addition, experience reveals the enjoyments of agreeable objects to be of limited duration and variable intensity—"the first bite always tastes best"—as well as liable to interruption, interwoven as they are, often inextricably, with the many strands of pain and discomfort manifesting the numerous needs of life. Cognizance of such limitations and conditions is obviously integral to the practical deliberation and planning on which we rely in choosing what to do. No one who steps out for a pleasant stroll has it in view to continue walking forever, nor do mature human adults base their decisions on the supposition that infirmity and death will never befall them. Together these limitations make up the general practical problem of life, setting the stage for deliberation. But all such limitations depend on the thought they limit, that of the enduring character of gratification itself. Their introduction does not efface the underlying notion, though it provides the occasion for the latter's reconception as that of uninterrupted enjoyment throughout one's existence, or of a fully enjoyable life. This is just the thought of happiness (*Glückseligkeit*) in the sense in which Kant standardly employs the term (*KpV* 22); happiness in this sense is "an ideal of the imagination" (*G* 418–419).

Happiness so represented is clearly an object of sensible desire. For as we have noted, the representation the self-sustaining efficacy of which one is conscious in the feeling of gratification is a representation of an object's existence and so one whose self-sustaining efficacy must also be a determining of one's faculty of desire to desire that object—in this case happiness, or (what comes to the same) whatever occasions the enjoyable experience in which it consists. This sensible desire is not itself a practical judgment, of course. Happiness as an ideal of the imagination is not yet happiness as an object of wish. But as was argued above, the judgments in which one brings particular objects one finds agreeable under the concept of the simply good all depend on a primitive act of practical judgment that subsumes what one finds agreeable *in general* under that same concept even if only to the extent that it can be an object of practical knowledge. And we have just seen that the agreeable

in general, as originally conceived in the feeling of gratification, is happiness as an ideal of the imagination. Hence every particular person—every subject with the capacity for practical knowledge—deems its own happiness simply good, at least in a qualified way. And since to take something to be simply good is to make it one's end (in the "problematic" manner described earlier), the foregoing considerations confirm, and furnish a basis for understanding, Kant's claim that happiness is an end for every human being, an end that belongs to the very being, or essence (*Wesen*), of a finite rational being, so that we may presuppose with certainty and indeed a priori that any such being will have happiness as an end (*G* 415–416).¹⁷

Doubts have sometimes been raised regarding this assertion, owing to the apparent difficulty of reconciling the idea of an embracing ultimate end with the disorder and fragmentation to which human practical life is often subject. But we can secure our conclusion against such concerns if we approach it from another angle by returning to the concept of a human person, but now considering a person, not from the side of receptivity and the material conditions of wish, but with regard to practical spontaneity and agency.

It belongs to the concept of a person that as the various judgments by which a person determines what it would be good on the whole to do depend on judgments about what is simply good, so the latter all depend in turn on a single idea of a common end, an end in which the objects represented in those judgments of simple goodness are all contained. For according to this concept, a person is not resolvable into a multitude of distinct desires or motives present together in a human being like travelers thrown together on a captainless ship, each with a different destination, each contending with the others for control of the helm.¹⁸ As a subject with the capacity for practical knowledge, a person is rather a

¹⁷They also confirm Kant's statements that all humans have an inclination for happiness (*G* 399), that they wish for it (*G* 418, *MS* 387), and that everyone presumes it to be good (*KU* 208).

¹⁸This is not just to distinguish a person from what is sometimes called a "wanton" (Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *The Importance of What We Care About* [Cambridge: Cambridge University Press, 1988]). According to the conception here articulated, persons are no more distinguished by the possession of a certain type of second-order desire than they are by having any other special sort of desire (e.g., "rational desire," desire in conformity with reason); rather, they are constituted by the capacity for practical knowledge. Only the latter can ground the self-conscious unity of agency implied by the concept of a person.

single subject of various practical judgments that pursues so far as is practicable the objects and actions deemed good in those judgments. But because there is *one person*—a single principle of action, as opposed to a mere collection of discrete principles of action present together in a human being—only insofar as there is *one action* to which the diverse actions assignable to the person all belong, and because diverse actions can all belong to a single action only insofar as they all belong to the pursuit of a single end, different actions can belong to a person only insofar as they serve a common end and hence only insofar as the practical judgments on which the actions are based themselves depend on the conception of that end.

What can make the idea that a person's various actions all have a common end seem questionable is the failure to separate it from certain gratuitous additional suppositions. There is no need to suppose that this end must be articulated and specified in such a way as to provide a "goal in life" or a "plan for living"; nor must we suppose that persons are always effectively guided by it in their conduct.¹⁹ All that must be involved in the conception of this end is the at least implicit practical understanding that the pursuings of the various objects deemed good in the person's practical judgments, being the actions of the same single individual subject who makes those judgments, belong to a single practical life and so are possible only insofar as they can be somehow included in that life as modes of the basic activity of living in which it consists. Having this conception does not imply that one has already determined what those pursuits, those activities, are, or how, specifically, they can be integrated and jointly realized; on the contrary, these questions cannot even be raised unless the conception is already in place.

Though this common end is clearly the same as the one to which agreeable objects were earlier said to be related in being deemed simply good (§III.2), it is not to be identified with the end of personal happiness.

¹⁹ Cf. John Rawls's discussion of Royce's idea that "a person may be regarded as a human life lived according to a plan. For Royce an individual says who he is by describing his purposes and causes, what he intends to do in his life," *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), 408. This characterization does, however, highlight the important point that our notion of a person includes an idea of self-determination, an idea implicit in the older sense of "person" as *dramatis persona*, a character an actor *adopts*, and registered in Locke's often-cited observation that "person" is "a Forensick Term" (*An Essay Concerning Human Understanding* II.xxvii.26). Persons' ends must be ends that they themselves set.

If, as we have been supposing, there are formal conditions of practical knowledge distinct from the material conditions identified above, this common end, so far as it is an object of practical knowledge, will neither be confined to one's own happiness nor include it without qualification. But since all judgments of the simply good depend on material conditions provided by the awareness of objects as agreeable, and since the first such condition—the consciousness of agreeableness itself—is an implicit representation of happiness as an ideal of the imagination, the common end implied by the concept of a person will necessarily *include* this ideal of the imagination, even if only to the extent that the latter can be an object of practical knowledge.

What is of interest to us here, however, is the primitive exercise of wish, in which one deems one's happiness to be simply good, whether without qualification or only to the extent that it can be an object of practical knowledge. This act is necessarily unilateral. On account of its reliance on material conditions, belonging to receptivity, it is the act in which a particular person first constitutes itself as such, distinct from every other. It therefore can never arise through actual agreement among diverse persons, but instead is already presupposed in every such practical agreement. Its object is one's own in a correspondingly exclusive sense: in contrast to the capacity for practical knowledge, one's end of happiness belongs to one not just as a person, but as the particular person one is. Since this indeterminate practical representation of happiness is one the possession of which is implied by the very concept of a person, and since it furnishes the genus in any specification of what one's wished-for happiness consists in, we may characterize it as the *formal, generic conception of happiness* and so distinguish it from whatever substantive conception may be worked out through such specification.

5. *Form and Matter of Happiness*

To the extent that the generic conception has been empirically specified through the exercise of wish, it is possible to deliberate with a view to determining what it would be good on the whole to do. In this determination the specific ends problematically related to oneself in wish are, to the extent that one's empirical judgments of practicability allow, assertorically related to oneself in choice: what one wishes to do, or would do,

thereby comes to be what one means to do, or will do. This process of practical deliberation leading from wish to choice obviously depends on the recognition that it is possible for objects of wish to be impracticable. There could be no deliberation were the subject not conscious of limitations in its productive capacity, on account of which potential practicality is not immediately actual. Yet as we have noted, deliberation also evidences the practicality inherent in mere wish. Indeed, the inner act of deliberation is the characteristic form of the efficacy proper to wish itself, an efficacy that can be described as practicality's own self-actualization, and even (given practicality's self-awareness) as its self-conscious self-actualization. The seeking, in deliberation, to find a way in which the simply good can also be practicable and good on the whole is wish's striving to be choice.

In fact, wish goes further still. Deliberation does not stop at mere actual practicality, but seeks to make it strong and secure, by bringing the object of choice within a sphere of protection and control. No banquet, however sumptuous, will satisfy if we must enjoy it in the condition of Damocles. Even when sufficient means are found and wish converts itself into choice, if residual uncertainty clouds one's prospect for success in realizing the object, or if one is distracted from one's enjoyment of it by the apprehension that its dependence on one's choice is contingent and unstable, further deliberation naturally ensues to perfect the actuality of choice's practicality by raising it toward necessity. This perfection lies primarily in prudence (*Klugheit*), or skill and practical strength in deliberation, and secondarily in the resulting powers and means—acquired theoretical knowledge and technical skills, complemented by the necessary external goods—that facilitate both deliberation itself and the carrying out of the choices in which it results. Practical judgment thus naturally proceeds, in respect of its practicality, from possibility through actuality toward necessity (or, in traditional terminology, from power to act to habit).²⁰

²⁰These modes of the *practicality* of a subject's capacity for practical judgment, which reflect its dependence on the subject's capacity for *theoretical knowledge*, are to be distinguished from the modes of its *freedom*, which reflect its dependence on the subject's capacity for *practical knowledge*. The latter are grounded in the freedom of the power of choice, or in the capacity to exercise the power of practical judgment in accordance with its own inner principle, the form of practical knowledge—a capacity that proceeds from the wish to judge in accordance with this form (the "good will," a formal wish lying in the

To secure the practicability of the agreeable objects one deems simply good—whether by limiting wish to ensure that one’s end is possible, as was recommended by many ancient thinkers, such as the Epicureans and the Stoics, or through the stratagem, more in favor among the proponents of the Enlightenment in Kant’s own day, of developing technical knowledge, talents, and other natural abilities so as to increase one’s capacity effectively to pursue one’s ends—is to acquire and to maintain what we may call *practical self-sufficiency*. It is thus implied by the idea of a person who is in a position to wish at all—that is, a person in whom wish and choice are distinct, mediated by deliberation—that such a person wishes for this self-sufficiency. This wish belongs to the very form of wish. It is coeval with the recognition that one’s powers of agency are limited, that the practicable can fall short of the simply good and so is distinct from it; and its object is the restoration of their unity, the unity of potential and actual practicality—a unity in which practical cognition realizes the necessary connection between itself and its object that it always already thinks (practically) in understanding itself to be causally efficacious. Because this wish for self-sufficiency is necessary in the way just described, its object must be included in the totality of the object of wish, though for the same reason self-sufficiency belongs to the form rather than the content of that object. So far as it is represented as an ideal of the imagination, happiness has no place at all for self-sufficiency; but as an object of wish, it inherits such self-sufficiency as an essential formal component through being subsumed under the concept of an end and thereby practically represented as one’s effect.²¹

form’s own practical self-consciousness) through actual judgment toward necessity in judging in accordance with it, a necessity expressing not prudence but virtue (*Tugend*). And both of these types of modality, which are *subjective* in the sense that they pertain to the free power of choice in an individual subject whose nature does not preclude the possibility of limitations and defects in the exercise of this power in the two respects indicated, are to be distinguished from the *objective* modalities, which reflect how an object is related to the capacity for practical knowledge itself. Thus, neither of the two sorts of necessity just noted are to be equated with the objective practical necessity distinctive of practical cognition of the unconditionally good described earlier (§III.3), which is a priori and so precedes acts of choice rather than growing out of them. Nor is either of the two sorts of potentiality to be equated with objective practical possibility, which lies in permissibility, or compatibility with the formal conditions of practical knowledge.

²¹The idea that self-sufficiency is a formal element of happiness as an object of wish is reflected in Kant’s characterization of happiness as a condition in which “everything goes according to wish and will” (*KpV* 124): in saying “goes according to” rather than merely

Happiness, then, in its generic conception, as the totality of what a person finds agreeable so far as this is made an end, or object of wish, includes as part of its form two distinguishable components: the agreeableness of the object, and the causal sufficiency of the subject for its production. These two components together constitute a self-sustaining double-sided relation between the subject and the object, a relation of mutual interdependency, one side corresponding to the subject's receptivity, the other to its spontaneous productive power. Kant is well known, of course, for the emphasis he places on the great uncertainty, abundantly displayed in experience, that attends the pursuit of happiness, an uncertainty that can rise, he holds, even to the point of impossibility if happiness is made the be-all and end-all of life (*G* 395–396, *KU* 430). But far from casting doubt on the formal features just noted, this considerable contingency rather reflects their essential presence, confirming that the generic conception contains in its very form the idea of a certain self-sustaining stability, which happiness must have in order to be conceived as an end at all. Nor should we suppose that the idea of self-sufficiency is best exemplified by a Robinson Crusoe or a rugged individualist, or through some exaggerated image of the self-made man. Self-sufficiency does involve a certain independence, the ability to stand on one's own two feet, as we say, and to manage one's own affairs, putting it in proximity to what nowadays is often called "personal autonomy." But persons who become attached to an inflated ideal of individualism or to some other excessive conception of self-sufficiency do so through the specific objects they opt to include in the content of their end rather than on account of anything belonging to its form. In that regard, they are like persons who suppose that the enjoyments that belong to a happy life find their paradigm in the pleasures of sensation. As part of the generic conception of happiness, the original conception of self-sufficiency is itself generic and so distinct from any specific conception an individual person may develop. Self-sufficiency can take a collective form to the

"is in agreement with," Kant suggests that in such a condition wish and will in some sense *determine* how things go. The presence of self-sufficiency in happiness is also reflected in his understanding of happiness (*Glückseligkeit*) as a condition analogous to that of blessedness (*Seligkeit*), "a complete well-being independent of all contingent causes in the world" (*KpV* 123n; cf. 25, 129); but though happiness is akin to blessedness on account of its self-sufficiency, this self-sufficiency is not absolute, not altogether independent of fortune (*Glück*), and therefore no more than *Glückseligkeit*.

extent that persons join their wills, entering into communities and other cooperative engagements, and it will have an essentially collective dimension where, as in the human case, persons are naturally sociable and born into families.

We are not yet in a position to locate the ultimate ground of this twofold character of the generic conception of happiness. But from the investigation of the form of practical knowledge to follow it will eventually become apparent that the two components express complementary aspects of the discursive-cognitive character of the act of practical knowledge. This act, as we shall see, is self-sustaining and to that extent independent, yet also self-enlarging (or self-determining) and to that extent such as can stand in relations of interdependency with other acts of practical cognition.

6. Generic and Specific Conceptions of Happiness

A noteworthy difference between the formal concept of happiness, on the one hand, and the substantive conceptions of happiness and of the practicable good that a particular person works out, on the other, is that whereas the former is implicated, as we have seen, in the concept of a person, the latter depend on experience. That the agreeable, so far as it can be practically known, is simply good is not a matter of empirical discovery. But obviously it is only by relying on certain empirical judgments that particular persons can identify the objects they find agreeable, ascertain the extent of their powers, and acquaint themselves with the circumstances in which they must act. None of this is simply a matter of reflecting on the general concept of a person.

Once this difference is noted, it is apparent that the goodness that specific practical judgments deem specific ends and the actions furthering them to have depends on the goodness that the end of happiness, generically conceived, is deemed to have, in that the former judgments simply specify the objects in which that happiness is to be found and the actions requisite for the successful pursuit of them. They do not represent any further basis for regarding as good the ends and actions they deem good beyond what is already provided by the recognition of the relation those specific ends and actions bear to the end of happiness generically conceived. For they all rely entirely on the same a priori recognized sensible criterion, pleasure in the agreeable, which is marked out by its

self-sustaining form as material for the simply good. These judgments are therefore all informed by the subject's original, primitive attitude toward this common character of agreeableness. So the judgment that a particular object one finds agreeable is simply good depends for its own validity on the validity of the general, primitive practical judgment as to the goodness of one's happiness, formally conceived, in which what one finds agreeable in general is (qualifiedly or unqualifiedly) deemed simply good—a judgment that depends in turn for its validity on its conformity to whatever formal conditions of practical knowledge there may be. And since goodness is just practical cognizability, corresponding to this dependence in point of validity is a dependence in respect of the goodness of the object: the goodness of the particular object depends on the goodness of the happiness. The converse, on the other hand, does not hold. One could not think to oneself: It wasn't until it became clear to me that happiness, for me, lies in music (or in being with my family, or in commerce, etc.) that I came to care about happiness at all or to regard it as anything good. This order of goodness is a reflection of the dependence Kant asserts in his famous declaration that the good will is the one good on which all other goods depend (*G* 393–394). **As practical reason, or the capacity for practical knowledge, the will proceeds from the universal to the particular and stands like a kind of Midas to its object: from the highest end to the lowest means, it makes everything it touches like itself—good if good, bad if bad.**

The dependence just described has an implication that is of potential significance for our purposes in this study. If it should turn out that the validity of a practical judgment concerning one's happiness generically conceived is contingent on its satisfying certain formal conditions of practical knowledge, then so is the goodness that any specific practical judgment ascribes to a particular agreeable object belonging to that happiness substantively conceived. Whether there are such conditions and if so what they may be are matters to be taken up below.

7. *A Caveat*

Before we turn to these questions, however, a comment is in order to ward off a possible misunderstanding of the point of the reflections to follow. The fact that Kant frequently speaks of the moral law as an a priori law valid for all rational beings (e.g., at *G* 389, 408, 425) often leads readers

to think that he supposes it possible, even necessary, to derive this law from the concept of a rational being. One might therefore be led to regard the endeavor on which we are about to embark, in which we shall undertake to identify presuppositions of practical judgment with a view to tracing this law to practical reason, as part of an attempt to provide such a derivation. But here the following points should be borne in mind.

The considerations to be spelled out below (§§IV–V) do not rule out the possibility, intimated earlier (§II.4), of a merely prudentially rational being—a being that lacks a free power of choice but is nevertheless capable of intention, of efficacious action specification, and hence of an efficacious self-conception, including a conception of the totality of the somehow practicable activities in which its happiness consists, and is in addition capable of rational deliberation in the specification of this conception and the means for its realization. Such a being, should it be possible, would act on the basis of practical thought, but not on the basis of practical judgment or practical knowledge: it would not employ in its practical thinking the concept of the good outlined above, or (what amounts to the same thing) the concept of sufficient or supporting reason. The practical conception of happiness would provide a basis for prudentially rational decisions, but the pursuit of this happiness would not itself be regarded as rationally supported and so would not be an object of free choice, or action specification determined by reason. Though this conception of happiness would be efficacious, it would not be efficacious *through* any practical judgment determining the pursuit of that happiness to be good; its efficacy would derive rather from sources external to practical thinking, sources from which it obtains its content. And since this content would be determined, not through any practical judgment empirically determining the concept of the simply good, but directly by the experience of agreeable objects, it would be this bare experience of the agreeable, the pleasing effect these objects have upon the mind, that would provide the source of efficacy. In such a being, reason would have only the subalternate or derivative function of further specifying given action specifications, never that of supporting them though providing an original, or immediate, specification, a specification through which a practical subject can be said to be self-constituting, or a person.

a merely
prudentially
rational being

But the fact that the considerations to follow do not rule out the possibility of such a being does not represent a departure from Kant's understanding of the relation the moral law bears to rational beings.

Kant does think that the *ground* of obligation can be found only in a priori concepts of pure reason (*G* 389), but he does not suppose that mere analysis of such concepts can ever produce an interest in morality in a practical subject in which that interest is not already at least incipiently present through the moral law's being given—through its announcing itself, as it were—in practical consciousness. He says explicitly that we lack the insight to rule out the possibility of a being such as the one just described—a prudentially rational being without the least inkling of even the possibility of such a thing as the unconditionally commanding moral law: “Were this law not given within us, we would not be able to reach it, as such, by any subtle reasoning [*durch keine Vernunft herausklügeln*]²², nor to talk the power of choice [*Willkür*] into accepting it” (*R* 26n). Precisely because this law is not just a thought or a theoretical judgment, but a command (in other words, because it is not just a *representation* of law, but a *representation of law that is itself that very law* and so capable of determining our free power of choice to choose to act in accordance with it), there can be no intimation of it in a particular subject other than in the self-conscious efficacy belonging to that subject's own representation of this law, an efficacy that constitutes the law's practical reality. Essentially the same point can be found in Kant's claim that the categorical imperative, though an a priori practical proposition, is not analytic, but synthetic (*G* 420), and in his well-known doctrine of the “fact” of reason (*KpV* 31). *Kant does appear to regard it as a matter of moral conviction or faith (Glaube) that all rational beings do take an interest in morality (see A829–830n/B857–858n), but he does not confuse such conviction with rational certainty based merely on an analysis of the concept of a rational being in general.*²³

²²Kant's use of *herausklügeln* here—given its association with *Klugheit*, or prudence—suggests that he means specifically to deny the possibility of founding this law on the basis of prudential considerations, however elaborately or subtly they may be developed.

²³Thus, in the argument leading up to the formula of humanity, Kant describes as a “postulate” his assertion that every rational being represents its existence as an end in itself (*G* 429n). Since he identifies humanity with rational nature, the points we have just noted also have a bearing on what he says about humanity: As a person's consciousness of the moral law develops, it tends increasingly to be accompanied by the at least implicit conviction that humanity includes the capacity for morality—the capacity that gives humanity its dignity (*G* 435). But in the bare concept of humanity so far as it is conceived simply as the rational nature belonging to a rational being in the minimal sense just indicated, the idea of such a capacity is not to be found (*R* 26n).

When Kant speaks, therefore, of the moral law as an a priori law valid for all rational beings, he is not claiming to have rational insight into the proposition that all rational beings are bound by the moral law in a sense that would imply that they necessarily have the capacity to be moved by it to act. Kant does seem to think that all rational beings are capable of recognizing from the very idea of a rational being whose rational cognitive capacity is *itself* a source not only of thought and cognition but also of *action* that the first principle of action for *such* a being is the moral law. More specifically (and in anticipation of the considerations to be elaborated in the sections below), we can take Kant's thought here to be that, since reason is the capacity to cognize things according a conception of law, a capacity based in the original a priori conception of law in general, and since the nature of rational beings in general thus lies in this capacity, which, as a cognitive capacity, is implicitly self-conscious, all rational beings are able to recognize, on reflection, that for a being whose nature lies in the capacity for such law-grounded knowledge to act from that capacity itself (and so to act from its rational nature) would be for it to act from its conception of law, which (according to the argument presented in the *Groundwork*) would be nothing other than to act from the moral law, as expressed in the categorical imperative's formula of universal law. All rational beings, in other words, are able to recognize that the law expressed through this formula is the unique representation through which a rational being can constitute itself as a *practically* rational being at all. In this sense, the moral law, so far as it is expressed in this formula, is valid for all rational beings. But to say this is not to rule out the possibility that there may be rational beings who lack the capacity to be *moved* by this thought of themselves acting from a conception of law—beings, in other words, in whom the capacity to cognize according to a conception of law is not also present in the form of a capacity for *practical* cognition, even though, as a capacity for *theoretical* cognition, it may play an indispensable role in the specification of such a being's practical *thought*.

Of course, on the other hand, the fact that this possibility is not ruled out provides no basis at all for supposing that such beings—beings to whose nature would belong rational animality but no animate rationality, beings whose actions would express rational desire but not desiderative reason—are really possible. It provides no basis, in other words, for supposing that they are anything more than mere figments of our analysis.

the nature of a rational being lies in the capacity of "reason" narrowly construed



FROM PRESUPPOSITIONS
OF JUDGMENT TO THE IDEA OF A
CATEGORICAL IMPERATIVE

IV



The Formal Presuppositions of Practical Judgment

1. Judgment and Cognition

In the preceding sections we have been developing the idea that willing, in the sense Kant has in view, consists in a certain sort of judgment. If this account is correct, then we may reasonably expect that an examination of the concept of judgment, giving due regard to ingredients to which Kant himself calls attention, will throw further light on his conception of the will. It would be a considerable undertaking, one beyond the scope of this study, to survey the many types of judgment Kant distinguishes and to look for a generic concept that covers them all, or a basic form to which they are traceable, though it does seem clear that all of them are at least in some way related to cognition in its human (discursive) form. Like other philosophers of his time, Kant classifies under the heading of judgment a wider array of mental operations than we commonly do today, including some that are in no sense acts of cognition. But since judgments of the type concerning us here are exercises of the will, or practical reason, understood as the capacity to know the good, and hence judgments in which a certain sort of knowledge—practical knowledge—consists, we may confine our attention to judgment so far as it counts as an act of the cognitive capacity. That is to say, we may focus on what Kant calls “logical” judgments, or judgments produced in the knowledge of objects by the understanding or reason, our discursive cognitive capacity, and we may leave to the side what he terms “aesthetic” judgments (such as judgments of taste), which, though they do involve representations that can figure in the knowledge of objects,

consider these representations in their relation to the judging subject through the operation of the capacity to feel pleasure or displeasure in such representations and hence do not themselves amount to cognition (*KU* 203–204; cf. 188–189). It should not be forgotten, of course, that judgments of this latter type can be *prerequisites* of cognition: we have already noted in effect that judgments of the agreeable, which make up one class of “aesthetic” judgments, are necessary material conditions of the empirical application of the concept of the good in practical cognition (§III.2). Yet as we also noted, such judgments do not themselves have the universal communicability characteristic of knowledge and so cannot lie in the exercise of the cognitive capacity, but belong instead to receptivity. Accordingly, in the discussion to follow “judgment” will normally signify “logical” judgment, the act of the discursive cognitive capacity.

To treat our topic in the proper order, we shall approach it from above, working our way downward until we reach the specific concept of practical judgment. Kant holds that “in the end there can be only one and the same reason, which must be differentiated only in the application” (*G* 391), so he must also hold that all rational cognition, both theoretical and practical, shares a common form. Once the general character of rational cognition is in view, we can consider what is distinctive about practical cognition.

2. *Thinking and Judging*

We may begin by noting two components Kant recognizes in the concept of judgment. The first is the idea of self-consciousness, which is already present in the more general concept of thinking. Self-consciousness, in the sense intended here, is not to be conceived as any form of self-observation, including the sort of inward observation—or introspection, as it is often called—that would presuppose a capacity of the mind to perceive its own operations, a capacity that, according to Locke, might be suitably described as “internal Sense.”¹ It is also to be distinguished from what might be called “aesthetic” self-awareness, that is, from awareness of one’s state of mind by the feelings of pleasure and displeasure attending one’s thoughts and other representations. The fundamental point

¹ *An Essay Concerning Human Understanding* II.i.4.

of contrast between self-consciousness and such other forms of self-awareness (which, as forms of *self*-awareness, all presuppose it) is that what self-consciousness is an awareness of is not anything that can exist prior to, or independently of, that consciousness. Nor on the other hand can it be an effect or product of this consciousness, something the latter brings into existence. Rather, it is identical with it, so that there can be nothing belonging to its constitution that lies beyond this consciousness. Self-consciousness, then, is not itself any form of knowledge of an object, whether theoretical or practical; it is nevertheless fundamental to all thinking and knowing and indeed fundamental in such a way that everything integral to the concepts of thought and cognition is comprehended in self-consciousness.

Reflection on this identity—the identity of self-consciousness with the consciousness of which it is a consciousness—reveals that self-consciousness is essentially one: not just something that *has* unity, but unity itself, unity of consciousness lying in consciousness of unity. For on account of this identity, self-consciousness excludes even the *possibility* that it is a mere multiplicity of independent items—a mere aggregate, say, or a mere flux (whatever the constituents might be—sensations, mental states, etc.). No bare multiplicity could be conscious of itself *as* multiplicity. Consciousness of multiplicity depends on consciousness of distinct items. And since self-consciousness cannot be anything separate from that of which it is the consciousness, the self-consciousness of the one item would have to be distinct from the self-consciousness of the other, since these items are themselves distinct. But a multiplicity of individually self-conscious items does not amount to a consciousness of a multiplicity. Self-consciousness, therefore, is inherently unitary. So far as it contains distinguishable components within it, they are originally related as components belonging to a single whole. We may call this unity the unity of thought.

It follows that the possibility of a self-conscious whole in which diverse components are contained—for example, a thinking of some thought, containing both the thinking of a subject and the thinking of a predicate—requires that the constituents not only be themselves self-conscious, but also contain within themselves the same self-consciousness. There must be a *single* self-same consciousness *in* each of the components, through which the latter are related to one another. The consciousness of the whole must accordingly precede the specific consciousness

of the components, as the consciousness of the form of relation in which the latter stand to one another in the whole; and this consciousness of the form of relation, or form of the whole, must be *in* each of the conscious thinkings that make up the components, as what enables the latter to be conscious of themselves *as* components of the whole and indeed as components of the *same* whole. In order for me even so much as to conceive of the sun as round, for instance, my understanding of the sun and my understanding of what it is to be round must be exercised in the same act of thinking. There must be a single self-conscious act—a determinate use of my understanding of the subject–predicate relation—in which, *in* thinking of the sun I understand myself to be thinking of what in that act I am conceiving as round, and conversely *in* the conceiving as round I understand that what I am thinking as round is what in that same act I am thinking of in thinking of the sun. It is only through this relation to the common whole then, that the components of a thought are related to one another, or combined in a single consciousness.

The necessary involvement of self-consciousness in thinking enables Kant to characterize thinking as spontaneous, as originating in a capacity “to bring forth representations oneself” (B75). For without self-consciousness there could be no thought or representation of oneself at all (nor, for that very reason, any self), nor therefore any conception of a capacity to bring forth representations oneself. But as Kant’s interpreters have often emphasized, it is important not to overlook his characterization of the way in which thinking and self-consciousness are related. In his famous pronouncement in the first *Critique* that “The *I* think must be able to accompany all my representations” (B131), Kant takes note of a certain conceptual point about the relation between thinking and self-consciousness. If we read this assertion in conjunction with his later remark that “the *I* is merely the consciousness of my thinking” (B413), we can express the point by saying that thinking is an at least *implicitly* self-conscious activity. Though thinking need not involve clear and explicit awareness of itself, the possibility of such awareness is implicated in the possibility of thinking itself.

Kant also says that the “capacity to think,” or “the understanding in general,” is the “capacity to judge” (A69/B94),² and that the *I think*

² Cf. A80–81/B106; P §22.

“contains the form of every judgment of the understanding in general” (A348/B406). As these remarks indicate, it belongs to the idea of judgment that judging is thinking. **Judging too, then, is at least implicitly aware of itself**, and that is to say that it is aware of itself as such: not only as thinking, but specifically as judging. Judging does not comprise two parts, one lying in thinking and the other in something other than thinking, something beyond the scope of self-consciousness; nor is it a type of thinking that is distinguished by the presence of some accidental feature that may belong to thinking when the latter takes place in certain conditions but not when it takes place in others (for instance, the feature of being influenced in a certain way through external stimulation by objects affecting the senses). Rather, judging is a self-constituted type of thinking, and as such it involves an implicit self-consciousness that includes a grasp of its own concept and extends to all that is integral to it as the act that it is. The different forms of judgment, such as theoretical and practical, must accordingly be likewise self-constituted.

judgment as “self-constituted”

The second component is specific to the concept of judgment. **Judgment differs from bare thinking (mere conceiving) in that the act in which it consists cannot be arbitrary.** Though, as Kant says, “I can *think* whatever I will, provided only that I do not contradict myself” (Bxxviii), the same cannot be said with regard to judgment. All thinking, Kant holds, lies in the combining of representations in accordance with the unity of consciousness (B145, *P* §22). In judgment, however, where thinking is engaged to cognize an object, representations are not combined “in a haphazard way, or arbitrarily [*beliebig*]”; and in the case of what Kant calls determining judgment—the specific type of judgment that will concern us below and in which the capacity to judge is employed in the rational cognition of an object—this nonarbitrary combination even involves “something of necessity” (A104–105), intimating the possibility of comprehending the ground of its nonarbitrariness.

“Bare”
thinking
vs.
Judgment

To say that judgment differs from bare thinking in that its combination of representations is nonarbitrary is to characterize it in negative terms. But we can also distinguish it in a positive way by noting that this combination is nonarbitrary just in that it sustains itself. Judging and bare thinking have in common that each is a self-conscious combining of representations, which, as such, has unity of thought and so contains nothing internal to it that prevents it from being sustained; but judgment is positively self-sustaining. Someone who is ignorant of the size

of the sun can *think* of the magnitude of its volume as having any of many possible values, but while each of these possible thoughts excludes all the others in the sense that it is incompatible with them, so that they can be actually thought or entertained in the understanding only severally, never together, none of them *holds itself* in the understanding, excluding the subsequent thought of the others. When on the other hand one *judges* the sun to be of a certain determinate magnitude, the act sustains itself, not allowing itself to be replaced by any other judgment regarding the sun's magnitude that is incompatible with it (though on account of its contingency it does allow, and must allow, the subsequent actual entertainment of *thoughts* that are incompatible with it). Similarly, if in deliberating about how to achieve a certain end one is indifferent between two individually sufficient but incompatible ways of doing so, the bare thought of pursuing the end in the one way does not exclude the subsequent thought of pursuing it in the other; indeed, vacillation in practical reflection is a characteristic mark of indifference. But if one reaches a decision and judges that one of the alternatives is good on the whole, subsequent judgment that the other is good on the whole is thereby excluded. (None of this is to deny that a judgment may eventually be forgotten or rejected as erroneous.)³

judgment as
self-sustaining

³The distinction between bare thinking and judging is not the same as Kant's distinction between problematic and assertoric judging. Kant classifies judgments, in respect of modality, as problematic, assertoric, or apodeictic, according as the act is a *possible* (i.e., *potential*), *actual*, or *necessary* assertion, or, in other words, according as it is a combination of representations conscious of itself as one that *might* sustain itself, as one that *actually* sustains itself (excluding its logical opposite, or contradictory, from sustaining itself), or as one that *necessarily* sustains itself (allowing no judgment whatsoever, not even a problematic judgment, to stand as its logical opposite) (see A74–76/B100–101). Thus problematic judgment involves more than bare thought: to say that a combination contains nothing in it that prevents it from being sustained is not yet to say that it is capable of sustaining itself.

It is possible, in fact, to conceive of a type of thought that lies in a combination of representations that is actually sustained (and so contains nothing that prevents it from being sustained), yet differs from judgment in being sustained only by something other than itself. Mere practical thought—bare intention (§II.3–4)—would be an example of this type of thought. For as practical, such thought is efficacious, which it could never be were it not itself somehow sustained. But whereas practical judgment is self-sustaining, indeed self-consciously self-sustaining, bare practical thought is sustained only by something outside the combination of representations in which it consists. This external support can be provided only by something that, like practical judgment, is both self-sustaining and efficacious. But inclination possesses these two characteristics. It is therefore possible to conceive of bare intention as sustained by inclination, insofar as the latter is efficacious

The conception of judgment delineated here differs significantly from one familiar way in which judgment is frequently represented in philosophical discussion. So far as a judgment is understood to be self-sustaining, it cannot be a “mental event,” such as a deciding what to think (or believe), a making up of one’s mind, or a putting together of representations. Kant does characterize judgment as an act, and as a combination, or synthesis, of representations. But he does not suppose, as some philosophers do, that to be an act at all is to be a certain type of process or event. So these characterizations do not imply that he has in mind what, say, Frege does when he describes judging as “a psychical process.”⁴ Indeed, it already follows from our earlier consideration of self-consciousness that thinking in general, and hence judging, as a type of thinking, cannot be conceived in that way: though on account of its self-consciousness judgment in Kant’s sense must be an act, for the very same reason it cannot be any type of process or occurrence in the mind, though of course all our judgments, as exercises of our discursive cognitive power, come to be in time (B1). Judgment is an act in the sense that it is an actuality. It is an actualization of the understanding or reason, but not in a sense that implies that it is a transition or a coming-to-be: the act of combination in which a judgment consists is not a *putting* together of

energetic not
synthesis

with respect to practical thought. Inclinations are fundamentally different from practical judgments and from practical thought, for though they are like practical judgments in being *self-sustaining*, in that their efficacy is just the tendency, self-strengthened into habit, of an agreeable sense-representation of an object to sustain the state of mind to which it belongs (and so also to prompt and to sustain bare practical thought where doing so contributes to the sustaining of that state of mind), and though such efficacy, like that of practical thought and judgment, is something of which the subject is *aware*, this awareness is not internal to that efficacious representation itself, but lies only in the feeling that accompanies it (e.g., the pleasure that attends the sensation produced by the object’s existence). The efficacy of an inclination is therefore not practicality, not an efficacy that is at least implicitly self-conscious. The representation’s efficacy is known only through the feeling it produces in the subject, not through the spontaneous act of thought that constitutes a practical representation’s efficacy in respect of its object. Thus in the case of inclination the awareness of the representation’s efficacy is itself a *result* of that efficacy, whereas in the case of practical thought and judgment the efficacy and the at least implicit awareness of it are the same.

⁴ Gottlob Frege, “Negation,” in P. T. Geach, ed., *Logical Investigations* (New Haven: Yale University Press, 1977), 44; cf. 42n: “If a judgment is a deed [*Tat*], it happens at a certain time and thereafter belongs to the past.” Cf. Peter Geach, *Mental Acts* (London: Routledge & Kegan Paul, 1957), 9.

judgment as
"holding"

representations, but a *holding* of them together. So we might say that to judge in this sense is to hold. We do describe ourselves as "making" judgments and as "reaching" them, and such makings and reachings can be counted as mental events or psychical processes; but the making or reaching of a judgment in the sense of interest to us here is no more the same as the judgment itself than is the making or reaching of anything else—events aside—the same as the thing made or reached.

3. Validity

So far, the two essential characteristics of judging have been considered separately. But because judging, as a self-constituted type of thinking, contains an implicit self-consciousness that includes a grasp of its own concept and extends to all that is integral to it as the act that it is, the two features must be understood together, in the idea of judgment as self-consciously self-sustaining. While the combination of representations in which thinking consists includes the implicit consciousness that there is nothing in it that prevents it from being sustained, judgment includes in addition the awareness of itself as positively self-sustaining. This distinctive character of judgment can be succinctly captured in Kant's terms, for it is nothing but judgment's *validity* (*Gültigkeit*), that is, the inner validity qualifying it as cognition and on which is based all objective validity, or validity of cognition in relation to its object. Just as we noted earlier that the efficacy by which *practical* thought (intention) can make its object actual lies precisely in its understanding itself to be efficacious, so we now observe that the capacity to sustain itself by which *cognitive* thought (judgment) can hold itself in the understanding lies precisely in its understanding itself to be self-sustaining.

Since a judgment's self-sustaining character belongs to its self-consciousness, it must rest in a distinctive *positive* unity of consciousness. Through this positive unity, each act of judgment agrees both with itself and with every other. For judgments sustain themselves individually through their agreement with themselves and collectively through their agreement with one another. That is to say: As a combination of representations, a judgment can sustain itself only insofar as the diverse components belonging to it are in agreement with one another on account of the positive unity of the act of judgment; and each judgment being one of many, judgments can together sustain themselves only insofar as all of

them, as acts of a single capacity, are positively united with one another in relations of necessary agreement, relations of mutual support and confirmation. This positive unity of consciousness goes beyond the unity of bare thought in that the latter, though it implies the absence of conflict among the components, grounds no agreement. To mark this difference, we might call judgment's positive unity the unity of knowledge. For such unity constitutes the very form of knowledge. As Kant points out, knowledge is essentially a whole, or system, of representations (A97, A645/B673), in which these representations are not arbitrarily joined together, but are connected in relations of agreement (A104–105). The original idea of knowledge is accordingly implicit in this positive unity of self-consciousness, and for this reason we can recognize in advance that every piece of knowledge can both sustain itself and confirm the others, while none of them can ever undermine itself or provide the least ground for doubting any other. The capacity to judge is thus nothing other than the capacity to know, the cognitive faculty.

Now as Kant repeatedly emphasizes, the finitude of the human (discursive) cognitive power entails that our knowledge of objects—including both knowledge of “what is” and knowledge of “what ought to be”—depends, directly or indirectly, on sensible conditions distinct from the capacity to judge itself. In the absence of such conditions, the synthetic judgments in which that knowledge consists and in which the concepts of its objects are determined (enlarged) would not be possible. So although judgment is self-consciously self-sustaining, it does not follow that the exercise of the capacity to judge does not depend on the presence of external conditions. It is implied by the very idea of theoretical knowledge that such knowledge would not be possible in the absence of sense representations that in turn depend on the object of such knowledge being present, directly or indirectly, to the senses. And we have already noted in effect that the “aesthetic” awareness of the agreeableness of certain objects is a necessary material condition of the empirical application of the concept of the good in practical judgment.

It is important, however, to describe this dependence more precisely, to secure an understanding of it that does not infringe the spontaneity implicated in judgment's self-consciousness. Knowledge stands in opposition to two quite different types of limitation or defect, mere ignorance on the one hand and error on the other, and it is only in respect of its opposition to the first of these that knowledge depends on sensible

conditions external to judgment. Though it is thanks to the presence of these conditions that we are able to advance from ignorance to actual knowledge, it does not follow that it is owing to their presence or influence that the judgments in which this knowledge consists are valid, or correct. Indeed, as the considerations to follow are intended to show, these external conditions are not themselves the determinants of the capacity to judge, and it is not to them that we must look to find the factor that directs its exercise.

Because the combination of representations in a judgment is not arbitrary, it must be in some way determined. But because a judgment's nonarbitrariness lies in its self-consciously sustaining itself through its implicitly self-aware agreement with itself and with all others, it cannot be determined from the outside, or result from anything that, by acting on the capacity to judge, would determine how it is exercised. Judgment therefore cannot lie in the passive operation of a capacity of receptivity. If the power to judge were passive, the nonarbitrary combination of representations in a judgment would have to be determined by sensibility, from a source external to the judgment and its self-consciousness. But then the self-sustaining character in which the judgment's nonarbitrariness lies could not itself be included in the judgment's self-consciousness—as it must be, belonging as it does to a self-constituting act—but would instead be something of which we could be aware only through feeling or inward observation, that is, through a kind of inner experience, and it would accordingly never be *in* the act of judging itself that we would be aware that we were judging rather than merely thinking. Nor would such awareness be in truth an awareness of our judging something, but only an awareness of our not being able to help but think it. Judgment therefore differs in kind from any sensible modification of consciousness, whether it be sensation (the effect of an object on the mind) or a feeling of pleasure or displeasure (a sentient awareness of the efficacy of a representation respectively to sustain the state of mind to which it belongs or to eliminate it) (cf. *MS* 211n, *KU* 220).

In the face of the recognition, just now reaffirmed, that the capacity to judge is not passive and inert, any lingering temptation to regard sensible modifications as conditions that do not just *enable* but actually *determine* this capacity to combine representations in some determinate way in a particular act of judgment would depend on the background supposition that without such determination the capacity would deliver, or at

least be liable to deliver, a judgment opposed to, or incompatible with, the one resulting from that determination. It would depend, in other words, on the supposition that it is thanks to such external constraint that this capacity's exercise is saved from error. This supposition, however, is not only unsupported but actually in conflict with the very idea of judgment as a combination of representations conscious of itself as a self-sustaining act of a single cognitive capacity, whose exercise can be originally conceived only as in thoroughgoing agreement with itself.

The nonarbitrariness of judgment, then, cannot be due to its having been determined by sensibility, from some external source. This independence from determination by sensible conditions pertains to judgment across the board, practical as well as theoretical, and hence in particular to choice, the type of practical judgment with which we are here chiefly concerned. In the latter case, Kant identifies this independence with freedom, stating that the free power of choice can never be determined by sensible impulses.⁵ Since judgment cannot be determined from without, its nonarbitrariness—its self-conscious sustaining of itself, or validity—must have its source within the very capacity of which it is the exercise, the cognitive faculty itself. In the act of judgment, then, this capacity is *self-determining*, even though this determination depends on sensible conditions.

The preceding reflections on the form of judgment have brought to light that judgment includes (i) an implicit understanding of itself as the exercise of the capacity to know, an understanding of itself as this capacity's spontaneous and self-determining actualization, in sensible conditions, in an act of knowledge. They also show this understanding to be *necessary*—in that it is conscious of itself as identical in all possible judgment—and so the basis of (ii) a universal recognition—a recognition in every judgment—of the validity, or cognitive standing, of every judgment and the *necessary agreement* among them that this implies. That is to say: Since each judgment's understanding of itself as knowledge is based in that judgment's consciousness of itself as an exercise of the

⁵ See *KpV* 32, *MS* 213. Though the free power of choice cannot be *determined* by sensible impulses, Kant holds that it is nevertheless *affected* by them (and to that extent is susceptible to their *influence*). This affection was characterized in other terms in our earlier reference to the material conditions of the exercise of wish (§III.1–2).

capacity to know, and since all judgments (all *possible* judgments) are recognized to be exercises of that same capacity, the ground on which each judgment's validity as knowledge relies—namely its being *an exercise of the capacity to know*—is equally a recognized ground of the validity of every other judgment. Hence every judgment contains the same implicit recognition that every judgment, on the same universally recognized ground, is valid. But insofar as judgments have this validity, qualifying them as knowledge, they must be in agreement with one another, so implicit in each judgment is the understanding that all other judgments are in necessary agreement with it as well as with one another.

4. Error

Before we proceed further, a comment is in order on what may seem to be a problem facing the account of judgment developed so far. According to this account, judgment is the exercise of the capacity to know. But such a characterization might appear to leave no way of understanding how error is even possible. It belongs to the very idea of a capacity to \emptyset that this capacity can never be the source of anything but \emptyset -ing itself, so how could the exercise of the capacity to know lie in anything other than knowledge? The foregoing treatment of judgment may thus appear, on account of its having ignored the possibility of error, to have fallen into error itself.

There is a reason, however, why error has not been mentioned in the account. In order properly to address the concern just raised, we need first to appreciate the grounds for this omission. Our aim is to articulate the formal presuppositions that lie a priori in our practical judgments, and with a view to this we have so far been considering the constitutive features of judgment in general. To bring these features into view, it is necessary to confine our attention to the concept of judgment implicit in the self-conscious exercise of the discursive cognitive capacity and to leave out of consideration everything pertaining to our judgments that we can know only by taking notice of the particular judgments we actually make, hence everything we can know only by comparing those actual judgments with one another or with what they presuppose. But it is only through such comparison that we can discover that any of our

judgments actually conflict with one another or with their own presuppositions; and it is only through such discovery of actual conflict that we can discover that any of our judgments are actually in error, or wrong; and it is only through such discovery of actual error that we come to recognize that our judgments *can* be in error. In short, we can recognize the possibility of error through its actuality, which is discovered in conflicts between or within judgments; but the converse does not hold. The actuality of error is not intelligible through a prior comprehension of its possibility, for the concept of judgment, which is presupposed in the recognition of error (in that the latter is just the recognition of error *in judgment*), provides no basis for any such comprehension. On the contrary, what this concept enables us to comprehend is the possibility of a judgment's validity, its standing as knowledge, for according to this concept our judgments are exercises of the capacity to know, so it lies in their nature to be knowledge. Hence the actuality of knowledge is intelligible through a prior comprehension of its possibility, and in that sense knowledge, unlike error, is comprehensible from itself. Thus we have a coherent conception of infallible judgment, but we cannot conceive of judgment that is intrinsically erroneous.⁶

Let us return now to the difficulty. From conflict between or within our judgments, we recognize that it is possible for them to be wrong or in error. And this possibility may seem to threaten the idea that each judgment is an exercise of the capacity to know. If two judgments conflict, one must be wrong. But how could either be wrong if both are exercises of the same capacity to know? Clearly that capacity cannot itself be the source of error, for precisely what makes intelligible a judgment's standing as knowledge, its being valid as opposed to erroneous, is its

⁶This radical difference between the ways in which knowledge and error are comprehended and recognized is reflected in the fact that, at the most fundamental level, conflict between judgments (e.g., in practical disagreements about what we should do) *surprises* and *in itself* disturbs us, whereas agreement does not. Accordingly, though this fact is often overlooked (e.g., in arguments for moral relativism or moral skepticism), and though, where noticed, it is not always taken to have the significance here suggested (cf. Williams, *Ethics and the Limits of Philosophy* [Cambridge, Mass.: Harvard University Press, 1985], 132–133), it is nevertheless a noteworthy indication of the nature of judgment. Even ironic detachment or any other form of relish that may be taken in disagreement and that may be proposed as an alternative to skepticism or relativism presupposes and depends for its possibility on this fundamental fact and what it reflects.

being an exercise of just that capacity. Nor can we suppose that the capacity to know is being exercised in the case of one of the conflicting judgments but not in the case of the other. For as we have seen, all judgments lie in the exercise of this capacity. Without this common ground, conflict would be no more possible than agreement. Wrongness or error in judgment must therefore lie in the specific manner in which the capacity is exercised on a particular occasion. It must, in short, lie in the misuse of the capacity. Even though the exercise of the capacity to judge cannot, for the reasons noted earlier, be determined by an external source, nothing rules out the possibility that this exercise might be constrained or impeded. While such constraint is entirely unsuited to serve as a general condition of the correctness of judgment, it can readily be conceived as a factor to which, in particular cases, appeal might be made in accounting for error—a factor, that is, that can make error possible by influencing the exercise, or the use, of the capacity to judge. In this way, erroneous judgment can nevertheless be an exercise of the capacity to know.

wrongness as
misuse of a
capacity

A related concern may linger. Though the possibility of erroneous judgment does not in fact threaten the idea that each judgment is an exercise of the capacity to know, it might still at first glance seem that the recognition of this possibility could undermine the understanding of itself as cognition that each judgment implicitly contains. It is certainly clear that the recognition of the possibility of error in judgment decisively shuts out any notion that a judgment's understanding or supposing itself to have cognitive standing excludes even the possibility of error. But it is also clear that the recognition of this possibility could not generally undermine the mere supposition of cognitive standing, if that supposition is indeed, as was suggested earlier, necessarily included in judgment in the sense that it is implicit in its constitutive consciousness of itself as self-sustaining.⁷ Giving up a judgment's supposition of cognitive standing would be tantamount to giving up the judgment itself, yet there can be no error in judgment without judgment. Thus, far from generally undermining the supposition, the possibility of error actually depends on it. It can, however, be said that once we come to recognize that it is possible

⁷ It is a further question, not considered here, whether the recognition of the possibility of error might lead to a skeptical suspension of judgment itself and in that sense remove the supposition of cognitive standing. (See, however, §II.7 and §III.7.)

for judgments to be wrong, the supposition of cognitive standing comes to involve a further supposition, namely that in the act of judgment the capacity to know is being properly exercised. For given the recognition that error is possible, that is, that the capacity to know may be misused, the supposition that one's exercise of that capacity is knowledge depends on the supposition that it is a proper exercise.

The results thus far obtained may be summarized as follows. According to its very idea, judging consists in the essentially self-conscious exercise of the capacity to know and so includes (i) the necessary self-understanding that judging is the exercise of this capacity. Therefore every judgment initially includes (ii) the understanding that all judgments are cognitions and hence in necessary agreement with one another. Once the possibility of conflict and error in judgment is taken into consideration, this understanding becomes (ii') the understanding that all judgments, so far as they are proper exercisings of the capacity to know, are cognitions and hence in necessary agreement.

5. Subjective and Objective Universal Validity

Let us return to the path we had been following. As we saw earlier (§IV.3), the necessary self-understanding of judgment described in (i) is the basis of the recognition of necessary agreement among judgments mentioned in (ii). This necessary agreement, as we shall now see, is of two types, which reflect two aspects of universality (or identity) in discursive cognition—subjective and objective—and two corresponding senses in which a judgment's validity can be regarded as universal.

For judgments to be in agreement is for them to sustain one another. Judgments that agree stand, directly or indirectly, in a relation of reciprocal confirmation, like mutually reinforcing effects. Such a relation implies, of course, that the related judgments are different; far from being the same as bare identity, agreement presupposes difference. But there are two senses in which we speak of judgments as differing from one another. In one sense, a judgment may be distinguished from another on account of a difference in content. In this sense, judgments are the same if they assert the same thing of the same thing, and they differ if they either concern different things or, if the same thing, say different things of it. But even where there is no such intrinsic or essential difference, judgments may be distinguished in another sense, on account of their external

differences, if they have been made by different judging subjects or by the same subject on different occasions. Judgments differing in the former sense can be said to differ merely objectively, or just in respect of a difference in their object or objects as these are represented in the judgments; judgments differing in the latter sense can be described as differing merely subjectively, or just in respect of a difference in the subject or subjects who make them. Corresponding to these two senses in which we can speak of judgments as differing, there are two types of agreement among judgments, which are complementary in that judgments that agree always agree in one of these ways, but never in both. One form of agreement holds only among judgments that, being intrinsically the same, differ only subjectively, or externally (in respect of subject or occasion). The other holds only among judgments that differ objectively, or internally (in content). The first might be called "subjective," since it lies in agreement across differences on the side of the cognizing subject or subjects, and the second "objective," since it lies in agreement across differences in the cognized object or objects. We may accordingly speak, in an extended sense, of agreement among judging subjects and agreement among objects judged, though in the primary sense the agreement is among judgments.

If agreement is so far from bare identity as to presuppose that the agreeing judgments are different, it is also so far from bare difference as to presuppose that the agreeing judgments have a common source, an identity in origin that, in a sense to be explained, grounds universality in consciousness or representation. Corresponding to the two forms of agreement just distinguished, there are two forms of universality essential to discursive knowledge, one informing its consciousness of subjects, the other its representation of objects.

The first form of agreement, among judgments that share the same content, is just the external side of the self-agreement by which an individual judgment sustains itself in the intellect. If different subjects recognize themselves to be asserting the same thing of the same thing, or if an individual subject is aware that judgments it made on different occasions have the same content, the judgments strengthen or reinforce one another. This aspect of judgments' self-sustaining nature is everywhere apparent and informs the practices and procedures that mark cognitive activity wherever it becomes scientific or methodical in character; it can be seen, for instance, in the repetition of experiments and the checking

of proofs in the sciences, and in politics and the law it is reflected in the publicity of deliberations in the framing of statutes and in the convention of precedent in their interpretation. This subjective agreement among judgments depends for its possibility on the *original universality of judgment's self-consciousness*, the universality of the judging subject's original self-awareness. This universality lies in the identity that belongs to that self-consciousness on account of its being the awareness of a single act of combination that constitutes the unity of knowledge and so is necessarily common to all cognition. And it is on account of this self-sameness of the fundamental act of cognition, this identity of the cognitive capacity itself, across different possible knowing subjects as well as different occasions of judgment that different subjects or the same subject on different occasions can, notwithstanding that difference in subject or occasion, reach the same verdict, asserting the same thing of the same thing, and find in this identity of content a confirmation of each judgment by the other. This universality in the self-consciousness of the subject's act, then, underwrites the supposition of necessary agreement—i.e., (ii)—in respect of this first form of agreement. It is this form of agreement that Kant seems to have in mind when he observes, in a passage we considered earlier (§III.2), that the concepts of good and bad require that “good and bad always be judged through reason and hence through concepts, which allow of being universally communicated, and not through mere sensation, which restricts itself to individual subjects and their receptivity” (*KpV* 58).⁸

The second form of agreement, among judgments that differ in content, lies in their coherence, in their standing together in relations of mutual support in one body of knowledge, even though they either concern different things or, if the same thing, assert different things of it. Like the first, this form depends for its possibility on the universality of self-consciousness, since diverse judgments could not confirm one another in any way at all were they not exercises of the same capacity. But it also depends specifically on an *original universality in judgments' representation of their objects*. It depends, that is, on the identity of consciousness in the original representation of the objects of cognition, on account of

⁸The case of theoretical judgment is parallel: in a judgment of experience we “will (*wollen*) that it should also be valid for us always and in the same way for everyone; for if a judgment agrees with an object, then all judgments concerning the same object must also agree with one another” (*P* 298).

which the cognitions of the represented objects are, as Kant puts it, *from the concepts* of them. That coherence among judgments depends on an original universality in their representation of their objects can be seen by drawing on an intermediate idea, that of objects' coexistence, a notion that brings with it the concepts of cause and law. It would not be possible for judgments differing in content to agree with one another were these judgments not grounded in knowledge concerning objects' law-governed coexistence, knowledge of objects' existence under laws that, in determining either how they exist (in the theoretical case) or how they are to exist (in the practical case), connect them in relations of interdependence. It is such knowledge that first makes it possible to infer from knowledge concerning the existence of one thing something concerning the existence of another, as well as to connect judgments about a thing's existence on one occasion with judgments about its existence on another. Only on the strength of such fundamental knowledge is it possible to hold, say, that if this stone here is being warmed by the sun shining on it, then so is that one over there; or similarly that if this merchant should charge a fair price to customers, then so should the next. Or again, such knowledge makes it possible to hold that if this stone is now being warmed by the sunlight, then it was cooler beforehand when in the shade; or that if an obligation to respect my neighbors' independence prohibits me from meddling in their affairs when I see that things are going well for them, then it will also enjoin me to help them recover it, so far as I can, should I discover that they have been robbed of it by some misfortune. But objects known to coexist according to laws are objects the knowledge of which is from concepts. For knowledge of such objects must represent them as things that can exist in the diverse conditions constituted by the different effects the other things with which they coexist have on them, and such representation—representation of identity across diversity—can only be through a concept. Reversing our steps, we can trace the connection again, in the other direction: In knowledge from concepts, each thing is known to be self-identical, or capable of existing in diverse conditions as the self-same thing it is, and on account of this recognized identity its action, or characteristic operation (what constitutes its existence in relation to other things), is known to be ever the same, though the conditions in which it exists are diverse. Such knowledge, however, is just knowledge of things coexisting according to laws, for because it recognizes the ac-

tion of each thing to be ever the same, it makes it possible for each thing to be known as something from which the same effect can be known to result wherever conditions are the same, and from which, where conditions differ, different effects can be known to follow in ways that reflect those differences in conditions. And it is on the strength of such knowledge of things' law-governed existence that judgments differing in content can agree with or confirm one another.

The two forms of universality just distinguished and correlated with the two forms of agreement are at bottom the same to the extent that each lies in a notion of a single, universally shared capacity, as a ground of interrelation. They differ in that the one form belongs to cognition's consciousness of the identity of the capacity to know, by which all cognizing subjects share a capacity to communicate uniting them as members in a community of knowers, whereas the other belongs to cognition's original representation of the possibility of its objects' coexistence, by which all cognizable objects share a capacity to interact uniting them as members in a law-governed system, an order of nature.

Because the universality of discursive cognition has the two aspects just described, there are two corresponding senses in which universality belongs to a judgment's validity, to its consciousness of itself as self-sustaining. This validity includes, namely, the judgment's awareness of its capability of sustaining itself, on the one side, *in every subject's cognition* and, on the other side, *in the cognition of every object falling under its concept*. Drawing on Kant's own terminology, we can mark the difference by speaking of *subjective* and *objective universal validity* (cf. *KU* 214–215).⁹ As the remarks to follow should help make clear, these two types of universal validity are implicated in the supposition of universal agreement—i.e., (ii)—and correspond to the two forms of agreement distinguished above.

A judgment has subjective universal validity, then, just if it is *valid for all subjects*—just if its validity has universality in respect of subjects. Such subjective universality does not, of course, imply that all cognizing subjects are in actual agreement. Actual agreement cannot be assumed in the case of contingent judgments, such as judgments of experience in

⁹Here “objective universal validity” has a wider sense than it does in the passage cited from the third *Critique*, where it is restricted to principles, or universal judgments. This extension is in the spirit of Kant's conception, however, and yields a certain symmetry, to be noted below, between the two sorts of universal validity.

theoretical cognition. For such subjects are distinguished from one another precisely in point of differences in the content of their knowledge, whereby one is ignorant of what another knows. Subjective universality implies rather that all subjects would agree in their judgment, or share the same judgment, provided that, with regard to the matter in question, they all had the opportunity to exercise the capacity to judge that they all share, and provided they all exercised that capacity properly.

And a judgment has objective universal validity just if its assertion, or the act of predication in which it properly speaking consists, is *valid for all objects falling under its concept*—just if its validity has universality in respect of those objects. Such objective universality does not, however, imply that the cognition of each object falling under the judgment's concept would actually agree with the cognition of every other in asserting the judgment's predicate of its object. Such actual agreement cannot be assumed in the case of contingent judgments, such as those of experience. For the objects of such judgments are distinguished from one another precisely in point of differences represented in the predicates of contingent judgments concerning them. In the case of such judgments, objective universal validity implies only that all the objects that can be brought under the judgment's concept would have its predicate asserted of them in the cognition of them provided that they were in such conditions as is the object of the judgment. The judgment that the water in the pond is frozen, for example, in relying on the concept water, implicitly involves the universal judgment that *any* bit of water, when in the conditions of the water in the pond, must be frozen—a judgment in which such conditions, though unspecified in the judgment itself, are implicitly regarded as sufficient to determine the water to be frozen.

As can be seen from the foregoing characterizations, subjective and objective universal validity perfectly mirror one another. A judgment by a particular subject has subjective universal validity in that any subject that can grasp its concept would, if in the conditions of the judging subject, share the same judgment (the same predicate, so to speak) and so be in agreement with the judging subject. A judgment about a particular object has objective universal validity in that any object to which its concept can be applied would, if in the conditions of the object judged, share the same predicate and so be in agreement with the judgment's object.

Though subjective and objective universality are both essential aspects of discursive cognition's validity, and though they mirror one another in the way just indicated, a certain difference in scope becomes apparent if we compare the cognitive, or logical, judgments that are the focus of our interest here with judgments of other types. Although any type of judgment that has objective universal validity also has subjective universal validity, the converse does not hold. Kant notes that the validity of what he calls aesthetic reflecting judgments—judgments of taste, or of beauty, for instance—has a universality that is subjective only; since these judgments presuppose no concept of their object, no understanding of the kind of thing it is, and hence are radically singular, their validity lacks objective universality even in the implicit sense just noted.¹⁰ Objective universal validity is thus *distinctively* characteristic of the judgments constituting cognition from concepts.

The objective universal validity of such cognition is a mark of its rationality. Cognition from concepts is rational cognition, cognition that lies in judgments of reason. In keeping with the traditional understanding of rational cognition as knowledge from the universal,¹¹ Kant holds reason to be “the faculty of the *determination* of the particular through the universal (of the derivation from principles)” (20:201), or, as he also puts it, “the faculty of principles” (A299/B356, *KpV* 119), through which we have “knowledge from principles,” knowledge in which we “cognize the particular in the universal through concepts” (A300/B357). Such determination, in which the particular is cognized in the universal through concepts, is carried out through a syllogistic inference of reason, yielding rational or a priori knowledge of the particular, whether it be theoretical knowledge of a particular effect, for instance the collapse of a house upon the undermining of its foundation by a man who knows that bodies fall when their supports are removed (B2), or practical knowledge of a

Reason and
knowledge from
universals

¹⁰ According to Kant's account of it, a judgment of taste is not based on any concept of its object, yet implicitly includes, along with an awareness of itself as the exercise of a certain capacity to judge (taste), the supposition that all such judgments concerning its object are valid and hence in necessary agreement with one another, or “universally communicable” (cf. *KU* §§8–9). Kant takes this subjective universality of such judgments' validity to reveal that they involve our cognitive capacity, even though they lie in a feeling of pleasure and so are aesthetic rather than cognitive judgments.

¹¹ Aquinas, for example, citing Aristotle, says, “the process of reason is from the common to the proper” (*Summa theologica*, Ia IIae, q. 94, a. 4).

certain action as something that ought to be done, such as returning a book one has borrowed (cf. *G* 412). But the idea of such determination figures in an implicit and proleptic way even in a singular judgment based on no inference, no actual derivation. In cognizing in experience that the water in the pond is frozen, one presupposes the possibility of specifying the general condition under which water is frozen and of thereby reaching a universal cognition—a principle—from which the singular judgment can be derived as a conclusion determined through the universal. For on this possibility rests the difference between knowing and being unable to help but think. **Discursive cognition, or cognition from concepts, then, is rational in that it either is from principles or at least presupposes the possibility of acquiring universal cognitions that would furnish principles from which it can be derived.** And the objective universality of its validity simply reflects the self-understood relation to principles that distinctively characterizes it as cognition from concepts, cognition that, lying in an act of combination of concepts in a synthetic judgment, presupposes both the concept it determines and the possible knowledge of a principle that can provide a basis for its act of determination.

6. *Practical Knowledge*

The conclusions reached so far have come to light through a consideration of features constitutive of rational cognition from concepts and so have application to both species of such knowledge, theoretical and practical. Indeed, they indicate that these forms of cognition are themselves but different applications of “one and the same reason” (*G* 391). We now narrow our focus to the case of practical judgment, beginning with a consideration of this difference in application.

We have noted that Kant distinguishes theoretical and practical knowledge by describing them, respectively, as knowledge of “what is” and knowledge of “what ought to be.” These descriptions may seem to depict the two forms of cognition as differing merely in respect of their subject matter, as if ought-to-be-ness (or goodness) were a special characteristic belonging to, or constituting, a certain kind of thing, which we are able to apprehend and investigate. In fact, however, they express a difference in the form of cognition, one most easily seen in Kant’s deepest and most fruitful way of articulating the distinction, by reference to

the difference in how the two types of knowledge are related to their objects. Rational knowledge, he says, “can be related to its object in two ways, either merely to *determine* this object and its concept (which must be given from elsewhere), or also to *make it actual*. The former is *theoretical* and the latter *practical knowledge of reason*” (Bix–x; cf. *KpV* 46, 89). According to this statement of the distinction, theoretical and practical knowledge have in common as rational cognition from concepts that they determine their object and its concept. Both begin with the concept of their object and proceed to determine it in an act of synthetic judgment; they thereby also determine the object, for since in determining their concept of the object each of these two types of knowledge determines itself (§IV.3), the necessary conformity between knowledge and its object must, even in the case of theoretical knowledge, be based in cognition’s determining the object rather than the reverse: “the objects must conform to our knowledge” (Bxvi). It is thus not by a difference in the direction of conformity between knowledge and its object (or “direction of fit,” as it is nowadays often called) that the two kinds of knowledge are distinguished from another. The difference lies rather in the direction of existential dependence. In the theoretical case the knowledge depends for its actuality on the actuality of its object; hence the object must, in order to be known, be “given from elsewhere” by affecting the mind. In the practical case the relation is the reverse: here the actuality of the object—as determined¹²—depends on the actuality of the knowledge. The purely formal character of this distinction can be clearly seen from its implication that so far as the efficacy by which practical knowledge works to make its object actual is complete, that object, as determined in practical cognition, is likewise an object of theoretical knowledge: “what ought to be” and “what is” are then one and the same.

¹²Since practical knowledge carries out a double function with respect to its object, in that it not only determines but also makes actual, there are two senses in which we can speak of the object of practical knowledge. In respect of practical cognition’s capacity to determine, its object is *what it determines*, namely *what it represents through its subject concept*; in respect of such cognition’s efficacy (through determining the free power of choice), its object is *what it makes actual*, namely *what it represents through its determination of the subject concept* (or what we might term the *content* of the knowledge). In the former respect the object of my practical knowledge that I should repay the money I have borrowed is myself, in the latter it is myself repaying it.

Though abstractly formulated, this characterization of practical knowledge agrees with the account outlined earlier (§II.6–9), which emphasized the causality as well as the rationality of practical knowledge, portraying a practical judgment as an exercise of practical reason that efficaciously determines the good. But consideration of the just noted difference in form between theoretical and practical knowledge helps bring into view a critical feature of the latter that was only implicit in the earlier discussion. It positions us to see that practical knowledge distinguishes itself from theoretical in that its determination of its *object*—the good—must also be the practical determination of the judging *subject*. It reveals, that is to say, that the subject of practical knowledge is necessarily the same as the object this cognition determines.

Kant's statement of the distinction between practical and theoretical knowledge does not, it is true, expressly identify the object of practical cognition with the subject. It does, however, imply that the object to be made actual through the efficacy of such knowledge cannot, just as such, be "given from elsewhere." And once this efficacy of practical cognition in respect of its object is taken into consideration, the identity of the subject and the object determined can be seen to follow directly if we bear in mind that, as was explained earlier, such efficacy is essentially self-conscious. Knowledge that is practical, or self-consciously efficacious, not only has causality in respect of some effect, but has this causality only through its understanding itself to have it. In such understanding, the conception of this causality is just that very causality itself, and therefore the subject of that conception must likewise be identical with the subject of the causality—in other words, the *cognizing subject* must be the *acting subject*, or agent, and indeed must be the agent precisely through being practically conscious, in its practical cognition of what it is to do, of itself as to be such a subject, or agent.¹³ (This consciousness, in which the *cognizing* subject constitutes itself as *agent* through regarding

¹³ Here the infinitival form serves to mark the content as that of practical knowledge. Although Kant often characterizes practical knowledge as knowledge of *what ought to be*, he also holds that "ought" (*Sollen*) expresses the relation practical cognition bears to an imperfectly rational being (cf. §V.4). If we abstract from this relation, we must distinguish practical knowledge from theoretical by describing it as knowledge, not of *what is*, but of *what is to be*, or as knowledge, not of *how things are*, but of *how they are to be*. But since, as we have just seen, practical cognition is self-cognition, we may also speak of it as knowledge of *what and how one is to be* and of *what one is to do and how one is to act* (or, alternatively, as knowledge of what it would be good to be or to do).

itself as *to be* agent, is itself at bottom identical in all practical cognition; as a formal a priori practical proposition that is “synthetic” precisely in that in it pure intellectual self-consciousness determines itself to be practical, it is pure practical reason’s own self-consciousness, or (what comes to the same) the act in which pure reason constitutes itself as practical, what Kant terms “the fact [i.e., deed] of reason” (*Faktum der Vernunft*) (*KpV* 31).) Thus whereas theoretical knowledge can be of objects distinct from the subjects who cognize them, practical knowledge is always knowledge cognizing subjects have of what *they themselves* are to do.

Of course, nothing in the idea of theoretical knowledge implies that it is not possible for the subject of such cognition to acquire theoretical knowledge of itself. Indeed, this must be possible if practical knowledge itself is to be possible, for the subject of practical knowledge cannot be distinct from the subject of theoretical knowledge (any more than there can be two reasons, one practical and one theoretical), and there can be no practical knowledge unless “what ought to be” can be the same as “what is.” Even so, an essential difference remains, for in no case is the object of theoretical knowledge in its original conception identical with the cognizing subject, yet this original identity is always understood in practical knowledge. The original concept of an object of theoretical knowledge is of something that, existing independently of the knowledge of it, must in order to be known be “given from elsewhere” by affecting the senses (something that can appear) and so not necessarily the subject itself; the original concept of an object of practical knowledge, on the other hand, is always of the subject itself, but conceived, and thereby constituted, as agent. Only in the case of practical cognition is it not even thinkable that the object be distinct from the subject.

Since the subject of practical knowledge and the object it determines must be the same, such knowledge must be not only *self-determining* in the sense noted earlier—the sense in which all rational cognition from concepts, whether theoretical or practical, is self-determining—but also *self-knowledge* on the part of the subject, though of course *practical* self-knowledge and so not knowledge of how one is faring or what one is doing, but rather knowledge determining what one is to be or to do, efficacious knowledge that works to bring about its object and the theoretical knowledge of it. Practical knowledge is therefore self-determining in a double sense. All rational knowledge is self-determining, yet such

self-determination is merely formal, the self-determination of cognition. But since knowledge determines its object as well as itself, and since practical knowledge is a *self-knowledge* on the part of the cognizing subject, practical cognition's determination of its object is also the self-determination of that subject.¹⁴ That is to say, it is a determination of the subject by the subject not only in respect of its bare capacity to know, but also in respect of its existence, its capacity to act. Hence, the self-determination distinctive of practical cognition—*practical self-determination*—is *material* as well as *formal* self-determination.

7. *The Form of Practical Knowledge*

As practical self-determination, practical knowledge constitutes a relation of the subject to itself that is not found in theoretical knowledge, a relation in which the subject makes itself its own object. In this regard practical knowledge is no different from practical thought generally. As practical self-specification, all such thought constitutes a relation to self on the part of its subject. But the self-relation of practical knowledge differs from that of bare practical thought in a critical respect. The difference is readily apparent if we consider the self-relation of practical knowledge in the light of the features we have seen to be constitutive of rational cognition in general (§IV.5). Doing so will reveal that this self-relation belongs to practical knowledge across the board, figuring not only in wish and choice, but also in practical judgments of the secondary types, such as one person's judgment concerning what it would be good for another to do, even though these latter judgments obviously lack the self-relation characteristic of bare practical thought.

As we saw, rational cognition is characterized by a universality that has both a subjective and an objective aspect, each grounded in a conception of a single shared capacity. On the subjective side, this conception is of an identical capacity to know, by which all cognizing subjects

¹⁴Theoretical cognition's determination of its object is also subjectively (i.e., in the cognizing subject) a determining of sensibility to represent the object in intuition. Practical cognition's determination of its object—the subject—is also subjectively a determining of the (free) power of choice to make actual what the cognition determines the subject is to be or to do. So while practical cognition's determination of the object is not itself the *making actual* of the object (in either sense of "object"), it is where possible the determination of the power of choice *to* make actual what it represents in its determination of the object.

share a capacity to communicate uniting them as members in a community of knowers; on the objective side, it is cognition's original representation of the possibility of its objects' coexistence, by which all cognizable objects share a capacity to interact uniting them as members in a law-governed system. But we have just seen that practical knowledge is distinctive in that, on account of its efficacy, its object and its subject must be one and the same. It follows that in the case of such cognition these two conceptions necessarily coincide, so that every subject that shares the capacity for practical knowledge must, precisely through possessing it, likewise be a member of the system of coexisting, interacting agents represented in such cognition. We also saw that the two-sided universality of rational knowledge entails that such cognition's validity is likewise universal in two corresponding senses, that rational knowledge is both valid *for* every subject capable of grasping the concept determined in such cognition and valid *of* every object falling under its concept. Given the self-relation of practical knowledge, it follows that in such knowledge subjective and objective universal validity necessarily coincide in the sense that the subjects for which the cognition is valid are the very beings to which it applies.

This necessary coincidence of practical cognition's subjective and objective universal validity entails that the act of practical self-determination is inherently universal, in that any particular act of practical cognition, whether a practical judgment of the primary or the secondary type, is always, as knowledge of the particular in the universal, based in doubly universal knowledge, knowledge *every* practically cognizing subject can have of what *every* such subject is, or ought, to do. Practical self-determination is, accordingly, a universal self-relation, whereas the practical self-specification characteristic of bare practical thought is merely particular, indeed singular. Practical self-determination is never the bare self-relation of an isolated practical *I*; it always has a footing in the self-relation of the *we* of practical knowledge.

We now have before us an articulation of *the form of practical knowledge*. As rational, practical knowledge has a double universal validity; in this regard it is no different from its theoretical sibling. As practical, or efficacious, it is marked by an identity of subject and object, a self-relation it shares with practical thought generally. But being at once rational and practical, it has a self-relation that is itself universal: its two forms of universal validity necessarily coincide.

These formal characteristics of practical knowledge—universality, self-relation, and their combination in universal self-relation, or self-related universality—are all recognizable as such solely in the self-consciousness of practical judgment.

8. The Presupposition of Universality

Because the two necessarily coinciding forms of universal validity together constitute the common form of practical knowledge, the particular practical judgments figuring in persons' wishes and choices bear a necessary relation to them. For these judgments are exercises of the capacity for such knowledge, even though they also have the particular self-relatedness characteristic of bare practical thought on account of their dependence on empirical conditions.

As we noted earlier, a particular subject's wishes, in which certain ends are deemed simply good, depend on the pleasure the subject finds in certain agreeable objects; and the subject's choices, in which certain actions are deemed good on the whole, are founded on its wishes while also depending on its empirical theoretical judgments of practicability. Wishes thus depend on conditions that are "internal" in the sense that the enjoyment of an agreeable object lies in the subject's sensible awareness of its own state as an animal being, while choices, in addition to presupposing the wishes internal to them, rely on conditions that are "external" in that judgments of practicability contain the subject's empirically-based cognizance of its capacity to produce effects in things outside it. (Each of these two types of awareness is of course itself internal to the subject, but external to the capacity for practical knowledge.) In both cases, however, the objects deemed good are represented—problematically in the one case, assertorically in the other—as to be made actual through the judgment itself by the very subject making the judgment, a judgment that in turn depends in the end on the operation of the capacity to feel pleasure and displeasure in that same individual subject.

Yet these particular judgments also bear a necessary relation to the universal. In the first place, the exercise of the capacity for practical knowledge in the judgment figuring in wish or choice has, in its own understanding of itself, objective universal validity. While empirical conditions are relied upon in the making of such a judgment, the end or

action is implicitly understood to be such as would be appropriate for any subject capable of practical knowledge, provided that it is in the conditions on which the judgment is based—that is to say, the corresponding conditions, or such conditions as those on which the judging subject is relying. In the case of wish, these conditions lie in a corresponding satisfaction of the sensible criterion for the use of the concept of the simply good; in the case of choice, they include both corresponding wishes (“internal” conditions) and corresponding judgments of practicability (“external” conditions). In either case, the judgment implicitly attaches its predicate to any subject who is in empirical conditions such as those on which it itself relies, whatever differences in conditions there might be in other respects. So while the practical judgment is explicitly a singular judgment, a judgment in which I specify what *I* would do or mean to do by determining what, in the conditions, it would be good for *me* to do, it implicitly regards itself as nested in a universal judgment, a judgment determining what, in such conditions, it would be good for *anyone* to do. It is accordingly “knowledge from principles,” practical knowledge in which “I cognize the particular in the universal through concepts.”

Second, the objectively universally valid self-determination also has subjective universal validity. It would be in agreement with *any* practically cognizing subject’s practical knowledge, so that, were all subjects who share the capacity for such knowledge to judge in the light of the same conditions and with regard to this same universal subject of the judgment, they could all share the very same universal practical knowledge through exercising this shared capacity, whatever differences there might in other respects be in the extent of their knowledge. Moreover, they could share this universal knowledge not only *in abstracto*, in the armchair of reflection, but also in use, in their interactions with one another in all particular situations in which this knowledge applies; for the derivation of actions from universal knowledge is as much an exercise of the capacity for practical cognition as is the cognition of the universal itself (cf. *G* 412).

As an exercise of the capacity for practical knowledge, then, a particular practical judgment contains in its self-consciousness an understanding of itself as in accordance with the form of such knowledge, an understanding lying in the presupposition that *it is possible for every subject with the capacity for practical knowledge to share (not only in abstracto*

but also in use) the practical judgment that every such subject is to act as determined in the particular judgment when in the conditions on which it is based. This might aptly be labeled the presupposition of the self-related double universal validity of practical knowledge, but it will be convenient to have a more succinct designation, so we may simply call it *the presupposition of universality*.

This presupposition articulates a certain possibility-condition to which practical judgment is subject, a condition lying in the possibility of the self-related universality implicated in the validity of practical cognition. This condition is fundamentally different from the “internal” and “external” conditions noted above, on which practical judging also relies. Whereas the latter both lie outside the capacity for practical knowledge itself, the condition expressed in the presupposition of universality must be internal to it. For every possible practical judgment, in presupposing or understanding itself to be in accordance with the form of practical knowledge, also presupposes that identical form: the idea of this form is what all these judgments, through their presuppositions, have originally in common. Indeed, since the capacity is understood through the form, we could just as well say that the capacity is in the form as that the form is in the capacity. Constituting the capacity itself, this form is prior even to its exercise; being presupposed in all practical judgments, it is not itself any such judgment. Unlike the conditions lying outside the capacity for practical knowledge, which are requisite for this power to have materials on which to exercise itself, this possibility-condition bears on a judgment’s very constitution, its validity and standing as practical knowledge. A practical judgment can count as practical knowledge, and its object as an object of such knowledge, only if this formal condition is satisfied. The material conditions, then, are conditions of practical cognition’s actuality, while the formal condition is a condition of its very possibility.

All of these conditions might be described as presuppositions of practical judgment, since each stands in a relation to practical judging recognizable from the latter’s own self-understanding. But because we have been able to articulate the presupposition of universality just from a consideration of practical judgment’s own form or constitution as an act of the capacity for practical knowledge, we can mark it off from the others by describing it as a *formal presupposition* of practical judging. To engage in practical judging at all is already to have one’s practical thinking

informed by it. Indeed, as the original understanding of the form of practical knowledge itself, it is just practical judgment's self-understood relation to *the first principle of all practical knowledge*, its relation, that is, to the original, purely formal universal to which all other practical knowledge, as knowledge of the particular in the universal, is related.

Although in the above formulation the presupposition of universality is set forth as unitary, it can also be considered as containing a pair of presuppositions, corresponding to the two types of universal validity figuring in the form of practical knowledge. In particular, it can be seen to include, first, the presupposition that there can be a universal practical judgment in which the subject's particular judgment is contained, and second, the further presupposition that it is possible for every subject capable of practical knowledge to share this universal judgment. Both presuppositions are of course satisfied in practical knowledge. But nothing said above rules out the possibility of instances in which the first of them might be empty, cases where there could be no such universal judgment, much less a universally shareable one. Nor has the possibility been excluded of cases in which such a universal judgment, though possible, could not be universally shared.

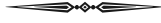
9. *Summary Observation*

Up to this point, we have been simply elaborating what is formally presupposed in practical judging. We have made no attempt to specify the concept of the good, to determine where goodness lies. We have not asked which substantive judgments are valid and which not. Nor have we investigated what any particular person or type of person might actually regard as good.

Yet precisely because the presupposition has been expounded independently of such considerations and merely through an examination of the concept of practical judgment, we are in a position to say that all practical knowledge—and hence also the good, as the object of this knowledge—will necessarily be in agreement with this presupposition, in that the latter is just a judgment's understanding itself to be in accordance with the form of such knowledge. We are also able to say that should certain claims about what is good turn out to be incompatible with the presupposition, then they cannot be sustained and so cannot be practical knowledge. For by being in conflict with the understanding

of practical knowledge that they involve merely in virtue of being practical judgments, they will in that sense be in conflict their own form and hence with themselves and so will lack the self-sustaining character essential to cognition, a character that they themselves, as exercises of the capacity for such cognition, purport to have. Therefore, if there turn out to be possible practical judgments that conflict with this presupposition, then the latter places a necessary constraint—an unconditional, categorical constraint—on the exercise of the capacity for practical knowledge. Action represented in a judgment in violation of such a constraint could be regarded as wrong or bad in itself, just as action represented in a practical judgment determined by and hence in necessary agreement with this presupposition could be regarded as intrinsically and unconditionally good.

We next consider the possibility that this presupposition provides a categorical constraint. In doing so, we shall return to the concept of the will, considering practical reason from the side of its practicality, while keeping in view what the foregoing investigation of its cognitive power has brought to light.



Constraints on Willing

1. The Form of Willing: As Constitution, or Essence

As was stated earlier, the will, as Kant conceives of it, is practical reason. Practical reason is the capacity for practical knowledge, the capacity exercised in practical judging. So willing is practical judging, and the form of practical knowledge presupposed in such judging is equally presupposed in willing, as *the form of willing*. This form, as we have seen, lies in universality; it is found in the idea of self-related double universality that resides in the presupposition of universality. As form, it is one and the same in all possible exercise of the will and so can also be characterized as the form of the will itself. Even though the presupposition of universality that figures in one volition, or practical judgment, differs from that figuring in another in that each relates a different volition to the form of universality, they share in common the idea of that form.

A particular will in an individual person is distinguished from practical reason itself in that it constitutes itself as a particular will through its willing a particular content, a particular end, determining the free power of choice in acts of wish and choice in the light of the internal and external conditions described earlier. As it is thus by its *acts* of self-determination that a particular will is distinguished from practical reason, which is a universal capacity, or power, a particular will can also be described as a particular actualization of the power of practical reason.¹ Maxims, the

¹ The distinction between particular wills and the will in general, or practical reason, may call to mind Rousseau's distinction between particular wills and the general will. Kant does sometimes speak in a way that suggests a willingness to liken practical reason to a

principles constituting the first *exercise* of a person's free power of choice in its wishing and choosing, are the fundamental acts of such self-determination.

But while it is thus by maxims that individual persons are first constituted as distinct from one another, maxims all share in common the form of willing, the idea of universality that resides in the presupposition of universality. As Kant says, every maxim has "a *form*, which consists in universality" (*G* 436). And corresponding to the different ways in which a maxim can be described, we can recognize two ways in which this universality is involved. Since a maxim is a "subjective principle" in the sense that it is a "principle on which the subject *acts*" (*G* 420n), its presupposed objective universal validity is also explicitly represented in its content, though only to such an extent as to cover the possible occasions of conduct of the individual person in question: its subject is still an *I*, not a *we*. But since a maxim is also a "volitional opinion" (*KpV* 66), or an exercise of the capacity for practical knowledge, its universal validity implicitly extends more widely, reaching, in the presupposition of universality, to persons generally. Even where the maxim prescribes a specific end or type of action in the light of certain specific conditions, and even where it depends on an influence on the will's exercise by a condition regarded by the subject itself "only as valid for the will of the subject" (*KpV* 19), it still involves the presupposition that all persons could act from a common recognition that every person is to act accordingly when in such conditions as those on which its prescription is based.

Thus far, our portrayal of the form of willing has merely underscored its essential involvement in all exercise of the will. To indicate such involvement, it is not necessary to speak of form in anything more than a

general will, as when he indicates that if everyone "does what he ought," then "all actions of rational beings take place just as if they arose from a supreme will, which embraces all private power of choice within itself, or under itself" (A810/B838). But while Kant's idea of the will as practical reason has a certain affinity with Rousseau's idea of a general will, it is clearly not the same; for though a Rousseauian general will is general in relation to the particular wills that are its members, it is constituted through a particular act—a social contract—and therefore is particular in relation to the universal power of practical reason. The undeniable affinity might be expressed by saying that the universal power provides the general form of a general will; like any other particular will, a general will is an actualization of the power of practical reason, but it constitutes itself as a will that is general in relation to further wills in that the latter are constituted through it and thereby belong to it as its citizen members.

minimal sense, a sense broad enough to cover even form that is immediately constitutive, having no normative or regulative function. Form in this broad sense can be present even where there is no possibility of a conflict between content and form. Logical consistency, for instance, belongs to the form of conception and thought. Since inconsistency lies in an opposition, a mutual exclusion, among thoughts, and since thoughts that exclude one another cannot together belong to the unity of thought, there is absolutely no conception that is not logically consistent. Similarly, according to Kant's doctrine of space and time as the forms of human sensibility, it is not possible for our outer perception to be other than spatial.

2. *As Norm and Principle of Self-Agreement (Morality)*

We speak of form in a more specific, normative sense when we speak of the form of willing. Form in this sense comes into view when we shift our attention from the activity of thinking in general to the acts of judgment in which the capacity for rational cognition is exercised, and it implies both that all validity of such judgments depends on their content's "conformity" with form, and that the form can determine the exercise of the capacity it constitutes so that this conformity is maintained.

That form thus characterized is not to be equated with immediately constitutive form is clear. A capacity's exercise cannot be conceived as determined except insofar as it is possible at least to conceive of the capacity's being exercised otherwise than as it is determined, and in the case of a capacity to judge, this is as much as to say that the judgments arising through its determination are synthetic, judgments to which contradictory thoughts can be opposed. So there can be form in the normative sense only where content that conflicts with form is conceivable. Form in this stronger sense cannot properly be described as "empty" in either of two related senses in which this expression might be taken. For even though it is present universally, in all exercise of the capacity in question, form in this sense is not something to which all conceivable content conforms. Nor is it a mere abstraction, devoid in itself of all power to determine content to conform with it. Form in the stronger sense is thus both "contentful," or substantive, and determining.

As noted, form in this stronger sense has a specific normative character. It serves as a norm, or canon, a standard of validity regulating the

employment of a cognitive capacity. This normativity stems from a certain preeminence or priority that the form is recognized in the self-consciousness of judgment to have over the content. Form has this priority since, being recognized as necessary, it is in every case the same, common to all possible employment of the capacity, independently of how the capacity is exercised, whereas the content is always contingent upon that exercise. In Kant's memorable phrase, the form is "firm even though it is neither dependent on anything in heaven nor supported by anything on earth" (*G* 425). And given that it is possible at least to think or to conceive of a conflict between the form and the content, the recognition of this priority of the form amounts to an understanding that because form is necessary whereas content is contingent, maintaining consistency or agreement between form and content must lie in conforming the content to the form rather than in adjusting the form to fit the content. And since judgment's validity—its self-consciously self-sustaining character—depends on such agreement, validity depends on content's conformity with form. Form in this sense is thus a norm, or standard of validity. Moreover, this standard is not an external rule, one that could have an effect on the capacity's exercise only through being imposed on it from without. It is an internal standard with its own regulating and determining power. For since the self-sustaining character of judgment itself depends on content's conformity to form, the form is nothing separate from it. Indeed, as we have seen, the form is just the self-recognized condition of self-agreement under which judgments can sustain themselves at all. It is, as it were, the life-principle of judgment. So the form must itself be capable of determining the exercise of the cognitive capacity. (In a nutshell, the relation between the two characterizations of form so far presented can be expressed by saying that it is *constitutive* of judging in general that it is *determinable—regulated—*by the form *constitutive* of its own validity.)

Since the will, as practical reason, lies in the capacity for practical knowledge, the form of willing is a form in this normative sense. Practical knowledge, or validity in practical judging, depends on maintaining the will's exercise and content in agreement, or conformity, with the form of such knowledge, and that form is just the form of willing. So if reason can be practical at all, or if there can be any such thing as practical knowledge, knowledge of the good, then it must be possible to exercise the will in such a way that its content agrees with its form. As the capacity

for practical knowledge, then, the will is a capacity to determine the concept of its object in conformity with its own form. And since the form is necessary and indeed essential to the will itself, the will can also be characterized as a capacity to determine itself in accordance with its own form, so that its content conforms to its form, where this self-determination—an instance of what we earlier called the formal self-determination of rational knowledge—is at the same time a determination of the will's exercise and content by that very form. In sum, since the form of willing is just the form of willing's own validity as knowledge, and since this form can determine the will's exercise so that the content agrees with it, it has normative authority with respect to all exercise of the will.

The form of the will, then, is prior to its content. It has priority as the a priori recognized form of validity in willing, and for this reason the will's exercise is subject to it as to an unconditional, or *categorical*, normative requirement. Because this fundamental form can determine the will to exercise itself in a manner that is in agreement with it—and indeed is the *only* thing that can determine the will to be in *nonaccidental* agreement with it—and because such agreement is an agreement of the will with its own form and hence with itself, this form can be described as a principle of *self-agreement*, a principle of unity, or identity.

It bears emphasizing, as an implication of what has just been said, that this principle of the will's self-agreement requires more than that the will's content be in accidental, or contingent, agreement with its form. Accidental self-agreement is not enough, since it falls short of the presupposition, involved in any exercise of the will, that this exercise itself is valid and as such self-sustaining. In order for the will's self-agreement to be compatible with this presupposition, it must stem from the will's being determined by the form essential to it as practical reason and thereby itself disposed to the proper exercise of practical reason. Therefore, the presupposition of universality involved in willing is the basis not only of the requirement of bare self-agreement, or what Kant calls "legality," but also of the requirement that the self-agreement have a necessary, or nonaccidental, character, which constitutes what he calls "morality," the source of the intrinsic goodness, or moral worth, of the resulting action.

In a will in which there is such necessary agreement, the form is mirrored in the content and the latter thereby expresses the former's determining power: *what* is willed is, at least in effect, that the content

be in agreement with the form of practical knowledge. Accordingly, in such a will it is always possible for the form itself to be viewed as if it were part of the content and thus contained as part of what is willed, and this indeed would appear to be Kant's meaning when he identifies such a will—"an absolutely good will"—with one "whose maxim always can contain itself, regarded as universal law, within itself" (*G* 447). And to the extent that a will in which there is such self-agreement is in turn efficacious, or successful in carrying out the action it wills, that same form will be present in the action as well.

3. *As Practical Law, or Law of Autonomy (the Moral Law)*

A yet more determinate characterization of the form of the will can be reached when we take into account the point just noted, that the will, in addition to being a capacity for judgment and cognition, is a causal power in that, as a capacity for *practical* cognition, it is the power persons have to determine, in the exercise of the free power of choice, their conception of how they are to be and to act and thereby to determine their own existence. Since the will is a causal power, its form can be described as a *law*, and in particular as a *practical law*, in Kant's understanding of these expressions. And borrowing another of Kant's terms, we can say that the character of the will in virtue of which its form is specifically a *practical law* is its *autonomy*. (Kant also maintains that this law is already known to us pre-philosophically under another name, as the *moral law*—a claim that will concern us in the sections to follow.)

According to Kant, laws necessarily, and hence universally, determine the existence of things (cf. *KpV* 19, *MAN* 468). They are represented through principles—universal cognitions from which particular knowledge can be derived—that specify how objects, or objects of a certain kind, act or operate. Given, then, that the form of the will is necessarily involved in all exercise of the will and capable of determining that exercise, and given that this exercise is of a person's causality, or power to determine its own existence, that form has an efficacy through which it can determine persons' existence and so can be characterized as a fundamental law of the will.

This law, however, is of a specific type. Corresponding to the distinction between theoretical and practical knowledge, there is an essential difference between the laws represented in these two types of cognition.

Occasionally, Kant marks the difference by simply exploiting his standard formulas for distinguishing theoretical and practical knowledge. Thus, echoing the contrast between knowledge of “what is” and knowledge of “what ought to be,” he speaks of laws “according to which everything happens” and laws “according to which everything ought to happen” (*G* 387–388; cf. 427). But as we have seen, he has a more illuminating way of articulating the difference between theoretical and practical knowledge (§IV.6), and this affords a more fruitful way of understanding the difference between the two kinds of law as well. A law represented in theoretical knowledge—a theoretical law—can determine the existence of the things subject to it independently of its being known by those same things. The laws determining the motions of the planets, the oxidation of hydrogen, and so forth, operate independently of whether the objects determined by them know them or have any cognitive capacity at all. It stands quite otherwise with practical laws. The self-relation of practical knowledge entails that a law represented in such cognition is one whose efficacy, or capacity to determine the existence of the beings subject to it, and therefore whose very being as a law, depends on its being represented in the knowledge—the practical knowledge—those very subjects have of themselves, of how they are to exist (*KpV* 43; cf. *G* 412). Hence it is *only through* the efficacy, the practicality, of practical cognition that a practical law is itself efficacious. Since a practical law’s capability of determining the existence of things—a capability essential to its being as a law—is a capability of determining that existence *only through* the practicality of the knowledge in which that law is represented, a practical law is nothing independent of the representation of it in practical knowledge, and indeed is nothing but practical knowledge itself. It follows as well that *the form of practical knowledge*, the form represented in the presupposition of universality, is just *the form of practical law*.

It was noted just now that the form of the will is a *law*, the will’s fundamental law. Since as we have just seen this law has its efficacy only through the efficacy of the practical cognition in which it is represented, it is a *practical law*. But since the form of the will is nothing but the form of practical knowledge, and since, as we have also just now seen, the form of such knowledge is the form of practical law, the form of the will is a *practical law that lies in the form of practical law*.

The distinctive practical character of this fundamental law—the dependence of its efficacy, or capability of determining the existence of

things, on its being known by the very beings whose existence it can determine—is directly tied to what Kant calls the will’s autonomy, its character as self-legislative, or as a law to itself, which he says is the highest principle of morality (*G* 440). For a law whose efficacy, and so whose very being as a law, depends on its being known by the beings whose existence it can determine is precisely a self-legislated law. Just as practical knowledge is self-knowledge, so practical legislation—the legislation of practical law—is self-legislation. A practical law is accordingly a self-legislated law, a law whose legislation is such that the subjects to whom the law is given are necessarily the very subjects in and through whom it is given by the practical reason, the capacity for practical knowledge, that they share in common. When Kant criticizes previous moral philosophers for conceiving of the will as heteronomous, he is thus claiming in effect that their moral theories are just that—*theories*, theoretically cognitive accounts in which the fundamental law is conceived as theoretically rather than practically knowable, as a law that determines a subject’s will or power of choice independently of the subject’s cognizance of it (*G* 441–444, *KpV* 33, 43).

Kant’s general characterization of a practical law is broad enough to allow us to draw a distinction between a practical law that is completely formal and as such knowable wholly a priori, and a practical law that is “material” in that it prescribes action in certain specific (inner and outer) material conditions. Since the fundamental law we have just been considering is recognized in practical judgment’s formal presupposition, in which the representation of the form of practical knowledge is equally the representation of the form of practical law, it must be a practical law of the first of these two types—a purely formal practical law, recognized a priori, through the self-consciousness essential to practical judgment in general. As the form of willing, it constitutes the form with which all exercise of the will must nonaccidentally agree in order to qualify as practical knowledge and practical law. We have noted that maxims, as practical judgments, implicitly regard themselves as having a universality that extends to all practical subjects, even though it may be only in the light of certain specific conditions that they prescribe their end or action. In view of this implicit regard, we can characterize maxims that prescribe their end or action in the light of certain specific conditions (“material” maxims) as putative practical laws of the second type. Provided they are determined by and hence in agreement with the form of

practical law—the will’s fundamental practical law—they are in fact the practical laws they purport to be.

The distinction between formal and material practical laws might on first inspection seem to entail that, aside from the one fundamental formal law, practical laws must be empirical, prescribing action in specific material conditions, conditions discoverable only through experience and hence not such as can be known to hold for every person. And this appearance might in turn foster the expectation that no a priori recognizable organization or system is to be found in practical cognition. It may therefore be useful, at this point, briefly to recall our earlier examination of fundamental practical judgments, bearing in mind that in practical knowledge the application of its formal condition follows the order of practical knowledge itself, which as rational cognition proceeds syllogistically from the universal to the particular. Doing so will enable us to describe how the two main organizing divisions among the practically cognized relations in which persons stand arise entirely a priori.

When we considered Kant’s statement that we know a priori that every human person has happiness as an end, we sought to determine the grounds on which it rests, and concluded that it could be traced to the concept of a person, which includes the thought of the first and fundamental act of the free power of choice wherein a particular person first constitutes itself as such through making happiness an end (§III.4). Though this act does depend on material conditions, it does not rest on any specific conditions, or conditions that could hold for some but not others and so distinguish one particular person from another. It is thus knowable a priori and constitutes the first application of the fundamental formal practical law.

Implicit in the act of making happiness an end, we noted, is the consciousness of the act’s dependence not only on formal conditions but also on the general material conditions just mentioned, conditions that include the operation of the subject’s capacity to feel pleasure and displeasure, the receptivity of the practical cognitive capacity. And through its awareness of dependence on the latter, this act is also conscious of itself as a *particular* act, a particular actualization of the power of practical reason, distinguishable from its form and hence also from other possible such acts, other possible persons, sharing that form. This consciousness makes possible not only the distinction between practical

judgments concerning oneself—that is, judgments of the primary type, wish and choice—and practical judgments concerning others, but also a distinction in kind between a priori practical laws bearing on particular persons in respect of their relation to themselves and laws bearing on them in respect of their relation to one another. The universal validity of practical cognition implies intrapersonal as well as interpersonal validity in judgments, and hence entails agreement among judgments made by the same person on different occasions as well as agreement among judgments made by different persons. Since the latter agreement is grounded in the identity, across all persons, of the capacity for practical knowledge (practical reason itself), whereas the former is grounded in the identity, across all judgments made by a particular person, of that capacity *in* a particular person (a particular will), there is a basis for a difference in kind among a priori practical laws. Conditions of interpersonal agreement in judgment constitute laws bearing on particular persons in their relation to one another; conditions of intrapersonal agreement constitute laws bearing on particular persons in their relation to themselves.

We have seen in addition that the fundamental act of making happiness an end includes two distinguishable though necessarily united moments: the wish to be engaged in the agreeable activities that are the objects of one's inclinations, and the wish for self-sufficiency in respect of these objects (§III.5). The fundamental practical law thus has application in each of these acts of the free power of choice, in each of these fundamental wishes. Through this twofold application, the fundamental law yields two a priori practical laws on each side of the distinction just noted between relations to self and to others. Together, these provide the basis of a system of practical laws relating particular persons to themselves and to one another.

In respect of the relation of particular interest to us here, that of persons to one another as members of what we might call a practical

²These two laws are, respectively, the expression in choice and action of the practical-cognitive attitudes of respect and love, which united together make up the original reciprocal relation holding among persons as subjects sharing the power of practical cognition. The former attitude relates to the capacity to represent, in practical cognition, the universal cause (the form of such cognition), the latter to the capacity to represent, in such cognition, the particular effect (the matter). (Cf. *MS* 448–450.) In exercising the former capacity, persons are independent of one another; in exercising the latter, they are interdependent.

world, the fundamental practical law divides into a law of natural justice and a law of beneficence.² Because self-sufficiency is necessarily an end for every person, an end already represented in the act that first constitutes a person, persons are subject to a fundamental negative law maintaining them in a material independence from one another that reflects their formal (cognitive) independence, a law not to interfere with others' pursuit of their happiness. And because for every person the practical representation of happiness includes, beyond the wish for self-sufficiency and independent of it, the wish for the material component of happiness, persons are connected together by a fundamental positive law, to help others in need, through which the formal interdependence belonging to them as (practical) cognizers is expressed in a material interdependence uniting them as agents. In this way, the essential division of the practical conception of the end of happiness is the basis of a division of the specific practically cognizable relations among persons derivable from the fundamental law into negative laws of justice and positive laws of beneficence.

It is noteworthy that the two laws just described arise *directly* from the form of practical knowledge through its application in the fundamental act of making happiness, generically conceived, an end.³ Thus, the practical judgment by which a person frames this end will amount to practical cognition provided that, through being made in accordance with the form of such cognition, it has the double universal validity discussed earlier. And since this judgment comprises both a wish for self-sufficiency and also a wish to engage in the activities one finds agreeable, each of these two component judgments will likewise have the double universal validity of that form. Hence, on the one hand, the wish for self-sufficiency will be nested in a practical judgment, to which every person can agree, asserting that such sufficiency is to be secured universally, in everyone's case. And on the other hand, the act of representing activities one finds agreeable as belonging to one's end will similarly be situated in a judgment, to which all can agree, asserting that, so far as is possible, the activities persons find agreeable are to be realized. Thus, the form of practical knowledge, through its application in the twofold fundamental

³They can also be brought to light *indirectly* by considering maxims that conflict with them, such as the ones Kant discusses in his examples of false promising and the refusal to assist others in need. These examples will be discussed below (§VII).

act of the power of choice, contains under it the two a priori practical laws just described.

Moreover, nothing said so far entails that these two laws, along with corresponding laws bearing on persons' relation to themselves, exhaust the system of practical law. At a subsequent stage in the application of the fundamental practical law, Kant introduces the empirically determined concept of human nature. To the extent that this concept is itself framed in accordance with the a priori idea of a natural end (cf. *KU* §64), we have reason to anticipate that further organization will inform more specific practical laws bearing on human beings considered merely as such, in their relations to themselves and to one another. Obviously, however, contingency and indeterminacy will increase as we descend toward the concrete in human practical cognition, proceeding from person to human being to citizen, from natural law to positive law, from ethics to casuistry.

4. As Practically Limiting Law, or Categorical Imperative

So far, we have progressed through a series of increasingly determinate characterizations of the form of the will: as constitution, or essence; as regulative form (or norm) and principle of self-agreement; and finally as law, specifically as practical law, or law of the will's self-legislation. We are now in a position to take one further step and to characterize this form more determinately still and in the relation Kant claims a practical law stands to *our* wills, the wills of human beings: taking this step will bring us to the idea of form as "practically limiting law" (*KpV* 32). To do this, we must take into account a certain fact about the nature of our human will that is known empirically, from our awareness of actual conflict between its exercise and the fundamental law that can determine it. Through encountering this conflict we discover that the nature of our will does not guarantee a necessary agreement between its exercise and that law. Moreover, even where our willing is in conformity with the law, the mere presence of accompanying feelings of reluctance, reflecting the presence of opposing inclinations that are being directly or indirectly thwarted, is by itself enough to reveal at the very least that we cannot know the will to be so constituted by nature as to guarantee an agreement between its willing and the law. We therefore recognize, Kant holds, that the human will is not perfectly or inviolably good in its

nature.⁴ In the case of such a will, he maintains, the specific relation of determinability in which it stands to its own principle is one of *constraint*, and this relation is accordingly called obligation; the practical necessity of the law in such a will amounts to necessitation; and the law is expressed in practical consciousness as a command, formulated as a categorical imperative (*G* 412–413, 439, *MS* 379).

5. *Morality and Prudence*

The feelings of reluctance just noted make it possible for what is in truth the principle constituting the very form and inner nature of the will to be viewed as an external constraint, a rule that, having no original footing in our makeup, must be traced to some other source, such as social custom or convention, and which we can regard as related to reason only to the extent that we can see that in our present condition it would be to our advantage to follow it. When the principle is thus viewed from the standpoint of the inclinations that can be thwarted, our earlier characterization of it as a principle of the will's self-agreement may seem off the mark. From this perspective, the principle of prudence or self-interest may strike us as a more suitable candidate for that designation, and more properly describable as a principle of practical reason. A brief comparison of these two principles, however, will show that in the most fundamental sense it is not the principle of prudence, but the fundamental practical law, expressed in the human will as a categorical imperative, that is a principle of self-agreement for the will, as the capacity for practical knowledge.

Prudence expresses its demands in hypothetical imperatives, and so far as a person's choices conform to these demands, they are consistent

⁴This recognition does not amount to a denial that humanity, our rational nature, is unconditionally good (see §VI.6–7 below); it does, however, require that a distinction be drawn between two fundamentally different elements within our human nature, which Kant marks by distinguishing between "predisposition" and "propensity" (see *R* 26–32). This distinction raises in turn the question whether such a conception of human nature can accommodate the idea of psychic harmony involved in the traditional idea of virtue, or whether it entails that for Kant the moral ideal must lie in the domination of inclination by reason, or in what Aristotle would classify as continence rather than virtue. This topic greatly exercised Schiller, Hegel, and others in Kant's day, and it has also received much recent attention; but though of considerable importance in its own right, it lies beyond the scope of this study. For a survey, see Allison, *Kant's Theory of Freedom* (New Haven: Yale University Press, 1990), chap. 10.

with the realization of the necessary end of happiness on which the imperatives depend. Now as we have seen, it is in the wish for happiness that a particular will in an individual person—a particular practical self—first constitutes itself as such (§III.4). So there is an undeniable sense in which this demand of prudence can be described as a condition of self-agreement in the will. In effect, it requires of a particular will that its choices be in agreement with the wish that constitutes its very identity as a particular will.

Such a characterization of prudential imperatives is potentially misleading, however, in that it fails to reflect the specific empirical condition that ushers in the requirement of prudence and defines the scope of its application. The general problem to which the exercise of prudent intelligence is a response would never have been encountered were it not for the fact that practical subjects are aware, from their *experience* of failure, that their productive powers are limited, that *would* does not entail *can*. This awareness is of course reflected in the empirical theoretical judgments of practicability that constitute the external material conditions of persons' choices. The original problem of prudence is therefore to practically represent, so far as is possible, the production of the fundamental object of wish (what one *would* do) in conformity with the recognized conditions of practicability (what one *can* do), making it a suitable object of choice. Thus, the requirement of prudence does not itself include a demand that we realize our happiness; on the contrary, it presupposes from the start the wish for happiness, on account of which persons are always already pursuing this end, so far as they can. What prudence requires is rather that one's pursuit of this necessary end be in agreement with the knowledge one has, or should have, of the actual circumstances of action. On such knowledge—empirical knowledge of “what is”—hypothetical imperatives in every case depend. There is no single form here with which the exercise of the will is always required to be in agreement in the pursuit of its end, but only the awareness of circumstances, infinitely rich in their variety, though of course through practice and experience uniformities are revealed and established, upon which general counsels of prudence can be based (cf. *G* 416–419).

The same point can be seen more directly from the concept of happiness itself, for the demand that choice agree with empirical theoretical cognition merely reflects the nature of the object prudence serves. According to its very concept, happiness lies in a certain relation between

individual human persons and the conditions in which they exist. As we have seen, the enjoyment constituting the material of happiness lies in the feeling, or inward sensible manifestation, of a self-sustaining relation between subject and object, in which the inclinations expressing the subject's lifepower stand to representations of their objects' *existence* in a relation of mutual furtherence. Thus the very feeling in terms of which happiness is conceived lies in a consciousness of a harmony between the subject's sensible desires and its experience. But this experience is empirical theoretical knowledge, which, as knowledge of "what is," depends for its actuality on the actuality of its objects, objects that can be so cognized only through being "given from elsewhere" by affecting the senses; it thus depends on the actual external conditions of the subject's existence. To suppose that prudence is a principle of absolute self-agreement in willing would thus be to overlook the "hap" in "happiness," the "Glück" in "Glückseligkeit," the essential element of contingency that belongs to happiness on account of its being a matter of how individual human persons are consciously related to the external conditions in which they exist.

The general injunction of prudence just stated comprises two more specific requirements, one negative, the other positive. Negatively, prudence forbids me from exercising choice in a way that conflicts with my theoretical knowledge of what, given the actual conditions of action, is possible, requiring that the contents of my choices be jointly possible, or practicable. But theoretical knowledge of something as possible is just the recognized possibility of theoretical knowledge of it (as actual). To know that a person *can* do something is in effect just to recognize that the possible judgment that the person *will* do it (or *is* doing it or *has* done it) can be integrated into one's body of theoretical knowledge. If I know that I cannot make an omelet without breaking eggs, this is thanks to my implicit appreciation that the things I already know exclude from the things I know or could ever come to know—that is, from in principle knowable things—that I have made (or am making or will make) an omelet without breaking eggs. So what prudence forbids are choices the theoretical knowledge of whose contents is not possible in the sense just indicated. Whatever the objects of wish might be, prudence forbids any exercise of the power of choice that conflicts with this cognizability condition, there being no other way in which conflict can be avoided. Positively, prudence enjoins one to make the will's conformity to theoretical cognizability

complete and necessary, so far as is possible, through making it one's end to ensure that the totality of the contents of one's wishes as well as choices remain within the bounds of theoretical knowability. (This positive requirement leaves open to what extent this agreement is to be secured through maintaining moderation in the inclinations on which wishes depend, and to what extent through the development of one's natural capacities and talents with a view to increasing one's capacity effectively to pursue ends.)

In demanding that the will's contents be in agreement with the condition of theoretical cognizability, the requirement of prudence parallels the law of the will described above, which as we have seen demands that the will's contents all be in agreement with the form of practical knowledge, the condition of practical cognizability. And in this latter case too negative and positive requirements can be distinguished. Negatively, the law forbids the willing of actions and ends that conflict with this form; positively, it enjoins one to make the agreement complete and necessary, so far as is possible, through making it one's end to ensure that all the contents of one's will (wishes and, through them, choices) are in conformity with this form.

The principle of prudence is thus like the law of the will in that both are normative principles governing the exercise of the will in its determination of the power of choice. In each case, agreement with the principle is agreement with a condition of cognizability. Prudence keeps choice within the bounds of what is theoretically knowable, morality keeps it within the bounds of what is practically knowable. But as we have seen, theoretical and practical knowledge differ in that the former is of what is given from elsewhere, whereas the latter is essentially self-knowledge (§IV.6). Hence the two principles differ in that prudence conforms the will to the subject's theoretical knowledge of what lies outside the will (outer nature), while the law of the will conforms the will to its practical knowledge of itself—knowledge of its form or constitution as the capacity for practical knowledge (inner nature) (cf. *KpV* 57–58). In short, prudence keeps the will in agreement with its external conditions, morality keeps it in agreement with itself. In the case of both sorts of consistency the will is exercised rightly. But only in the latter case is the rightness, the validity, intrinsic; in the former it is relative to, or dependent on, theoretical knowledge. And similarly intrinsic wrongness is to be found only in willing that conflicts with the law of the will.

None of this, of course, touches the point that in the human case the consciousness of the fundamental law of the will is attended by feelings of reluctance, which, though themselves reflections of the law's efficacy, reveal that the human will is not by its nature perfectly good. The human will can accordingly be in agreement with itself only through constraining itself. The principle of morality remains a principle of the will's self-agreement, then, but only through also being a principle of *self-constraint*. So conceived, the principle brings with it the ideas of ethical duty, moral feeling (respect), and virtue, the notions of discipline and culture, and a variety of other familiar traditional ethical concepts that receive treatment in Kant's doctrine of virtue (cf. *MS* 379–382).



INTERPRETATION

VI



The Categorical Imperative

1. Overview

The last two sections explicated Kant's identification of the will with practical reason by elaborating his conception of practical reason as the capacity for practical knowledge. By outlining an account of practical judgment as the exercise of this capacity, we identified the form of practical knowledge, and in a series of steps we traced practical judgment's presupposition of this form down to the point where it emerged as the formal practical (or self-legislated) law of willing, encountered in the human will as a categorical imperative. With this key in hand, we now return to Kant's exposition of the categorical imperative in the *Groundwork*, with a view to gaining a better understanding of its main ideas. We begin with a brief survey of Kant's three formulations of the imperative, noting their relations to the idea of practical knowledge and to one another, and then proceed to a closer examination of the formula of universal law, of particular interest on account of the special position to which Kant assigns it in the exercise of moral judgment (*G* 436–437) and also because it has so often been regarded as particularly problematic.

2. The Three Formulas

It is clear from the conclusions reached so far that the form of practical knowledge, lying in the necessarily coinciding subjective and objective universality of such cognition's validity, is precisely captured in Kant's

idea of the autonomy, or self-legislation, of the will, a lawgiving in which the subjects to whom the law is given are necessarily the very subjects in and through whom it is given by the practical reason they share in common. In the *Groundwork*, this idea of a legislation to which every will is subject just through being at the same time legislator is explicitly introduced in the third and final formulation of the categorical imperative, “the idea of the will of every rational being as a universally legislating will”: according to this idea, Kant explains, “the will is not merely subject to the law, but so subject that it must be regarded as self-legislating and for this very reason first subject to it (of which it can consider itself the author)” (G 431). But while it is not until this final formula that we see the form of practical knowledge unmistakably shining forth, this does not mean that this form is not also operative in the first two formulas, “act only according to that maxim through which you can at the same time will that it become a universal law” (G 421) and “so act that you employ humanity, in your own person as well as in the person of every other, always at the same time as an end, never merely as a means” (G 429). On the contrary, there is good reason to suppose that it is.

An indication can be found in what Kant himself says about how the three formulas are related. After completing his initial presentation of them, he refers to them as “three ways of representing the principle of morality” and tells us that they are “fundamentally only so many formulas of the very same law, one of them of itself uniting in itself the other two” (G 436). Though this remark is somewhat obscure, the thought it expresses is elaborated and clarified in the ensuing paragraphs, as Kant recapitulates his exposition of the three formulas and traces the relations between them. There he argues that the first two formulas are “fundamentally the same” (G 437–438) and then, immediately after explaining why this is so, advances to the third formula, arguing that “from this it follows incontestably” that every rational being “must be able to regard itself, with respect to all laws to which it may be subject, at the same time as universally legislating” (G 438). Kant’s thought here cannot be that the third follows neither from the first nor from the second alone, but only from both taken together. For he has just argued that the first “says just as much as” the second. His point is evidently rather that the third contains explicitly united in a single idea—the idea of the will as self-legislating—distinguishable yet mutually entailing aspects, one of which is prominent only in the first, the other only in the

second. Such an interpretation also fits with his comment that the difference between the formulas is “subjectively rather than objectively practical” (G 436), for by this he seems to mean that it is a difference, not in what they represent or prescribe, but in the manner or the extent to which these formulas, as different ways of regarding the same law, are factors in the latter’s efficacy in a particular subject’s will.

What is newly introduced in the third formula, then, is the *explicit* recognition of the will’s autonomy, the *articulate* awareness that rational beings, as beings with wills, are subject to laws only through also being the legislators of those laws. This autonomy, as we noted earlier (§V.3), is just the self-relation of practical knowledge (as self-knowledge), through which its subjective and objective universal validity necessarily coincide. But though Kant portrays this explicit recognition as arising in the third formula, it is also true that the idea of the necessary coincidence of the two sorts of universal validity is implicitly at work in the first and second formulas, as he understands them.

Consider the first. It is clear that in its explicit mention of universal law this formula highlights the objective universal validity of practical cognition. But the idea that persons subject to the law are also legislators is not altogether absent. The subjects to whom the imperative is addressed are directed, not just to act on a maxim that could *be* a universal law, but to act on a maxim they can *will* as a universal law. It is true, of course, that in considering whether it is possible to will a maxim as a universal law, the thought is not yet explicitly in place that the law to which one would be subject, were the maxim a universal law, would be such a law only *through* one’s willing it, but since the will is a self-conscious causality, the idea of a possible willing of a maxim as a universal law cannot be separated from the thought of a possible or problematic connection between the act of willing and the establishment of the law willed, nor therefore can it be divorced from the thought of possible legislation. Thus, in employing this formula one relies not only on the idea of a law to which all are subject, but also on the thought of oneself as a being who, while *subject* to that law, is also *willing* it. Indeed, in considering whether one can will one’s maxim as a universal law, it is *through* conceiving of oneself as willing the law that one conceives of oneself as subject to it. This is the very thing one must be able to do in order to be a self-legislating subject, for such a subject stands only under practical laws, which according to their very idea are laws to which one cannot

conceive oneself as subject except through conceiving oneself as willing them.

There remains, of course, a noteworthy apparent difference between the first formula and the third, in that the first seems to rely only on the thought of one's *own* agreeing to everyone's acting on one's maxim, rather than the idea of *everyone's* agreeing to everyone's acting on it. In other words, the first formula seems to hold a maxim to the idea of objective but not subjective universal validity. This difference, however, is "subjectively rather than objectively practical." Here we should first recall that Kant makes clear in his discussion of the formula of universal law that the question it raises, the question whether it is possible to will, or agree to, everyone's acting on one's maxim, is to be answered by determining whether the attempt to do so results in a *contradiction*. It is therefore to be decided *a priori*, by reason, not by consulting how such a law would bear on one's own particular wishes, which distinguish oneself from other persons. This means that every person who asks this question will reach the same answer so far as they are properly exercising their capacity for practical knowledge. Kant cannot, therefore, be supposing that the first formula differs from the third in such a way that, while different persons' proper application of the third in connection with the same maxim would always yield the same result, the proper application of the first would yield different results depending on who applied it, one person being able to agree to everyone's acting on the maxim, another not. This consideration is not by itself enough to show the two formulas to be equivalent, of course, since it is one thing to say that if one person's attempt to will a maxim as a universal law results in a contradiction then so will any other's, but another to say that such an attempt must itself involve an attempt to represent other persons as also willing the same law. However, as we will see when we examine the first formula more closely below (§VI.3), when one considers whether a maxim can be willed as a universal law, the idea of everyone's agreeing to, or willing, that law is tacitly implicated in one's very thought of everyone's being subject to it.

One other apparent obstacle to supposing that the idea of self-legislation is implicit in the formula of universal law is Kant's claim that the latter can also be expressed in terms of the idea of a universal law *of nature*: "act as if the maxim of your action were to become through your

will a *universal law of nature*" (G 421). The significance of this claim will be considered below (§VI.4). For the moment, it is enough to observe that even here, where it is conceived as a law of nature, the universal law is thought of as coming to be *through* the exercise of one's will.

In the case of Kant's second formula as well, we can clearly discern the involvement of ideas it does not directly express. As we will see below when we consider it more closely, in its mention of humanity, or rational nature, this formula directs attention to persons' *capacity* for practical knowledge, a capacity whose exercise begins in the act of setting an end (§VI.6). It thereby accentuates such cognition's subjective universal validity. But as we will also see, the idea of practical cognition's objective universal validity, even though not highlighted in the formula, figures essentially in the argument by which Kant arrives at it (§VI.7). Examination of this argument will confirm his assertion that the first and second formulas are "fundamentally the same."

When we consider the first two formulas, then, bearing in mind what Kant says about them, we will find both of the coinciding forms of universality alive in his understanding of each, though with this difference, that in the first formula prominence is given to the *objective* universal validity of practical knowledge, the validity of such knowledge in respect of its objects, while in the second emphasis is placed on practical cognition's *subjective* universal validity, the validity it has in respect of cognizing subjects. In other words, rational beings are first principally considered as *cognized* and so as subject to law, following it in their action; they are then principally regarded as *cognizing* and so as having the capacity to recognize law and thereby spontaneously to agree to a rule proposed as a candidate for law—to deem it good, relating it to an end—if it is indeed eligible. In subsequently coming explicitly to recognize that rational beings, as spontaneously cognizing subjects, are cognizable only under laws they themselves recognize as valid for themselves, we explicitly grasp the necessary coincidence of the subjective and objective universality of the cognition's validity, therein seeing that the cognition is practical cognition, self-cognition, and that the law cognized is practical law, self-legislated law, authored solely by the will itself. The third formula thus comes into view as we comprehend at once the two sides or faces of practical cognition and practical law, faces that were first thrown into relief through being separately emphasized in the first two.

Once these points are noted, we can see in the order of the formulas the analytic sequence of practical cognition's ascent, within its philosophically articulated form, to full self-consciousness. This sequence complements and mirrors the upward progress, in the first section of the *Groundwork*, from common moral rational cognition to philosophical by way of the three propositions set forth in the analysis of the concept of duty (*Pflicht*).¹ It also reflects the relations the formulas bear to the history of practical philosophy. For it represents a development *from* the idea of a *law of nature* (a law of theoretical reason), which is central to the natural law tradition and in which the notion of objective universal validity stands in the foreground, *through* the second formula's idea of humanity as an end in itself, in which prominence is given to the idea of subjective universal validity and therewith also to the idea of *free, uncoerced agreement* between persons that lies at the core of the social contract tradition, *to* the third formula's idea of all subjects jointly giving themselves a *law of freedom* (a practical law), in which both forms of universal validity are equally explicit and in which these two traditions are united, each finding its truth through the other.²

¹ One great disadvantage under which the anglophone reception of Kant's practical philosophy has labored is the lack of an adequate English term for *Pflicht*. Translators all agree in using "duty," and they really have no option, as "obligation," the only viable alternative, is needed for *Verbindlichkeit*. When we hear the English word "duty," we are prone to think of debt and subordination to a superior, the relation of servant to master, serf to lord. Small wonder that English-speaking readers have found it difficult to comprehend how an analysis of morality that begins with this concept can arrive at the idea of autonomy, or self-legislation. It is otherwise with *Pflicht*. This term, cognate with the now archaic English "plight" (in the sense of *pledge* or *engage*, as in "trothplight"), contains in its root meaning *both* the passive idea of being bound and also the active, participative ideas of binding or engaging oneself and siding with. Kant's ear is true; the concept he is expounding contains in germ the self-relation that lies originally in the self-consciousness of practical knowledge. And like the three formulations of the categorical imperative, his three propositions (*G* 397–401) articulate the different moments belonging to the form of practical knowledge, but as they present themselves *in concreto* (and so in the aspect of necessity rather than universality) in pre-philosophical moral consciousness. Thus, when he takes up the common concept of *Pflicht* with a view to developing the idea of the good will's morally worthy (unconditionally good) action, he proceeds *from* the first proposition's representation of such action as *determined (necessitated) by* the law of *Pflicht* (not by inclination), *through* the second proposition's portrayal of this determination as *self-determination*, or as the conformity of the maxim of willing to the will's own a priori principle, *to* the unification of these two in the third proposition's idea of *Pflicht* as the necessity that an action has through the subject's respect for the law constituted by that principle.

² These moments correspond, approximately, to the stages of dogmatism, skepticism, and transcendental self-critique that figure in Kant's portrayal of the history of speculative

3. *The Formula of Universal Law*

A closer examination of Kant's first formula will confirm that the universalization test figuring in it can be traced to the formal presupposition of practical judging identified above (§IV.8). This will make it possible to understand the test along the lines proposed at the outset, as serving to bring the form of willing explicitly into view, with the result that, in the case of morally impermissible action, an intrinsic wrongness is made manifest.

Kant arrives at this formula—"act only according to that maxim through which you can at the same time will that it become a universal law"—through an analysis of the concept of a categorical imperative (G 420–421). According to this analysis, such an imperative comprises two elements, reflecting the cognitive and desiderative moments of practical cognition. It includes, first, the law to which one's maxim is to conform, a law that must be unconditioned (since the imperative is categorical, not hypothetical) and that therefore can lie only in "the universality of a law in general," or in what might be called the form of law (cf. §V.3). In addition, a categorical imperative contains "the necessity that the maxim be in conformity with this law," a necessity that is practical, but which, owing to the imperfection in the nature of a will subject to imperatives, amounts to necessitation, the law in such a case being represented as a command (cf. §V.4). In keeping with this analysis, we can regard the formula of universal law as comprising two elements: first, the thought of a certain possibility-condition, the possibility of being willed as a universal law; second, the command to act only on maxims that satisfy this condition. We have, then, two points to consider in comparing this formula with practical judgment's formal presupposition.

philosophy. Just as the dogmatism of traditional rationalist metaphysics tends to arouse skepticism and empiricism, so the presumption that ethical obligations trace to a law of nature can lead to doubts that encourage the attempt to understand them through the ideas of contract and agreement. In both cases, attention shifts from the objects of cognition to the subject and its capacity to know. And just as dogmatic rationalism and skeptical empiricism, notwithstanding their opposing attitudes of certainty and of doubt, both represent genuine and complementary aspects of human reason's capacity for theoretical knowledge, so the natural law and social contract traditions, despite the apparent conflict between the ideas of law and freedom, both express genuine and complementary aspects of human reason's capacity for practical knowledge.

As to the first, we have already noted that the presupposition of universality implicit in practical judgment and hence in choice and its maxim³ also expresses a possibility-condition (§IV.8). In the case of choice this presupposition lies in the thought of the possibility of all subjects with the capacity for practical knowledge sharing the practical judgment asserting the goodness of every such subject's acting according to that same choice when in the conditions on which it is based—that is, according to the same maxim, or rule of choice. We have also noted that the principal apparent difference between this possibility-condition and the one expressed in the formula of universal law is that the latter seems to include only one of the two conditions figuring in the former, in that it directs our attention to the condition that reflects the objective universality of practical cognition's validity, but does not mention the condition of agreement among practically cognizing subjects that would reflect its subjective universality.

We can take a first step toward determining whether this difference is more than merely superficial by asking how the subject to whom the formula of universal law is addressed is to understand the scope of the universal law the formula demands the maxim be willable as. The universality in the representation of a law lies in the assertion of a predicate in respect of *all* that falls under the subject concept. But what is that concept? Obviously Kant cannot be thinking it to be, in the first instance, the concept of, say, a Prussian, or a philosopher, nor even the empirically determined concept of a human being. The formula is supposed to express a pure a priori principle, valid for every rational being (*G* 420n, 425), and Kant's intent is clearly that it is merely the concept of such a being that is to figure in the subject position in the thought of the universal law the formula requires one to be able to will one's maxim as, for in his analysis of the concept of a categorical imperative he explicitly

³ As was pointed out earlier, Kant says that maxims have their source in the free power of choice (*MS* 226) and are principles "on which the subject *acts*" (*G* 420n). We can thus regard the maxims of a particular subject as belonging to—indeed, as together making up—the specific practical self-conception that this particular will constitutes through its wishes and choices. As principles of choice and hence of practical judgment, maxims are a particular will's general self-determinations and purport to be sufficiently grounded in general material conditions of the sort to which wish and choice, as species of practical judgment, are always essentially subject: "internal" conditions (pleasure in the agreeable) in the case of wish, and in addition "external" conditions (empirical theoretical judgments of practicability) in the case of choice.

indicates that this law contains no condition that limits its universality (G 420–421). In considering whether one can will a maxim as a universal law, then, one is to conceive of that law as a law for all rational beings—a law, that is to say, for all beings to whom the formula of universal law can be addressed. But such beings are necessarily beings conceived as having the capacity to determine whether they can will a maxim as a universal law for all such beings. And beings with that capacity are just beings who share the capacity for practical knowledge and hence have the capacity to agree to everyone’s acting on one’s maxim if it has the validity of practical knowledge. According to Kant’s understanding of it, therefore, even though the formula of universal law does not explicitly direct attention to the condition of agreement among subjects that reflects the subjective universal validity of practical knowledge, it does nevertheless employ an idea of universal law in which the beings subject to the law are conceived to be precisely the beings who have the practical cognitive capacity that would make it possible for them to share the universal practical cognition in which that law would be represented.

The points just noted, though important, do not quite bring us to the conclusion that this possibility of all subjects’ sharing such universal practical cognition—the possibility represented in the presupposition of universality—is just the possibility to which attention is directed in Kant’s formula, the possibility of willing a maxim as a universal law. For they leave open the question whether the universal law mentioned in the formula is meant to be considered as nothing more than a merely theoretically knowable law of nature—a question to which an affirmative answer may seem to be suggested by Kant’s statement that the formula of universal law can be recast in terms of the idea of a universal law of nature (G 421). To reach the conclusion that the two possibility-conditions are the same, such an answer must be ruled out.

We can do this easily enough by adopting the following reasonable interpretive assumption, further support for which will be provided below (§VI.4). We may suppose that the law mentioned in this formula is to be understood in a manner that fits with the propositions concerning laws that Kant laid down earlier as the starting point from which he elaborated his account of a categorical imperative: “Every thing in nature works according to laws. Only a rational being has the capacity to act *according to the representation* of laws, that is, according to principles, or a

will" (G 412). Given this initial characterization of the distinctive way in which rational beings stand to laws, it is reasonable to assume that the law mentioned in the formula is to be conceived accordingly, namely as governing beings that on the one hand are included among the things in nature, all of which operate according to laws, but on the other are distinguished by the fact that, as *rational* beings, their capacity to act according to a law *lies in* their capacity to act from their representation of it.

When the formula is so interpreted, the possibility-condition it expresses is the same as that of the presupposition of universality. For the representation of law from which rational beings are capable of acting must be an efficacious conception of that law *as a law* and thus lie in the practical knowledge that the way of acting it generally represents is practically necessary, or good. It must therefore include, in its consciousness of its own validity as knowledge, the thought that this mode of conduct can be recognized as good by all subjects who share the capacity for practical knowledge. Consequently, it must be a judgment that is at least implicitly conscious of itself as one that all such subjects do indeed share and from which they accordingly act, provided that they are cognizant of the relevant conditions, properly exercising the capacity for practical knowledge, and not hindered in their judgment or action. Thus any subject who acts according to a law only through representing it as such presupposes that, subject to the conditions just noted, any subject who can share that same representation of the law, that same recognition of the goodness of that way of acting, does so and acts accordingly. So when the law mentioned in Kant's formula is understood in the way just suggested, we can spell out the formula's possibility-condition as follows: for it to be possible to will a maxim as a universal law is for it to be possible for everyone capable of practical knowledge to act from the shared idea of the goodness of their all acting from the same representation of the maxim as a law. A law in this sense—a practical law—would thus just be practical knowledge, having the self-related double universal validity expressed in practical rational judgments' presupposition of universality. Though this double universality is not fully explicit in Kant's initial presentation of the formula of universal law, it can be seen clearly in one of his later statements of it: "act according to a maxim that at the same time contains in itself its own *universal* validity for *every* rational being" (G 437–438; emphasis added).

The second element contained in the formula of universal law is the command to act only on maxims that satisfy the stated possibility-condition. It is clear from the foregoing account of how the formal presupposition of practical judgment constitutes a constraint on willing (§V) that this second element does not amount to the addition of anything beyond what is already involved in that presupposition. Because each maxim, as a practical judgment, purports, through its essential if implicit self-understanding, to be a valid exercise of practical reason, it already supposes itself to be in conformity with the possibility-condition just described. Hence it can be in agreement with its own presupposed form only through conforming to this unconditional condition. And to the extent that this condition can really determine and even constrain the exercise of the power of choice so as to secure such conformity, there is real practical power lying in the form of willing itself, the power of an unconditional command of practical reason expressed in a categorical imperative. Therefore the categorical imperative, to the extent that it finds expression in the formula of universal law, originates in the form that all maxims, as practical judgments, share in common.

4. Practical Laws and Laws of Nature

Although it lies beyond the scope of this study to provide a comprehensive treatment of the formula of universal law, the interpretation just outlined will receive further articulation as we proceed. We can begin to bring it into sharper relief by marking a few of its distinctive features. One of them lies in the way it conceives of the universal law mentioned in this formula. Another has to do with the sort of reasoning it allows to enter into the universalization test. It will also be useful to consider what implications the practical-cognitivist interpretation outlined here has for our understanding of how the formula of universal law and the formula of humanity are related.

It should be evident already that according to the account here presented, the fact that there is a formula of the moral law that involves a conception of universal law at all traces to the fact that the judgments (in the form of maxims) whose formal presupposition this formula serves to express have an at least implicit objective universal validity, which marks them out as judgments of reason, as “knowledge from principles,” in which we “cognize the particular in the universal,” so that the willing of

a maxim is always “at the same time” the implicit willing of it as a universal law. But the most noteworthy feature of the way in which the law mentioned in the formula is here being understood traces to the fact that we are taking these judgments to be practical rather than theoretical. If we take the judgments to be practical, and if we suppose that the formula serves to express their presuppositions, then we must suppose that the law mentioned in the formula is a practical rather than a theoretical law (that is, a law of the sort described above in §V.3). Thus it was assumed in our above discussion of the formula’s possibility-condition that this law is to be understood as one that would govern beings whose capacity to act according to a law lies in their capacity to act from their representation of it. Making this assumption enabled us to bring our earlier account of the presuppositions of practical judgment to bear on Kant’s formula by taking the law mentioned in this formula to lie in practical knowledge, knowledge with the distinctive self-related double universal validity that we have seen expressed in the presupposition of universality and that we have found to be grounded in the self-determining character of practical judgments, as judgments in which subject and object are one and the same.

In view of its obvious importance for the interpretation here presented, the assumption that the law mentioned in the formula is a practical law merits further discussion. Indeed, further consideration is in any case called for, given that this assumption may seem to conflict with Kant’s statement that this formula can also be expressed in terms of the idea of a universal law of *nature*. Examining this remark will enable us to underscore a further important aspect of the understanding of the idea of law provided by the interpretation here proposed.

After his initial presentation of it, Kant says the formula of universal law can also be stated as follows: “act as if the maxim of your action were to become through your will a *universal law of nature*” (G 421). And he then uses this idea of a law of nature in the immediately ensuing discussion of the four famous examples of duties derivable from this formula. He later suggests that in offering this variant formulation he means only to indicate that the universal conformity to law of actions bears a similarity or analogy to an order of nature (G 431, 437). But it is often supposed that in speaking of a law of nature here Kant has in mind a law or regularity of the sort that might be investigated in an empirical science dealing with the special constitution of our human nature, such as psychology

or anthropology—for example, that human beings laugh when amused, that they cry out when in great pain, or that their behavior tends to conform to that of their peers and associates. It is a characteristic feature of such laws that their efficacy does not lie in their being practically known or represented as the laws they are by the beings whose existence is determined in accordance with them. Indeed, these laws are operative regardless of whether those beings know them in any way at all. Problems arise, however, if Kant is taken to be thinking of laws that have this feature. If for instance we consider the maxim that figures in his second example, the maxim of making a false promise when in need of a loan that one foresees one will not be able to repay, we will find it difficult, to say the least, to see why the attempt to conceive of this maxim as a universal law of nature must result in a contradiction, as Kant says it must, if the law in question is of the sort just described. For if the circumstances of need that according to the maxim are to trigger a false promise occur rarely enough or are sufficiently difficult for others to detect (a possibility whose likelihood increases in the case of maxims containing narrower specifications of the triggering circumstances, to say nothing of maxims in which the specifications include, out of prudence, say, or a feeling of shame, the stipulation that one have assured oneself that the chances of others' detection of one's inability to repay are very small), then the law might not be discovered, in which case it would evidently be possible for those who would act on it to do so and to be successful in their deception. But no one thinks that whether this maxim is indeed contrary to duty is contingent on whether such a law would be discovered.

Such problems do not arise if we suppose that the idea of a law of nature that Kant introduces is to be understood in a way that does not require us to depart from the idea suggested above, that the law mentioned in the formula of universal law governs rational beings, beings whose capacity to act according to laws lies in their capacity to act from their representation of them. If this supposition is correct, then the possibility will not be excluded that a universal law of nature can be a universal law of our *rational* nature, or of what Kant calls humanity (*G* 429), and accordingly the consideration of whether a maxim can be willed as a universal law of nature can be understood to involve the attempt to conceive of it as a law that all rational beings necessarily follow out of their shared recognition of its validity, its rational necessity, or in other

words as a law that is followed by all rational beings in the way it would be followed by a perfectly rational being, a being who is conscious of principles of practical reason not as imperatives, but simply as the laws that all such beings necessarily, or *naturally*, follow out of their shared knowledge of those laws' rational necessity (cf. *G* 412–414). Such a law would manifest itself in experience insofar as the actions of the rational beings who determine their nature through their practical knowledge of this law would themselves belong to the order of nature knowable through experience, but the law itself would not be knowable *only* through experience. There would be another way of knowing it, indeed a more original way, in view of which we could say—borrowing the terms of one of Aristotle's famous distinctions—that, in the case of such a law, what is “better known to nature” is also better (but *practically*) known to us.

This way of understanding the variant formulation's mention of a law of nature receives ample support from what Kant says elsewhere concerning the concepts of law and nature. Particularly striking confirmation can be found in two passages from the *Critique of Practical Reason*. The first occurs in a discussion comparing the theoretical and practical employments of reason. After dividing the general concept of nature—which he defines as the existence of things under laws—into the concepts of sensible and supersensible nature, Kant first explicates the notion of supersensible nature through the idea of a practical law, according to which, as we have seen (§V.3), the existence of the things determined by such a law (rational beings) depends on the knowledge of them (and therefore also, of course, on those beings' own knowledge of that law itself). Stating that the supersensible nature of rational beings is their existence according to laws belonging to the autonomy of pure reason, he then relates such beings' supersensible nature to their sensible nature as *natura archetypa* to *natura ectypa*. In short, the moral law “is to give to the world of the senses, as a *sensible nature*, (as far as rational beings are concerned) the form of a world of the understanding, that is, of a *supersensible nature*” (*KpV* 43). Through such efficacy of the moral law, our a priori practical knowledge of ourselves as rational beings—our knowledge of what we are to do—would be reflected in our a posteriori theoretical knowledge of ourselves as sensible beings—our knowledge of what we *do* do.

This order of ideas is also in evidence some pages further on, in the “Typic of Pure Practical Judgment,” where Kant describes the role the

idea of a law of nature plays in practical judgment, and in doing so throws light on why he introduced this idea in his alternate version of the *Groundwork's* first formula. Practical laws, as we have seen, are laws of autonomy; as such, they govern the exercise of freedom, a form of causality that is not sensibly conditioned. Hence, to represent such a law of freedom *in concreto*, practical judgment must employ a law of nature to serve as its "type"; but to play this role, the law of nature must not be represented through a schema of the imagination (or as a law of sensibly conditioned causality), for to depict the law by such means would be to view the actions in accordance with it as heteronomously determined by temporally antecedent sensible conditions (*KpV* 69). The law of nature must rather be regarded in abstraction from the imagination's schema, merely as "the *form of lawfulness* in general," thus enabling it to be conceived as the expression (or ectype) in sensible nature of a law of freedom (*KpV* 70). There are, then, strong textual grounds for supposing that the reference to "a universal law of nature" in the variant rendering of the *Groundwork's* first formula expresses the idea of a practical law that determines rational beings' supersensible nature and is thereby also manifested in experience as a law of their sensible nature.

Lest the mention here of perfectly rational beings and a supersensible, archetypal nature give the impression that a level of idealization has been introduced that carries us too far from the conditions of human practical life and from Kant's own often repeated statement that the human will is not by nature perfectly good, it is worth pointing out that the idea here being proposed can be elaborated in a way that will allow us to keep such conditions explicitly in view. For while Kant denies that the human will is by nature perfect, he also holds, of course, that through moral culture it may be perfected, and in particular that the very consciousness of obligation that provides the original basis for inferring that the human will is not perfect already implies the presence of a capacity (freedom) through which the power of choice can be brought to a condition of virtuous observance of the moral law. Thus, we can understand his mention of a universal law of nature as expressing the idea of a law of our rational nature that is "recovered" as a law of our "second nature," a law that would be reestablished in the achievement of the highest good, where virtue itself is universal through human persons' having cultivated their free powers of choice to such a condition of readiness to act in accordance with practical reason that this "art," or cultivated state

of character, issues in action with a degree of universal regularity converging on that of nature's own uniformity and so "makes possible a system of freedom like a system of nature" (see *MS* 218). Thus, as Kant says later in the *Groundwork*, "morals considers a possible realm of ends as a realm of nature" in order to bring the practical idea of the former into actuality (*G* 436n); this aim explains, of course, why the idea of a law of nature is introduced into the formula of universal law.

But regardless of whether we elaborate the interpretation by including this explicit accommodation of the lack of natural perfection in a human being's will, it is clear that when understood along the lines here suggested, Kant's introduction of the idea of a law of nature is suitable to the moral ideal he articulates in the third formula of the moral law—the ideal of a system of autonomous rational beings united together in a realm of ends. This ideal is not introduced, of course, until a later stage in the argument, but we should nevertheless expect to find it implicit in the formula of universal law, since, as we have noted, Kant says the different formulas are "fundamentally only so many formulas of the very same law" (*G* 436). No such suitability is to be found, on the other hand, in the bare concept of a law knowable only through experience, a law whose efficacy does not lie in its being practically known. For since to be determined according to a law of this sort constitutes the heteronomy of natural necessity, the concept of such a law lacks—indeed precludes—precisely the element of autonomy, or self-legislation, which Kant identifies as the basis of morality (cf. *G* 446–447).

5. *No Means–End Reasoning*

The interpretation outlined above has another distinctive feature, closely related to the one just discussed. Following up the idea that morality delimits a form of goodness and badness that differs from others in that it is intrinsic to the action to which it belongs, we have been attempting to understand the universalization test as serving to make explicit what is already present in a maxim. In particular, we have been pursuing the thought that in morally impermissible action, the impossibility of willing the maxim as a universal law simply reflects a conflict between the maxim's content and its form. According to this way of understanding it, the test involves no prudential deliberation, nor any rea-

soning that connects, under the principle governing hypothetical imperatives, the willing of an end to the willing of the necessary means (*G* 417–419). Means–end reasoning may figure in the determination of maxims, and it must do so where the maxim is a rule of choice as opposed to an exercise of bare wish. But on the present interpretation it has no place in the test itself.

It has often been supposed, of course, that the universalization test does rely on prudential or consequentialist considerations, either in idea or at least in Kant’s actual deployment of it. This supposition has gained currency in part because Kant’s own explanations of how contradictions are involved in cases of impermissible willing can easily appear to rely on reasoning about the relation of means to ends. In discussing the maxim of false promising, for example, Kant suggests that a universal law of false promising would render such action vain and futile (*G* 403). As we shall see, however, it is possible to understand these explanations without supposing them to rely on such reasoning (§VII).

The supposition is bolstered by other factors as well, including certain preconceptions that readers often bring to the text. One of these is the long-standing and widespread tendency to interpret the formula of universal law through the lens of an attenuated conception of practical reason, as first and foremost the capacity to determine the means necessary to achieve given ends, a capacity belonging even to bare practical thought. When the formula is seen through this glass, the application of the universalization test will be understood as a kind of formal operation that modifies the exercise of a more basic deliberative capacity, an operation that first comes into play when, as the saying goes, we adopt the moral point of view on our action. It will then seem natural, as it did to J. S. Mill, to suppose that in attempting to show how familiar duties are derivable from this formula Kant is inevitably constrained, perhaps in spite of himself, to turn his eye to the consequences that would result if a maxim were universally followed. Since we have set such conceptions of practical reason to the side, however, we need not at present occupy ourselves further with the difficulties they engender.

Interpretations that invoke means–end reasoning can also gain favor through the influence of a certain common tendency of practical reflection that can make them seem natural and appealing. Moral philosophy seeks not only to explain the source of moral obligation, but also to contribute,

through its elucidation of morality's principle, to the interest we take in conducting ourselves in accordance with it. In this latter regard, however, philosophy faces a perennial predicament, in that this interest, and philosophy on its behalf, must work against certain empirical conditions that tend to erode it. It is guided from the start (as for instance in the traditional representation of the highest good) by the idea that virtue has an intrinsic worth, through which it also has happiness as its natural effect. Accompanying this idea is the complementary thought that vice contains an inner discord, which tends to subvert its prospect of attaining the ends it sets for itself. But while these causal connections are tightly drawn in practical cognition's original ordering of its concepts, experience offers such an abundance of instances to the contrary that a wish naturally arises for some reassurance that, at least in the appropriate conditions, these linkages are secure. Testament to the strength of this concern are the countless reaffirmations of these connections to be found in the canonical texts of traditional ethical philosophy, not to mention the pages of popular fiction. At every turn, ordinary practical consciousness exhibits its interest in seeing goodness encouraged by its own success and badness shamed by its own failure. To the extent that our thinking is implicitly informed by such a wish, we will be ready, when considering Kant's universalization test, to suppose that its exhibition of a maxim as wrongful should be closely bound up with the circumstance that, under the condition of universality, action on the maxim is futile or ineffective. This tendency is entirely innocent, but to ensure that our attention is not distracted by it when we apply the test, it will be important to keep squarely in view that, while universality may indeed, in a certain sense, annihilate a maxim's efficacy and thereby render action upon it futile or impossible, this can only be the consequence and reflection of the maxim's wrongfulness, not the ground or explanation of it.

Interpretations that take the universalization test to include scrutiny of consequences or of the relation between means and end will be canvassed in due course when we consider the first formula's application (§VII.7). For now it will suffice to observe that since the connecting of means with ends depends on theoretical judgments, reliance upon it generally introduces a contingency that is not easy to reconcile with Kant's idea that moral goodness and badness are intrinsic in character. As we noted earlier (§V.5), prudence requires of the will that it be exercised in accordance with conditions lying outside itself, represented in theoret-

ical judgments of practicability, whereas morality lies in willing in accordance with the formal presupposition of practical judgment, a formal condition internal to the will. Thus if we assume that the universalization test relies upon prudential or consequentialist requirements, then we will have to regard contradictions brought to light through its application as depending in part on conditions lying outside the will and so will not be able to see them as revealing the sort of inner contradiction—between a maxim’s form and its content—in which we have been supposing Kant takes intrinsic wrongness to consist.

6. *Comparison of the Formulas of Universal Law and Humanity*

If we understand the universalization test without assuming that it involves instrumental reasoning or a heteronomous conception of a law of nature, we have a much better prospect of making sense of Kant’s assertion that the formula of universal law and the other two formulas are so many ways of representing the very same law. Our chances are particularly improved in connection with the second formula, which differs much more strikingly in appearance from the formula of universal law than does the third. Since the third is based in the idea of rational beings as autonomous members of a realm of ends, it brings directly into view not only the notion of universal law but also the specific idea of practical law. The second formula, however, does not mention universal law at all; it simply enjoins us to employ humanity in our own person as well as in the person of every other always at the same time as an end.

This striking difference notwithstanding, the second formula can readily be seen to be “fundamentally the same” (*G* 437–438) as the formula of universal law once it too is understood in the light of the form of practical knowledge represented in practical judgment’s presupposition of universality. To see how this presupposition finds expression in the second formula, we need first to consider what Kant means by “humanity” (*Menschheit*). As it figures in this formula, this term expresses not so much the empirically determined concept of human nature as the a priori idea on which it is based (*G* 410n, 425; cf. *MS* 216–217). This idea includes no thought of specific characteristics that would distinguish human nature from the nature of rational beings generally; it is just the idea of “humanity as rational nature” (*G* 439). So “rational nature” (*vernünftige Natur*) is

the expression on which we should focus. Let us start with “nature.” The sense of interest to us is not, of course, what Kant labels the “material” (or “substantive”) sense, the sense we have in view when we speak of the whole of nature. “Rational nature” is clearly intended to signify the nature of rational beings, or what distinctively belongs to it. Our concern is accordingly with what Kant calls the “formal” (or “adjectival”) meaning of “nature,” the sense we intend when we speak of the nature of a thing or kind of thing. When used in this sense, Kant holds, “nature” signifies “the coherence of the determinations of a thing according to an inner principle of causality”—that is, according to a law⁴ (A418–419n/B446n). This is the sense he has in mind when he characterizes “nature in the most general sense” as “the existence of things under laws” (*KpV* 43) and as “the existence of things so far as it is determined according to universal laws” (*G* 421). Given this conception of nature, we may take it that *rational* nature is the existence of *rational* beings so far as it is determined according to an inner law of causality based in their practical reason. For as we have seen, Kant holds rational beings to be distinguished from other things of nature in that their capacity to act according to a law lies in their capacity to act from a *representation* of it, or from a principle, and this capacity he identifies with the will, or practical reason (*G* 412, 427). In other words, and according to the account of willing set forth earlier (§V.3), the inner law of causality determining the existence of rational beings must be a practical law, which, being nothing but practical knowledge, originates in the capacity for such knowledge, practical reason itself.

Now to trace humanity by these steps to practical reason may seem questionable. For in doing so we have nowhere invoked the notion of an end, yet it is by way of a discussion of ends that Kant arrives at the formula of humanity (*G* 427–429). Indeed, when he later reviews the three formulas, he explicitly indicates that, whereas the first formula is concerned with a maxim’s form, or universality, the second is concerned with its matter, that is, with its end. Moreover, he states in several places that the distinctive characteristic of humanity, or rational nature, is its capacity to set an end for itself (*G* 437, *MS* 392).

⁴Here “principle” (*Prinzip*) has a broader meaning than it does in Kant’s claim that “only a rational being has the capacity to act *according to the representation* of laws, that is, according to principles” (*G* 412), where it is equivalent to “fundamental proposition” (*Grundsatz*) (cf. *KpV* 19–20) and signifies a *practical* law (cf. §VI.3).

But if we begin with the thought that humanity lies in this end-setting capacity, we can reach the same conclusion. Kant states that it is an act of freedom to have an end at all, and that an end is an object of the free power of choice (*MS* 384–385). Given that he regards end-setting as having this connection with freedom, **we must take it that the exercise of the capacity to set an end is not an act of bare practical thinking, but a type of willing and hence an exercise of practical reason.** More specifically, and according to the account of willing propounded earlier (§§II–III), the act of setting an end for oneself is just the act of wishing, an act of practical judgment wherein some object is deemed simply good. But as we have seen, such practical judgment includes in its self-consciousness an understanding of itself as an exercise of the capacity for practical knowledge, and this capacity is practical reason.

That it should be possible to reach the same conclusion by these two seemingly different routes is a reflection of the fact that, as the capacity for practical knowledge, practical reason can be considered in two ways. As we have observed (§II.5), Kant holds that the act of making something an end for oneself is nothing other than a practical principle (*MS* 385). Such a principle is related to an end in the way practical judgments are related to their objects: it is a universal practical rule—a representation of a law—that both determines the end and works to bring it, as determined, into existence. **A rational being's capacity to set an end for itself is thus nothing other than its capacity to act from its representation of a law.** Kant accordingly defines the will both as the capacity to act according to principles (*G* 412) and as the faculty of ends (*KpV* 58–59; cf. *KU* 220). And he relies on the equivalence of these two ways of representing the will when he introduces the concept of an end to begin his argument leading up to the formula of humanity: after reminding us of his earlier characterization of the will as the capacity to act from a representation of a law, he goes on without further ado to say that “what serves the will as the objective ground of its self-determination is the end” (*G* 427). As is clear, these two ways of characterizing the will, or practical reason, are mirrored in the two ways of representing humanity that we have just been considering. Just as the will can be defined both as the capacity to act according to principles and as the faculty of ends, so humanity can be characterized both as rational nature and as the capacity to set an end for itself.⁵

⁵ Kant also characterizes rational nature itself in terms of that capacity (*G* 437).

Indeed, we can see in the formula of humanity's injunction to employ *rational nature* always as an *end* a movement from one to the other of these two sides from which practical reason can be regarded. It takes us *from* the representation of practical reason as rational nature, or as the *law* of rational beings represented in the formula of universal law, *to* the representation of practical reason as the spontaneous capacity for self-cognition that recognizes *itself* as the *end* for the sake of which it acts. In thus enjoining us a priori to employ our *nature* (what determines us, constituting us as practically cognizable) as our *end* (what we determine, constituting ourselves as practical cognizers) the second formula serves as the linking moment in the transition from the formula of universal law to the formula of autonomy and its idea of the will's being a law to itself.

Recognizing that humanity lies in practical reason, or the capacity for practical knowledge, accordingly puts us in a position to draw the following comparison between the first two formulas. The formula of universal law has us consider persons in respect of their capacity to *be known* (intelligibility), their capacity to be determined by the practical law in which practical knowledge consists. The formula of humanity has us consider persons in respect of their capacity to *know* (intelligence), their capacity to determine the practical knowledge in which practical law consists. We thus have reason to expect we will find that in the formula of humanity practical knowledge is viewed from the side of its subjective rather than its objective universal validity.

What is it, then, to employ humanity always as an end? In the sense of interest to us here, employing something is a matter of acting from a practical judgment in which such employment is represented. A person's use of a tool or other artifact as a means to some end, for instance, begins in a practical judgment in which such use, in view of its anticipated contribution to the end, is deemed good on the whole. So the second formula's specification of how humanity is to be employed is in the first instance a specification of how it is to be represented in practical judgment. And insofar as this formula articulates a genuine categorical imperative of reason, this specification expresses how humanity is necessarily represented in practical knowledge and hence how it ought to be represented in all practical judgment. So to the extent that the injunction to employ humanity always as an end formulates a categorical imperative, it must be founded in the recognition that in practi-

cal knowledge humanity is necessarily represented as an end. Given then that humanity is just the capacity for such knowledge, understanding the formula of humanity is a matter of understanding why practical cognition necessarily includes in its representation of its object the representation of *its own capacity*, and moreover the representation of this capacity *as an end*. We now take up these two points in turn, drawing on our earlier consideration of practical knowledge and its form.

As we have seen, practical knowledge, as knowledge, includes a consciousness of itself as lying in the exercise of the capacity for such knowledge. But we have also observed that this knowledge, as practical, or self-consciously efficacious with regard to its object, always lies in self-cognition (§IV.6). So the judgment in which practical cognition consists has in its subject position a conception of self, and this conception includes originally and hence necessarily the conception of the capacity for practical knowledge—that is, the idea of humanity, or rational nature. Humanity is thus necessarily represented in the subject position in all practical knowledge, prior to all acts of practical predication through which particular ends are adopted, even the act of making one's own happiness an end, through which particular persons first constitute themselves as such. As Kant remarks, it is “the subject of all possible ends” (*G* 437).⁶

We have also seen that practical cognition's consciousness of itself as cognition includes an awareness of itself as self-sustaining, every act of practical cognition being necessarily in agreement with every other as well as with itself. And this a priori consciousness includes the recognition that the capacity for such cognition, as the ground of the necessary agreement among its acts, must be essentially one. Therefore the idea of humanity, which practical cognition necessarily includes in its representation of the subject it cognizes, is the representation of an essentially unitary capacity, whose practical judgments all stand in a necessary agreement with one another that constitutes the unity of practical knowledge, an agreement that also belongs to the actions issuing from them. Being thus recognized as the

⁶In a brief exposition of Plato's doctrine of ideas in the first *Critique*, Kant says the idea of humanity is something the human being “carries in his soul as the archetype of his actions” (A318/B374). When this indeterminate concept is determined completely and in a fully self-consistent manner, Kant holds, it constitutes humanity in its complete perfection, yielding the representation of the ideal human being (A568/B596; cf. *MS* 386–387, 391–393, *KU* 233).

unitary capacity for self-sustaining practical cognition, humanity is always represented in such cognition as already actual and self-sustaining in and through such cognition itself. As Kant says, it is conceived as “self-standing” (*G* 437). Hence in practical cognition humanity is represented, not as *to be produced*, but as *to be sustained*, both in a negative sense (as not to be hindered) and also positively (as to be furthered) so far as practical cognition, as discursively self-determining, is capable of developing and perfecting itself. Thus, in addition to being necessarily represented in the subject position in practical knowledge, humanity is necessarily represented as sustaining itself in and through its own exercise in self-sustaining practical knowledge.

As we noted earlier (§III.2), the concept of an end falls under that of the good, the object of practical knowledge, and is distinguished from the concept of means in that it is the representation in such knowledge of what furthers *itself*. An end is therefore necessarily in agreement with itself, every element that may belong to it furthering every such element, so that an end always sustains itself, to the extent that external conditions allow. Since in representing its object practical cognition originally and hence necessarily represents humanity, its own capacity, as self-sustaining, it necessarily represents humanity as an end. The formula of humanity’s demand that humanity always be employed as an end thus lies in this necessity in practical cognition’s representation of its own capacity as self-sustaining, a necessity that constitutes recognition of humanity as an “end in itself,” or something whose existence has “absolute worth” (*G* 428).

The second formula, then, as well as the first, can be traced to the form of practical cognition. And as expected, we have found that in the formula of humanity practical knowledge is viewed from the side of its subjective rather than its objective universal validity. In respecting the second formula’s requirement to employ persons’ humanity always as an end, one exercises one’s will from a regard for their capacity for practical knowledge. To consider persons under this aspect is to highlight their capacity to know rather than their capacity to be known.

7. *The Equivalence of the Two Formulas*

When the formula of humanity is understood along the lines just indicated, it is not difficult to see that a will exercised in accordance with its

injunction must also be exercised in accordance with the formula of universal law. The implication can be traced as follows. To employ humanity always as an end is always to represent the capacity for practical knowledge as fundamental to any subject one represents in one's practical judgments. So far as this capacity is always so represented, one's choices will be based in practical judgments in which practical predicates are attached to a subject originally conceived through the idea of the capacity for practical knowledge, and this practical predication will constitute and determine a conception of humanity in one's own person. This determination can accordingly be an act of practical knowledge (as it purports to be) just insofar as it agrees with this original idea of humanity, the capacity for such knowledge, and so just insofar as it has subjective universality, or validity for every subject sharing this capacity. But the *capacity* for practical knowledge—whether in one's own subject or in that of another—can be represented in no other way than through the *form* of such knowledge.⁷ So agreement with the idea of humanity is just agreement with that form. Hence, insofar as the self-determining judgment that grounds choice is in agreement with this original end, it will necessarily have the form of such knowledge. Such a judgment will therefore have a validity that is objectively as well as subjectively universal, holding both of and for all subjects in which this capacity, humanity, is present, and it will accordingly be possible to will such a judgment, such a maxim, as a universal law. So willing that conforms to the formula of humanity must also conform to the formula of universal law.⁸

It is equally apparent that the converse implication holds as well. Where the will is exercised in accordance with the formula of universal law, maxims will be chosen only so far as they have subjective as well as objective universal validity, or a validity that extends in both of these

⁷ Kant makes the same point when he says, "it is just this suitability of [a rational being's] maxims for universal legislation that marks it out as end in itself" (G 438).

⁸ It will be noted that this implication has been traced by focusing on the employment of humanity in one's own person. This does not mean, however, that cases of willing involving one's employment of humanity in the persons of others have been ignored. For any employment (or misemployment) of humanity in the person of another is equally an employment (misemployment) of it in one's own person. As Kant observes, there would be no duties to others (nor any duties at all) were there no duties to oneself, in that one can recognize oneself as bound to another only through binding oneself (MS 417; cf. §VII.9).

respects to every subject sharing the capacity for practical knowledge. But to ensure that one's maxims have such validity is as much as to choose them only on the condition that they be in agreement with this universally shared self-sustaining capacity, and to choose them only on such a condition is nothing other than to employ that capacity as an end. Since this capacity is just humanity, the restriction of one's choice of maxims to the condition that they can be willed as universal laws amounts to the employment of humanity, in one's own person and in that of others, always as an end. Taken together, these two arguments support Kant's claim that the two formulas are equivalent, or "fundamentally the same" (G 437–438).

Despite the equivalence that thus comes into view when the formulas are interpreted in the light of the form of practical knowledge, their marked difference in appearance can of course make it seem that they must express substantively different principles. And in fact, Kant's claim has often been questioned, particularly with regard to the second of the two implications just traced. The formula of humanity is commonly held to be more demanding, yielding duties that cannot be derived from the formula of universal law, at least not so easily and plausibly. It is generally agreed that since this formula identifies a necessary end for the will, it can generate not only negative requirements, but also positive obligations, such as the duty of beneficence. The formula of universal law, on the other hand, has been thought capable of supporting nothing but prohibitions, merely forbidding us from acting on maxims that cannot be willed as universal laws. If the formula of universal law is a mere prohibition, its universalization test can perhaps show that certain maxims are permissible, but never, it seems, that a maxim is obligatory.⁹

⁹ See, for example, Henry E. Allison, *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy* (Cambridge: Cambridge University Press, 1996), 166–168; and Allen Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999), 100–102. A similar view is taken by Christine Korsgaard (though she sees the formulas as also differing in a way that she associates with a distinction between ideal and nonideal theory); Korsgaard does not deny that positive duties, or obligatory ends, can be derived from the formula of universal law, but she notes that Kant reaches them "by a curiously roundabout procedure in which someone is imagined formulating a maxim of rejecting them and then finding it to be impermissible" ("The Right to Lie: Kant on Dealing with Evil," in *Creating the Kingdom of Ends* [Cambridge: Cambridge University Press, 1996], 143–144, 151–154). Wood, however, questions whether such a procedure can succeed, arguing that the imper-

Such reflections have fueled speculation that the formula of humanity goes beyond the formula of universal law, that it is reached through a development or progression in the argument of Section II of the *Groundwork*, as Kant himself seems to suggest at one point when he speaks of a “progression” in the formulas (*G* 436).¹⁰ Yet in view of Kant’s assertion that the two formulas are “fundamentally the same,” we ought first to consider—as we have here in fact been doing—whether such differences as there are between them may lie not so much in their substance as in the way and the extent to which they make a single fundamental principle explicit, enhance its efficacy, or facilitate its application in particular cases. Differences of the latter sort could be enough to account for Kant’s suggestion that there is a development in the argument, and they might also help explain why his readers have generally favored the formula of humanity. We should also be alert to the possibility that some of the differences interpreters have cited may be merely apparent (§VII.6).

To remove the doubts just noted, however, we need to consider more closely the second of the two implications traced above, with an eye to the question whether the formula of universal law can reveal positive obligations.¹¹ Our prospects for making sense of Kant’s claim that the two formulas are at bottom the same will be greatly improved if we bear in mind that the action of a person always presupposes a practical

missibility of refusing to help others does not by itself entail the obligation to help them, not even the obligation merely to help some others sometimes, for the maxim of refusing to help and that of helping some others sometimes are related as contraries, not as contradictions. Kant’s “roundabout procedure” will be considered more closely below (§VII.4–6). Not all recent interpreters have questioned the equivalence of the formulas; for an account similar to the one offered here, see “Universal Laws and Ends-In-Themselves,” in Onora O’Neill, *Constructions of Reason: Explorations in Kant’s Practical Philosophy* (Cambridge: Cambridge University Press, 1989).

¹⁰In his very brief sketch of the progression, Kant relates the formulas to his categories of quantity in their reflective use as formal or logical criteria of the possibility of cognition (cf. B113–116). The first formula reflects practical cognition’s *unity* of form; the second the *plurality* of its material, or the ends (rational beings) which (and by which) it represents; and the third their *totality* in a system to which they belong through their all being originally practically cognized by each in relation to that same unity of form. For some recent discussion of the progression, see Barbara Herman, “Leaving Deontology Behind,” *The Practice of Moral Judgment* (Cambridge, Mass.: Harvard University Press, 1993), and Allen Wood, *Kant’s Ethical Thought*, chaps. 3–5 (esp. 182–190).

¹¹That this formula does support positive duties will receive further confirmation later, when we examine Kant’s explanation of how it grounds the duty of beneficence (§VII.4–6).

judgment in which an end is determined, and that such a judgment is itself a maxim. If we then draw on the point made above that humanity is a necessary end presupposed in all determination of ends in practical judgments, it will be possible to show directly—that is, without considering any maxim contrary to duty—that exercise of the will conforming to the formula of universal law entails full conformity with the formula of humanity. This can be done by exhibiting this entailment in the case of the fundamental practical judgment in which a particular person first constitutes itself as a person.

Since the path of reasoning just outlined is essentially the one that Kant himself follows, we may start by considering the brief argument by which he arrives at the formula of humanity (*G* 429). He begins with the observation that, as a human being, one necessarily represents one's existence as an end in itself, on the basis of one's recognition of oneself as a rational being. He goes on to note (or to put forward as a "postulate") that every other rational being represents its own existence in the same way "in consequence of the very same rational ground" that is also valid for oneself. He then concludes that the proposition that "rational nature exists as end in itself" is an objective principle and as such the basis for the formula of humanity's requirement to employ humanity in one's own person and in the person of every other as an end in itself.

Now the premises of this argument will likely seem suspect, and the conclusion will almost certainly seem to be an egregious *non sequitur*—reminiscent of the one we have seen Kant charged with in his derivation of the formula of universal law (§I.1)—if we are disposed to assume that here in his use of the phrase "the very same rational ground" we find the familiar equivocation. If we assume that the "same rational ground" in consequence of which every rational being represents its own existence as an end in itself can be nothing more than some such reflection as "*this* existence is *my own* existence" (where this existence is tacitly thought as distinct from that of every other rational being), then of course no stronger conclusion will follow than that every rational being represents *its own* existence as an end in itself. The conclusion Kant draws will follow only if "same rational ground" is understood more strictly, as indicating that he is thinking of a principle of reason, something that can serve as a ground for a judgment in which we "cognize the particular in the universal through concepts." It is highly implausible, however, that Kant's argument is afflicted by this equivocation;

indeed, elsewhere Kant himself explicitly criticizes the sort of confusion such equivocation involves (*KpV* 28).

A more satisfactory understanding of Kant's thinking here is available if we view it in the light of the conclusion we reached earlier, that in all practical judgment it is presupposed that humanity, or our rational nature, is an end in itself. If this is correct, then we can take the representations of which Kant is speaking (that is, each rational being's representation of its own existence as an end in itself) to be in fact nothing other than representations of the very same humanity, or rational nature, even though each of them is a representation of that same humanity in a different individual subject. Indeed, each such representation is a practical judgment in which the same original idea of humanity that they all presuppose is determined in the representation of humanity *in one's own person* as an end in itself. Each person's representation, in other words, begins from "the very same rational ground"—namely the idea of humanity as an end in itself—and determines it for the case of that individual person, thus reaching the representation, or practical judgment, that humanity in one's own person is an end in itself. Each person, Kant says, *necessarily* represents itself through such a judgment, and we can take this necessity to stem from this representation's being just the practical self-conception that constitutes a person.¹² It is the fundamental practical judgment through which an individual person originally constitutes itself as an individual person. But every such practical judgment, as such, presupposes its own objective and subjective universal validity. So in order to be in agreement with its own form—the form we have seen to be articulated by the formula of universal law—it must be consistent with the implicit universal practical judgment in which it is contained, and in which humanity as such, and hence humanity in every person, is deemed an end in itself. But to exercise one's capacity for practical knowledge in a way that is consistent with this necessary implicit judgment that humanity in the person of every other as well as in one's own is an end in itself is just what the formula of humanity requires. Thus, to exercise the will in accordance with the formula of universal law is to exercise it in accordance with the formula of humanity. The first "says just as much as" the second (*G* 438).

¹²If restricted to *persons*, as it is here, this claim expresses a conceptual truth; Kant's assertion extends to rational beings generally, and presumably for this reason is put forward only as a postulate (see §III.7).

It seems, therefore, that the difference between these formulas, while striking enough to create the appearance that they express substantively different principles, is in fact a difference in aspect. As they have been interpreted here, a full understanding of them reveals each to be equivalent to the other. To understand the point of the formula of universal law, and therewith its applicability in practical judgments generally and hence even in fundamental practical judgments, is to recognize that humanity, as the capacity for such judgments, is in each person an end in itself. And conversely (as we saw earlier), to respect the true intrinsic worth of humanity at all is to conform one's maxim to the form by reference to which humanity (as a capacity) must be represented, namely the idea of universal law.

8. *A Concern about Method Removed*

Despite the support they provide for Kant's claim that the first two formulas are equivalent, the foregoing arguments may seem to conflict at a deeper level with the basic order of concepts structuring his practical philosophy. If, as we have argued, the first formula says just as much as the second, then it must be no less capable than the latter of yielding positive obligations, duties to adopt certain ends; given that the second formula supports such duties through its identification of humanity as something of positive worth, it seems to follow that some recognition of this value must also be involved already in the first. Yet this implication may appear to clash with the idea, central to Kant's practical philosophy, that the principle of morality, as a principle of the will's autonomy, precedes any substantive conception of the good. This doctrine, sometimes associated with John Rawls's thesis that "the concept of right is prior to that of the good,"¹³ is widely regarded as decisively distinguishing Kant's approach from all teleological and consequentialist accounts of morality's principle. Kant emphasizes this feature of his view when in the *Critique of Practical Reason* he calls attention to the work's "paradox of method," namely that "the concept of good and bad must be determined,

¹³ *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), 31; see also his "Themes in Kant's Moral Philosophy," in *Kant's Transcendental Deductions: The Three Critiques and the "Opus postumum"*, ed. Eckart Förster (Stanford, Calif.: Stanford University Press, 1989), where he discusses the related idea that for Kant the rational is subordinate to the reasonable.

not prior to the moral law (for which it seemingly must even serve as the basis), but only (as was done here) after it and through it" (*KpV* 62–63). Given this ordering of concepts, and given the second formula's identification of humanity as an end in itself, or as something whose existence has "absolute worth," it might seem reasonable to conclude that Kant offers the first formula as the more fundamental expression of the moral law, one free of any representation of value, and that he regards the second as reachable "only . . . after it and through it." An interpretation that entails their equivalence might thus fall under suspicion of being at odds with Kant's method.

This concern, however, can easily be laid to rest. Kant's method, however paradoxical it may seem, is without question a sound guide for the interpretation of his ethics. Yet in the statement of it quoted above, Kant does not claim that the moral law is prior to the *concept* of the good. He only says that this concept must not be *determined* prior to the moral law but only after it and through it. He is not asserting, in other words, that the original, *indeterminate* concept of the good—the concept to be determined in practical judgment—is posterior to the moral law.¹⁴ His method merely requires, in accordance with the "fact" of reason, or our consciousness of the unconditional moral law, that the determination of the concept of the good be wholly in conformity with that law of the

¹⁴ On the contrary, in keeping with another methodological requirement—that definitions are to be framed in such a way as to leave disputed points undecided (*KpV* 9n)—Kant refrains from defining the concept of the good in terms of the moral law. He takes it to be uncontroversial that this concept is a rational concept of an object of desire and hence that the good is an object of some sort of knowledge, but he does not build it into his definition that the good is an object of practical knowledge. Thus, rather than telling us that the good and the bad are the objects of *pure* practical reason, he merely identifies them as the objects of practical reason in general: the good, he says, is "a necessary object of the faculty of desire . . . according to a principle of reason" (*KpV* 58). Taken by itself, this definition does not specify whether the principle it mentions is objective and as such a practical law, or whether it is merely subjective, that is, a maxim that a particular person follows because doing so has agreeable effects. The definition thus leaves undecided how the good is to be determined, whether by considering what the possible causes are of agreeable states of mind in the particular subject making the determination, or whether by considering what the possible effects are of the exercise of the will in accordance with a law of pure practical reason. In short, the definition leaves undecided whether there is any such thing as a practical law at all. It thus does not rule out the possibility that goodness is exclusively relational, and in particular that for a thing to be good is for it to produce or contribute to some further object that need not itself be good. In such a case there would be no practical knowledge; "good" would signify merely an object of practically deployed theoretical knowledge.

will's autonomy.¹⁵ He is thus opposing the method of heteronomous theories, which first seek to determine, independently of the moral law, what the (achievable) good consists in (happiness, say), and then attempt to explain or to vindicate that law by showing that following it will make the realization of this good possible.

That Kant should oppose this heteronomous method is precisely what we should expect, given the approach we have been following. For it has been a guiding thread of our interpretation that he takes morality to lie in the practical knowledge of *intrinsic* goodness, and that he accordingly regards an action to be genuinely moral (not just legal) only so far as it springs from such knowledge. Heteronomous views, in contrast, are founded on the assumption that morality lies in an action's contribution to some further good, which implies that the goodness an action has insofar as it is moral is never intrinsic, but always merely relative.

It should also be clear that in tracing the universalization test to the formal presupposition of practical judgment we have not infringed the method required by the will's autonomy. For this presupposition, as we have observed on more than one occasion, was identified without any consideration of what specifically is good, and so does not depend on any actual judgments about what it would be good to do or to pursue that any particular person or persons might make. It is because we adhered to this method that we were able to view the presupposition brought to light as belonging to the form of willing and to account for this form's normative authority and standing as a fundamental law of the will, a law of autonomy.

Nor, finally, do the foregoing arguments for the equivalence of the two formulas constitute any departure from the method. As we noted in

¹⁵This requirement has a noteworthy consequence. If, given that there is such a thing as a practical law, the determination of the good must be in accordance with that law, then the fact that there is a practical law rules out what seemed to be a genuine possibility in abstraction from consideration of the reality of that law, namely that the good can be determined by considering what the possible causes are of agreeable states of mind in the subject making the determination. Thus, given that there is a practical law—given, that is, that there is something unconditionally good, or of absolute worth—there can be no valid practical judgment that deems an action good merely on the basis of the action's agreeable effects. An antecedent consideration of the action's relation to what is unconditionally good (humanity) is always required.

tracing the first of the two reciprocal implications, it is not even possible to understand what it is for willing to agree with the idea of humanity except through an understanding of what it is for it to agree with the form of practical knowledge, for it is solely through this form that the capacity for such cognition is represented, and humanity is just that capacity. Only “after” and “through” that form, then, is it possible to reach the second formula’s idea of humanity as something of intrinsic, absolute worth.

Further confirmation that the present interpretation harmonizes with Kant’s approach can be gleaned from a comment he offers later in the *Groundwork* when discussing heteronomous theories. After classifying such views as empirical or rational according as they ground morality’s principle in happiness or in perfection, he remarks that if he had to choose between them, he would favor the concept of perfection for the reason that, although incapable of deciding the question concerning morality’s principle, it at least places the question in the court of pure reason and thereby preserves unfalsified for a “closer determination” the indeterminate idea of “a will good in itself” (*G* 443). Thus, while he rejects any approach that would trace morality’s principle to the concept of perfection, Kant by no means holds that any attempt to articulate this principle by starting with a concept of something good, or of something good in itself, must be heteronomous. On the contrary, the idea of “a will good in itself” is the very idea with which the *Groundwork*’s own investigation into the principle of morality begins.

It is illuminating in this connection to recall how that investigation unfolds. The first two paragraphs of Section I provide a systematic classification of the various elements that we might judge to belong to the highest good, either as constituent ingredients or as the factors in a human person’s makeup and circumstances that would contribute to it. These elements include, along with happiness and the other “gifts of fortune” on which it partly depends, also those factors that belong to a human being’s perfection. But in discussing the latter Kant deploys a crucial distinction between what he elsewhere calls “moral perfection” and “physical perfection” (*MS* 391–392; cf. 386–387). Moral perfection lies in the will and has a good will as its sole foundation, whereas physical perfection lies in “capacity” broadly understood, comprising all other goods belonging to a human being’s constitution—both “gifts of nature”

(including talents of mind and qualities of temperament) and also the acquired qualities of character and other skills that facilitate the pursuit of one's ends. Kant's primary purpose in surveying these various elements is of course to point out that only the good will is unconditionally good, that the goodness of all the other items is contingent insofar as they may be bad or harmful if the will that makes use of them is not itself good. But for our purposes here the following points are of particular interest. First, in this opening argument of the *Groundwork* Kant introduces the idea of intrinsic goodness and separates such goodness, as something found only in the will and its exercise, from the goodness of everything else, including the goodness of happiness and that of perfection in everything that lies outside of the will. Second, in making this separation Kant relies on the idea that the will is practical reason and further that the intrinsic goodness of the good will is based in the capacity for practical knowledge, the original understanding of the good; for without such understanding the good will could never be what Kant holds it to be, namely the sole factor capable of ensuring the good use of all those other things that are distinct from it and whose goodness depends on good use.

To find the principle of such knowledge, Kant then explicates the idea of this intrinsic, unconditioned goodness unique to the will through an analysis of the concept of a good will as it is found in ordinary moral consciousness in the more determinate concept of duty. The principle he eventually reaches turns out to be nothing other than the first of the three formulas of the categorical imperative that are later presented in Section II by way of the philosophical account of the will as practical reason. It is evident from Kant's own procedure in the *Groundwork*, then, that although he rejects heteronomous attempts to determine the good prior to and independently of the moral law, he holds that the idea of the *intrinsic* goodness of the good will is coeval with that law. An analysis of the concept of an unconditionally good will reveals its principle, and conversely to comprehend this principle is to recognize the unconditioned goodness of a will in thoroughgoing conformity with it.

This relation is clearly reflected in the interpretation developed above. For the original idea of this good will cannot be anything but the original subject of practical judgment, that is, practical reason itself, which, as we have seen in our discussion of the formula of humanity, is identical with

humanity.¹⁶ Moreover, in our attempt to bring the principle determining this good will's exercise to light, we have not followed the heteronomous method that Kant criticizes, but have sought to identify the formal presupposition of the practical judgments in which the self-determination of that presupposed concept of a good will consists, the presupposition with which those judgments must conform to qualify as the practical knowledge of the good that they purport to be.

When Kant says that the concept of the good must be determined after and through the moral law, then, his point is that it must be determined in accordance with the presupposition of the act of practical judgment in which this determination always consists, the presupposition that, as we have seen, binds it to the form of willing, or to the fundamental law of the will. And since this law and the practical judgments in accordance with it constitute practical knowledge, he is thereby expressing, in its bearing on the case of practical knowledge, his general "Copernican" proposition that objects must conform to our cognition. The moral law is the first principle of all our practical cognition, or knowledge of the good, and since the objects of this knowledge must conform to our knowledge of them, they are constituted according to the practical judgments in which we determine the good in accordance with that first principle. Far from clashing with the basic approach of Kant's practical philosophy, then, the attempt here undertaken to trace the formula of universal law to practical reason by investigating the formal presupposition of practical judgment is an explicit elaboration of the "Copernican" way of thinking expressed in Kant's statement of the "paradox of method."

¹⁶ If, as is here being suggested, the good will in its original representation is just humanity itself, then of course a distinction must be drawn between the good will as potentiality and the good will as actualized in morally worthy action and ultimately as perfected (strengthened) in virtue. It is the good will in the former sense that is present insofar as there is consciousness of duty at all (cf. *G* 412–413); it is the good will in the latter sense that Kant has in mind when he suggests that the true "vocation" of reason as a practical power is to produce a good will (*G* 396). Kant relates the two at one point, when he describes rational nature—that is, humanity, "the subject of all possible ends"—as "the subject of a possible absolutely good will" (*G* 437). This distinction only applies, of course, in the case of a will that is not by its own nature perfectly good (cf. *G* 412–414).

VII



Applications

1. Preliminaries

In the preceding sections we have developed an account of the categorical imperative that traces it to the form of practical knowledge. By interpreting this imperative as the expression of the first principle of such cognition, we have explained how it is related to reason. We have now to consider this principle's relation to familiar substantive requirements of morality. Since this principle lies in the *form* of practical knowledge, its relation to these requirements must be ascertained through considering its use, its application in conduct.

We noted earlier that the *direct* application of the will's fundamental formal practical law in the basic act of the free power of choice by which particular persons first constitute themselves as such yields a system of a priori practical laws, including laws of natural justice and beneficence, which persons legislate to themselves in respect of their relations to one another (§V.3). But as we also observed, Kant holds that, because the power of choice in a human being can be affected by sensible desires, human persons are conscious of practical laws as *constraints*, under the headings of obligation and duty. Thus, in illustrating how the fundamental practical law, in the guise of a categorical imperative, is the source of duties, he takes up maxims *contrary* to this law and shows how conflicts in the will arise when we attempt to square them with this imperative in its several formulas. Four examples are offered, arranged under two intersecting distinctions among duties, one dividing them into perfect and imperfect, the other into duties to oneself and duties to

others. Since the examples he presents have occasioned numerous difficulties of interpretation, particularly in connection with the formula of universal law, we shall take up a few such cases, focusing on this formula, and consider how contradictions result when we try to attribute the double universal validity of practical knowledge to maxims contrary to duty.

Of course, in view of the many interpreters and critics who have concluded that the formula of universal law yields unsatisfactory results when applied, it might seem advisable to concentrate instead on the more popular formula of humanity. Even if, as was argued above, the formulas are equivalent, it does not follow that they are equally serviceable for specifying obligations and duties. The very fact that the second formula is so commonly favored might itself seem to indicate that it is better suited for this purpose.

On closer scrutiny, however, the second formula's relative popularity points toward a rather different conclusion. Its appeal is too immediately palpable to spring from any detailed consideration of questions relating to its application. The exalted idea of humanity as an end in itself has a certain grandeur and easily captures our attention. The widespread appeal seems accordingly to stem mainly from the fact that the second formula, owing to the aspect of practical knowledge that it highlights, presents persons in a more attractive light than does the first. It is more pleasing to conceive of ourselves as cognizers than as cognizable, as beings who determine laws than as beings who are determined by them, even though in practical knowledge our very right and title to the former self-conception is identical with our subjection to the latter (cf. *G* 440). This difference, however, provides no reason for thinking the second formula to be best suited for the purpose of discriminating cases under the law. What it rather shows is that this formula has a distinct and complementary function, namely the motivational role of helping the law gain "acceptance," or efficacy in determining choice (cf. *G* 437). While this function is essential to the self-development by which a person's practical cognition can perfect itself and thereby approach the practical wisdom basic to virtue, it nevertheless constitutes a second moment in the imperative's application, one that presupposes the cognitive, discriminative moment with which we are here concerned.

In fact, there is no escaping the conclusion that, in spite of the difficulties, the formula of universal law is preeminently suited for this

principal role in application, nor does the formula of humanity afford a genuine alternative to it in this regard. As we saw when tracing the relations of mutual entailment between the two formulas, there is no way of determining whether the employment of the capacity for practical knowledge agrees with that capacity, as the formula of humanity requires, except through determining, in accordance with the formula of universal law, whether it agrees with its own form, since it is through this form alone that the capacity can be represented. In order, for instance, to consider whether others can “agree” with a certain way of treating them (*G* 429–430), we must conceive of them as each exercising their capacity to judge in accordance with that form. For this reason, presumably, Kant assigns to the formula of universal law a special position in the exercise of moral judgment (*G* 436–437) and refers to it as “the canon for the moral judgment of action in general” (*G* 424; cf. 403).

The cases to be considered will be selected with the following points in mind. First, and most importantly, the application will be best illustrated if we take up examples involving duties that are the simplest and most primitive and hence broadest in scope. It would not be appropriate, for example, to begin with cases involving duties that presuppose particular relations between the persons involved, such as the ties between citizens, family members, or friends. Nor would it be proper to begin with an instance of action undertaken in response to another’s presumed or anticipated wrongdoing, such as Kant’s much discussed example of the murderer at the door. Examples of either of these sorts depend upon specific, contingent conditions of action, but the cases we should consider first are those involving duties that attach to us most fundamentally, merely in virtue of our standing as human persons, or subjects with wills, sharing the power of practical reason. This principle of selection simply reflects the order of practical knowledge, which proceeds from the universal to the particular, and hence from instances that are simplest and best understood to more involved and difficult cases, which call increasingly on judgment and individual discretion, as ethics by degrees shades into casuistry (cf. *MS* 411). Kant too proceeds in this fashion. Even in the *Metaphysics of Morals*, which, being concerned with the application of the moral law to the human being, takes into account empirical knowledge of human nature (*MS* 216–217), Kant confines the exposition of the doctrine of virtue to its “metaphysical first principles” and hence to duties human beings have to themselves and to

one another considered merely as human beings (*MS* 375–376, 468–469). In the more fundamental investigation undertaken in Section II of the *Groundwork*, abstraction is required even from the empirically determined concept of human nature (*G* 406–412). In keeping with that requirement, we shall restrict our attention principally to cases involving duties that presuppose only the general relations persons bear to one another so far as they are considered merely as persons. Such cases can be considered entirely a priori and indeed must be, in that they can depend on no more than the fundamental exercise of the free power of choice in which particular persons first constitute themselves as such.

Second, although Kant chooses his examples with the aim of showing that the formula of universal law supports duties to oneself as well as duties to others, it is in the derivation of the latter that the contribution made by the distinctive self-related universality belonging to the form of practical knowledge is clearly in evidence. For this reason, we shall concentrate on duties of this type. In part for same reason, however, cases involving duties to oneself do raise some questions. What is most noteworthy about these duties for our purposes is that the idea of universality seems not to play any significant role at all in their derivation, at least as Kant presents it. It is thus not immediately clear how they can be accommodated by the foregoing account of the formula of universal law. We shall therefore briefly consider, at the conclusion of our discussion, how they differ from duties to others and the special way in which they are related to the form of practical knowledge.

Third, we need to select examples that illustrate the formula's relation to each of the two main types of duties to others that Kant distinguishes—perfect duties, which prohibit certain maxims of action, and imperfect duties, which prescribe certain ends. These fall, respectively, under the laws of justice and beneficence. Surveying examples of both types will enable us to consider how this division among duties is related to the formula of universal law, and in particular to examine Kant's observation that the distinction between these two types of duty corresponds to a difference in the ways in which contradictions result when we attempt to will as universal laws maxims that violate or fail to conform to them.

Finally, our task will be facilitated if we are able to compare our discussion with Kant's own illustrations of the formula's application, so it will be desirable to examine where feasible cases that he himself considers.

Doing so will also provide an occasion for addressing some of the difficulties interpreters have encountered in their treatment of his examples.

2. *Perfect Duties to Others: The Primitive Duty of Natural Justice*

To illustrate how the formula of universal law applies in connection with perfect duties to others, Kant selects a case of false promising, involving a maxim of securing loans, when one needs money, by promising to repay even when one knows repayment will not be possible. This example seems, in fact, to figure in an especially prominent way in Kant's treatment of the universalization test in the *Groundwork*. In addition to being discussed in connection with the formula of universal law (at *G* 422), it is the only example offered to illustrate the universalization test when it is first introduced in Section I (at *G* 402–403). It also seems, on first inspection at least, to be the clearest and most compelling of the four he presents. Moreover, a somewhat similar example involving a deposit is prominently positioned in the *Critique of Practical Reason* (*KpV* 27–28). Such considerations may tempt us to think that Kant developed the universalization test by reflecting chiefly on cases of wrongdoing involving promises and contracts. And this thought may lead us to surmise that in examining such cases, in which a general practice, such as that of promising, seems to be exploited, he first noted that the exploited practice depends on broad compliance and then hit on the idea that attempting to conceive of the wrongful conduct as universally engaged in serves to bring the action's illicit character into sharp relief, in that the resulting contradiction dramatically reveals its dependence on the very practice it abuses and thereby erodes. By such a train of reflection we thus arrive at a familiar interpretation of the universalization test, which sees it as inspired by reflection on the relations holding between social practices and the particular actions that depend on them.¹

Difficulties emerge, however, as soon as we broaden our view. To the extent that it depends on the socially instituted modes of interaction figuring in promises and contracts, false promising is not the most primitive case of action violating a perfect duty to others. Persons are liable to

¹ See, for example, Robert Paul Wolff, *The Autonomy of Reason* (New York: Harper & Row, 1973), 165–169.

mistreatment in a great many ways that do not involve such institutions. Borrowing a phrase from Kant, we can locate these other types under the heading “assaults on the freedom and property of others” (G 430). Though in many instances a false promise too might in a sense be describable as such an assault, the more primitive cases falling under this description differ from false promising in that they do not directly exploit any practice or convention, but achieve their aims in other ways, by physical force, for example, or by stealth.² As recent scholarly discussion has made clear, such cases pose difficulties for accounts of the universalization test that are rooted in reflection on conditions specific to instances of wrongdoing that involve the abuse of a general practice. If the test is interpreted along such lines, its application will seem particularly problematic in cases where persons merely violate the law of natural justice, assaulting the freedom of others by exploiting some advantage afforded by nature, fortune, or their own industry.

Even apart from such difficulties, we have good reason to doubt that Kant’s thinking will be adequately captured by such an interpretation or indeed by any other that is inspired by reflection on a specific type of case. Kant announces that he is moving beyond “a popular philosophy, which goes no farther than it can get by groping about by means of examples” (G 412). So interpretations constructed from particular instances, even instances that Kant himself discusses, are liable to prove unsatisfactory. To avoid the impasse just noted, we need only proceed in the manner proposed above, following the order of practical knowledge in applying the formula of universal law. Our prospects of understanding how the test is to be applied will then not be derailed by interpretive presumptions derived from reflection on conditions that are specific to cases involving promises and contracts. We may even find that a clearer understanding such cases will follow as well.

Let us consider, then, the primitive duty of natural justice. This obligation, as we have seen, is founded on the practical knowledge of

²Cf. Korsgaard’s discussion of the distinction between natural and conventional actions (“Kant’s Formula of Universal Law,” in *Creating the Kingdom of Ends* [Cambridge: Cambridge University Press, 1996], 84–85, 97–101). An assault on another’s property is of course possible only where the institution of property has been established, but if the assault is by sheer physical force, it will not exploit any practice figuring constitutively in this institution, at least not in the direct way that false promising seems to exploit practices constitutive of promising.

self-sufficiency as an end. Now according to its very idea, a person's self-sufficiency can never be augmented by the action of another, but only restricted by it. Another's action may of course contribute to a person's self-sufficiency in an indirect and negative way, by blocking some limitation or hindrance, such as the action of a third that would otherwise infringe it. But far from increasing self-sufficiency, such a contribution merely constitutes a favorable external condition to which it is subject, whereas everything integral to self-sufficiency must be the person's own achievement. The maxim we have to consider, therefore, is one prescribing action that restricts or infringes others' self-sufficiency. Before applying the test, however, we should specify this maxim more completely, both in respect of its action and with regard to its end.

A suitably definite representation of the action can be reached by relating it to the specific respect in which self-sufficiency is liable to infringement by others' action. This can be done by characterizing the action as the limitation of what Kant calls outer freedom. For as the following considerations indicate, outer freedom is just what self-sufficiency requires, as a negative condition, in relation to others.

outer freedom

Freedom in the sense of interest here is not merely the general practical freedom presupposed by morality and practical knowledge, the freedom that belongs constitutively to the free power of choice and lies in the bare capacity to choose in accordance with such cognition's form (cf. *MS* 213–214, 226). Rather, it is this presupposed self-determining capacity in a self-developed condition. Freedom in this sense is an end, something to be achieved, and as such it is a necessary object of wish. Conceived negatively, it is the independence of the free power of choice from influence on its exercise.

As its name indicates, outer freedom contrasts with inner freedom. Freedom of the power of choice, when conceived in the negative manner just noted, as independence from influence, can be divided into two spheres, inner and outer, in accordance with a corresponding difference among sources from which influences may arise. **Inner freedom lies in the independence of this power from influences arising within oneself, namely inclinations.** Outer freedom consists in its independence from influences stemming from other persons—or, in Kant's words, an "independence from the necessitating power of choice of another" (*MS* 237).

inner freedom

Now because the sort of necessitation that outer freedom precludes has an external source, it can influence a person's free power of choice

only in its outer exercise, or in the choice of action, not in its inner exercise, the setting of an end in an act of wish. For such necessitation can occur only through a person's coming to judge that its capacity to realize its end is negated or limited by another's action. The judgment may be based on the experience of physical constraint, the immediate blockage or hindrance of one action by another, or it may be arrived at a priori (in a comparative sense), as happens where the necessitation takes the form of coercion, yet in either case the necessitation is mediated by a judgment of practicability. But it is precisely on account of its dependence on such judgments that choice differs from bare wish. Thus, the inner exercise of a person's power of choice, or its exercise in the setting of ends, is not subject to such necessitation at all. In respect of such exercise, therefore, there is no sense in which outer freedom is subject to limitation and hence no sense in which it is an end to be secured. In this relation it can be conceived only as an outward reflection of the inner freedom integral to the general practical freedom presupposed by morality—as, say, the capacity to refuse to obey any unjust rule or command laid down by another person, for instance a prince's order, backed up by a threat of death, to bear false witness against an honest person (*KpV* 30).

So it is only in respect of one's choice of the actions to be undertaken in the pursuit of one's ends that outer freedom can be limited. In this respect, outer freedom lies in the independence of one's capacity to pursue one's ends from hindrance to its exercise stemming from the power of choice of another. Now it is clear that this capacity can be subject to such infringement, even in instances where the other's action exploits no social practice or institution. The possibility of this hindrance does, however, depend on certain more general conditions. Most fundamentally, and obviously, persons whose actions are subject to such interference from one another must coexist, or be present together in a single world. This coexistence, moreover, must be conceived in the first instance as stable and hence as ongoing. For persons' practical cognition, as cognition, lies in activity, and as practical, it is essentially self-related, so the ends represented in such cognition and hence also the pursuit of them must be activities as well. But while this condition of coexistence is necessary, it is plainly not sufficient. It is possible, of course, even where persons are conceived merely as so related, to comprehend how one person's action might prevent or hinder another's. One person might divert a river, for instance, thereby preventing others downstream, of whose existence the first is

perhaps entirely ignorant, from using its water. Any such hindrance, however, can only be incidental if the persons cannot also stand in relations of mutual recognition. Our concern here is with choice and its maxim, and we are accordingly interested in intentional rather than incidental hindrance to choice: hindrance that is already represented, in the alien power of choice from which it springs, as a hindrance to choice. To comprehend the possibility of such hindrance, we must conceive of persons coexisting as members of what we may call a *practical world*—that is to say, we must conceive of them present together in the world in such a way that mutual recognition and mutual influence are possible. Mutual recognition—persons' recognition of one another as persons—is possible where persons can in principle know what actions are intended by one another. Mutual influence is possible in a negative sense where persons can each act in ways that are opposed to actions they recognize the others intend, preventing or hindering them; mutual influence is possible in a positive sense where persons can each act in ways that are in agreement with actions they recognize the others intend, enabling or facilitating them.³ Where diverse persons coexist as members of a practical world, the outer freedom of one is limited to the extent that another chooses to act in a way it recognizes will prevent or hinder the former's action and succeeds, at least to an extent, in the attempt. (One such person's action may also, of course, prevent or hinder another's merely incidentally, as in cases of ignorance or defective execution; but as noted, hindrance of this sort lies outside our present concern.) Where a person's actions constitute such impediments to others' actions they can accordingly be described—to borrow Kant's phrase again—as “assaults on the freedom . . . of others” (*G* 430).⁴

Now since the material ends a person pursues in acting are all united in the fundamental end of happiness, generically conceived, outer freedom amounts to independence from hindrances by others to one's pursuit of that basic end. Thus any assault on this freedom, to the extent that it is successful, limits a person's capacity to realize this end. And

³Of these two types of influence, it is really only the possibility of *negative* influence that must be assumed to understand the possibility of hindrance to outer freedom and with it the possibility of the law of natural justice; the possibility of *positive* influence is a condition that will come into play when we consider the law of beneficence (§VII.4).

⁴In this passage, Kant mentions assaults on property as well as on freedom. But since property is a specific, socially instituted form of freedom, we may omit mention of it here to focus on the primitive case.

since this capacity is just what self-sufficiency consists in, outer freedom is nothing other than the independence from other persons' powers of choice requisite for self-sufficiency, and it can therefore be regarded, in a negative sense, as self-sufficiency itself in relation to others.

To complete our specification of the maxim, we need to articulate the wish from which it would spring, the end that its action of limiting others' outer freedom would serve. Since the duty to which the maxim is opposed is founded on the practical knowledge of self-sufficiency as an end, the maxim would originate in a particular person's misuse of the will regarding this object; and since persons first make self-sufficiency an object of judgment in the act of wish in which they include it as a necessary component in their end of personal happiness, the misuse will have to lie in the way in which this object is practically represented in that wish. The illicit wishing includes two moments or stages, a lapse followed by a transgression. The primitive wish consists in a practical judgment that ascribes simple goodness to one's self-sufficiency, yet does so without being seated in a judgment ascribing such goodness to the self-sufficiency of persons generally. The latter judgment is lacking, of course, since *ex hypothesi* the wish's content is not determined in accordance with the form of practical cognition, even though this form is presupposed in the very act of wishing itself. Now because this wish is not situated in a materially universal practical judgment, it is not flanked in the subject's practical thinking—not even implicitly—by any practical judgments pertaining to others' self-sufficiency that could limit the further practical judgments it may yield. Hence, once other persons with whom one coexists are brought into consideration in practical reflection, this wish will be able (subjectively) to support the further wish that the self-sufficiency and outer freedom of others be limited so far as necessary to further such sufficiency in one's own case. This latter wish clearly violates the form of practical knowledge, yet because it is still an exercise of the practical cognitive capacity, it not only presupposes its own conformity with such cognition's universality but can even spawn a representation of itself as universal, by prompting the subject to project it onto other persons.⁵ It may accordingly, in response to the reflection of

⁵ Because this wish violates practical cognition's form, its projection onto others effects the appearance of an inversion of the true relation between the form and the act of cognition, in that here the representation of universality is framed on the basis of the act, rather than the act on the basis of the representation of universality. In fact, of course, the universality

itself that it thereby finds in its representation of others, assume a defensive aspect, as a wish that one's own outer freedom be *secured against* the threats to it that those others are now represented as posing.⁶ Because the wish to limit others' freedom and self-sufficiency is just as unlimited as the primitive wish from which it springs, it finds immediate expression in a maxim of action, namely the practical judgment that deems it good on the whole to do what one can to limit others' outer freedom, and hence their self-sufficiency, their capacity to realize their ends, where doing so augments, or extends, one's own outer freedom and so also one's own self-sufficiency. (In this maxim, one's limitation of others' freedom is in the first instance conceived merely as a reduction of their ability to limit one's capacity to realize one's ends, hence merely as an indirect and negative augmentation of one's own freedom; but it can subsequently come to be conceived as including the compelling of others in ways that turn the exercise of their capacities to one's own benefit, furthering one's freedom directly and positively.)

We are now ready to apply the test. On the interpretation developed above, the application of the formula of universal law is a matter of considering whether it is possible for every person, or subject capable of practical knowledge, to share the practical judgment asserting the goodness of every person's acting according to the maxim in question. Thus in the present case we are to consider whether it is possible for every person to deem good every person's acting to limit others' outer freedom, where practicable, with a view to augmenting their own. Since here all persons are on the one hand deeming good both the limitation of others' freedom and the extension of their own freedom, while on the other hand, insofar as they agree with the similar judgments of others, also deeming good the limitation of their own freedom and the extension of others', they are all deeming good both the extension and the

represented through this projection is not the form of practical knowledge; it is a represented universality of act, not of validity.

⁶ It is easy to see that by this self-induced confusion the illicit wish works to sustain itself as *ostensible cognition*, in that its surreptitious inversion of cause and effect fosters a very familiar vicious circle of self-justification, which supports in turn suspicion and the resulting enmity that prepares the way for assaults on others' freedom. So far as it is illicit, practical judgment is naturally prone to such sinister dexterity. Clothing itself in a suit of false credentials that give it an appearance of legitimacy, it strives, in the inner court of conscience, to sustain itself in the face of its own lingering implicit consciousness of its violation of practical cognition's form.

limitation of both their own and others' freedom. These judgments are inconsistent insofar as the extension of a person's freedom is incompatible with the limitation of that same freedom. And since all persons necessarily wish for self-sufficiency and the freedom it involves, the only way of avoiding this contradiction is by not extending one's own freedom in ways that limit or assault the freedom of others.

It may be noted here that although the inconsistency just pointed out reveals the practical incognizability of the action represented in the maxim of assaulting others' freedom, it is nevertheless possible to conceive of this maxim's being universally adopted without any actual conflicts or incompatibilities arising between person's choices, so far as these are considered in abstraction from the underlying wishes they express. This possibility is left open because choice always depends on judgments of practicability. In a condition of civil society in which externally legislated laws of outer freedom were promulgated to all and universally recognized as enforced with sufficient stringency for all, persons would not regard assaults on others' freedom to be practicable, or at least would not deem them good on the whole. In such an environment, everyone's following the maxim of assaulting others' freedom just where doing so would further their own could in principle result in their strict compliance with these laws and hence in the complete absence of assaults—even of attempted assaults—on others' freedom.⁷ Here there would be perfect harmony at the level of choice. The practical incognizability just brought to light would nevertheless remain at the deeper level of the wishing that the maxim of action expresses. For the same inconsistency emerges when the universalization test is brought to bear just on the wish that one's own freedom be secured and that others' be limited so far as necessary for that end. Indeed, this difference between the levels of choice and wish is just what the story of Gyges' ring vividly illustrates. As we noted earlier, it is always through its bearing on wish that the form of practical cognition bears on choice.

For related reasons, framing the maxim in variant terms as one of merely *trying* to assault others' freedom would not eliminate the inconsistency, even though we can easily enough conceive of everyone's

⁷ Much the same result could in principle hold even in a state of nature, to the extent that persons shared the same theoretical cognition, including judgments of practicability, and were aware of one another's wishes and choices.

attempting this (as we do when we think of the Hobbesian natural condition). For the underlying wish would remain the same. Nor would such a way of representing the action be genuinely practical. Those who are modest or diffident of success (uncertain in their *theoretical* judging) may describe themselves as trying to do something rather than as simply doing it. But such qualification is extrinsic to practical thought itself and plays no essential role in its operation. To try to do a thing is to mean to do it.

3. *The Maxim of False Promising*

Having now described in outline how the universalization test applies in the primitive case of a maxim of assaulting others' freedom, we are in a position to consider its application in a more complex instance where further concepts, such as that of a promise, are presupposed. So let us return to Kant's example of the man who acts on a maxim of false promising, a maxim specifying that when one sees oneself to be in need of money, one is to borrow it, promising to repay even when one knows repayment will not be possible. Kant says this principle—or "presumption of self-love," as he also calls it—"could never be valid and agree with itself as a universal law of nature, but must necessarily contradict itself," since the universality it would have as a law "would make the promise, and the end one might have in it, itself impossible, in that no one would believe that anything was promised him, but would laugh at all such utterance as vain pretense" (*G* 422).

When we first reflect on Kant's discussion of this maxim, our thoughts are easily drawn to the salient features that give the proposed conduct its specificity, distinguishing this case from the one just considered. Thus it is often noted that the action depends on a shared possession of the concept of a promise, and that, as "a lying promise" (*G* 402–403, 429), it counts as a type of deception. But as we shall see, these facts, though integral to the example, need to be considered with care. Because our attention can easily fix on one or another of them, we face a greater risk here than we did before of overlooking the specific *type* of wrongdoing Kant has in view and so a greater risk of misunderstanding how the universalization test is being applied. On account of this complication, it will prove important, in thinking about this example, to bear in mind that an action can be open to criticism in more than one

respect. If, for instance, a prince were to coerce one of his subjects to bear false witness against another, his attempt to harm the latter would still be wrong even were it not also an act of coercion, and the coercion would still be wrong even were it done not to harm but to assist. We shall accordingly approach this example in stages, taking ourselves by successive steps more deeply into the case.

The observation that making a promise to someone depends on a shared possession of the concept of a promise might encourage us to think that the notion of a general practice, or convention, figures centrally in Kant's reflections. Thus, according to one commonly adopted line of interpretation, which we touched on earlier, Kant is relying on the idea that promising is possible only in a community where a practice of promising is in place, a practice that false promising exploits and erodes.⁸ As we noted, this approach faces obvious difficulties if offered as a general recipe for thinking about violations of perfect duty to others. But now that we have found another way of understanding cases in which no practice or convention is exploited, we may briefly revisit this reading, treating it as specifically tailored to abuses of the practice of promising.

It seems fair to say that, to the extent that this practice is itself something good, and to the extent that promises do in some way depend on its existence for their very possibility, while false promises tend to undermine it, it may be possible, by developing this interpretation, to bring into view one respect in which action on the maxim of false promising is open to criticism. It is difficult, however, to find in Kant's discussion of the example any indication that he is drawing specifically on the notion of a general practice or convention. Moreover, such an interpretation points us toward the thought that it is the practice of promising or the community it defines that is injured by action on this maxim. But this community includes the man who acts on the maxim as well as those whom he may attempt to defraud. So this reading leaves it unclear why such action should be considered a violation of a duty to others. We may therefore set this line of interpretation to one side, for the present at least, and look for a way of applying the universalization test that does not rely specifically on the notion of a practice.

⁸ See, for example, John Kemp, "Kant's Examples of the Categorical Imperative," *Philosophical Quarterly* 8 (1958): 63–71, and Robert Paul Wolff, *The Autonomy of Reason*, 166–167. Such a reading may also be reflected in §135 of Hegel's *Elements of the Philosophy of Right*.

Focusing directly on the idea of a promise would seem to afford a more favorable prospect. For though he makes no mention of a practice of promising, Kant does explicitly remark that when we attempt to think of the maxim as a universal law, we find that the promise itself would be made impossible (*G* 422; cf. *G* 403). So let us regard the case from this vantage point, beginning with a brief consideration of the concept of a promise, and in particular the original concept, that of a promise made in good faith, a concept that does not itself presuppose, but rather is presupposed by, the concept of a false promise.

But first we should observe that “promise” is used in two senses, which correspond to the two senses of “judgment” noted earlier. On the one hand, “promise” can be used to refer to the sort of action to which attention is directed when someone says, “You promised.” In this sense, a promise must already have occurred and so be a thing of the past before any breach of fidelity can take place. But the word is used in another sense when we say, “You broke your promise.” In this usage, a promise must be present at the time of infraction, since nothing past can be changed or altered. Though different, the two senses are clearly related; a promise in the first sense is the making, the coming-to-be, of a promise in the second. It is the latter sense that interests us here.

As Kant articulates it, the concept of a promise presupposes the concept of a contract. A promise is one of two acts, the other being its acceptance, that together make up the components of a contract, a single act of the united will of the parties who enter into it.⁹ Each party agrees with the other, yet the agreement they reach is not a mere sum of the two acts. As a single joint act, the contract is rather the whole through which the component acts are constituted in their relation to one another. Precisely because it lies in an act, this agreement is in a sense artificial: the parties do not just find themselves in agreement, they *make* an agreement.¹⁰ And because it lies in a *single* act, it counts as

⁹ Cf. *MS* 272–273. Otherwise put, a contract is a joint choice. As choice in general is the result of deliberation, joint choice is the result of joint deliberation. Thus, as Kant notes, the constituent acts of a contract (promise and acceptance) must be preceded by preparatory acts of negotiation (offer and approval).

¹⁰ Agreement of this type constitutes thoroughgoing interdependence among judgments (in a sense to be explained below) and so is both stronger than bare agreement between independently judging practical subjects and also different in form from the latter. If two persons concur, say, about the value of a certain dietary practice or regimen of personal hygiene, their judgments can reinforce one another, but they are not thoroughly interde-

a practical agreement not only in the sense that practicality, or self-conscious efficacy, belongs to each of the two acts—the promise and the acceptance—that the agreement comprises, but in the sense that practicality belongs to the agreement itself. A contract is thus an actual joining of wills to constitute a plural practical subject, a practical *we*.

As an act of willing, a contract is an act of practical cognition. Hence it involves the implicit recognition that its terms have universal validity (and so would be equally suitable, *mutatis mutandis*, for an agreement between the two parties were their positions reversed). But as a joint act, it is distinctive in that it constitutes an actual system of practical knowledge bound together not only by interdependencies between the merely objectively differing judgments contained in it but also by interdependencies between those of its judgments that differ merely subjectively. In a contractual agreement of the sort described in Kant's example, for instance, there is not only interdependence between (i) the cognition (shared by both parties) that the person accepting the promise is to lend the money and (ii) the cognition (again shared by both) that the person making the promise is to repay it, but also an interdependence between (a) the practical affirmation of these two thoughts (as interdependent) that constitutes the promise itself and (b) the practical affirmation of them (as interdependent) that constitutes its acceptance. These two relations might be called, respectively, objective and subjective interdependence, for they are founded on the objective and subjective forms of agreement among judgments discussed earlier (§IV.5). But they constitute a stronger interrelation than do the latter in that here the related judgments not only sustain or confirm, but *entirely depend on* one another, so that in each case neither can stand nor fall alone, but only both together.¹¹ This act thus realizes, in the microcosm of the contract, the

pendent. Yet while these forms of agreement between subjects are different in this respect, they are alike in that they are symmetrical. In this latter regard they differ from certain derivative forms of agreement, in which one person's judgment is related asymmetrically to another's, as in cases of deference to expertise or to experience (where the judgments are still independent) and in chains of command (where there is thoroughgoing dependence, but in one direction only).

¹¹ Objective interdependence among practical judgments figuring in persons' choices constitutes the unity of action and hence binds together even the choices of ends and the choices of necessary means to them that individual persons make in practically determining what they are to do (§III.1). But in such a case the relation between means and end is nevertheless asymmetrical—indeed, doubly asymmetrical—in that the means depends on

shared practical knowledge thought in the presupposition of universality, but in a distinctive way, as a type of practical knowledge in which, on account of the *subjective* interdependence of the judgments figuring in it, the original *identity in the capacity* for practical knowledge shared by all persons is actualized in an *identity in the act* in which the parties to the contract participate.

It can be seen from the foregoing that the actual joint practical cognition constituting a contract is equally an actual (positive) practical law—an instantiation of the form of practical law—to which the participants are bound and through which they are bound to one another. It is also clear that a promise is nothing but that very law, but in its relation to the party making the promise, just as the acceptance is that same law in relation to the other side. Fidelity to a promise is thus fidelity to a law, and if a promise is broken, that law is broken also.

We noted earlier that a person's outer freedom can be subject to another's assault in conditions where one person can know what another means to do and can influence the latter's intended action. Promising too depends on these conditions, as persons join their wills in practical agreements so that their freedom in the pursuit of their ends can be mutually furthered or at least secured against interference. But as a component of joint willing, promising depends on a further condition, namely the mutual communication of practical judgments between the parties to the agreement, and it therefore directly depends as well on their ex-

the end just on account of a dependence in respect of goodness, whereas the end depends on the means just on account of a dependence in respect of practicability. In joint willing, in contrast, the objectively interdependent judgments are related *wholly symmetrically*, in that each of the actions they represent depends both for its goodness and for its practicability on the other. In such willing, each action is at once both the other action's end and its necessary means.

To say that there is such reciprocity of end and means in jointly willed action may at first glance appear to imply, absurdly, that in cases where the jointly willed actions are completed at different points in time (as in the present instance, where the lending must, of course, precede the repayment), an end will be achieved prior to its necessary means. But in this type of practical knowledge, none of the interdependent actions figuring as constituent ingredients of what is jointly willed are considered complete until all are: such cognition, we might say, "regards nothing as done so long as something more remains to be done" (Bxxiv). Lending, properly conceived, is a temporary making available; it comes to an end precisely when the repaying does. The same reciprocity figures in exchanges of labor: if you help me harvest my field today and I fail to keep my promise to help you harvest yours tomorrow, you will be entitled to reparation, preventing me from securing the benefit I would otherwise have received from your assistance.

ercise of their capacity to communicate. For as joint practical cognition, the act of agreement comprising the promise and its acceptance includes its own distinctive form of self-consciousness, which lies in the parties' joint recognition that they are together making the same objectively interdependent practical judgments determining how they are to coordinate their conduct. Such recognition depends on the parties' exercise of their capacity to communicate their judgments to one another, in declarations, or professions, made by each side to the other, and in each side's understanding what the other has expressed. Only through exercising this capacity can persons achieve the subjective as well as objective interdependence in practical cognition requisite for a contract.

Beyond this relation to joint willing, the capacity to communicate bears an essential relation to the form of practical knowledge and even to the form of discursive knowledge in general. For communication constitutes a shared self-consciousness (a *we*) in the form of the *mutual recognition* by different subjects of a subjective agreement among their judgments and is therefore the *distinctive* form of interaction—a reciprocal relation of expression and understanding—made possible by the cognitive capacity. The capacity to communicate is thus the capacity through whose exercise alone different cognizing subjects can recognize one another as such. This capacity and its integral relation to the discursive-cognitive capacity itself are accordingly already presupposed in the subjective universal validity constitutive of discursive cognition in general.

Now a false promise (a promise only in an extended sense of the term) precludes the joint willing on which a genuine promise depends, aiming to secure instead a mere outward semblance of such willing, to induce the person to whom the promise is addressed to suppose, erroneously, that a genuine agreement in willing has been reached. In making a false promise one wills to miscommunicate one's will, professing a certain practical judgment one does not in fact hold, and one does this, moreover, in the very act of securing an agreement in profession, or in utterance, characteristic of the communication figuring in a genuine contract. An agreement is to be established in what is said, in the thoughts expressed in the professions, but this agreement is to be merely outward, not the reflection of an agreement in practical judgment. In a false promise, then, the capacity to communicate is employed, not to communicate a practical judgment, but rather to effect a mere semblance of communication and of joint willing by inducing in the person to whom

the promise is addressed the false belief that a practical judgment has been expressed.

This general characterization of a false promise clearly fits the action described in Kant's example. Further specificity is needed, however, to capture the full extent of that action's intended dissemblance. The man proposes to profess a judgment that the requested loan, upon being granted, is to be repaid at a determined time. And he does this in the recognition that he does not actually think that, all things considered, he should repay. But this is not all that he recognizes. It is not just that he is undecided or indifferent as to whether he should repay; he knows repayment will not be possible. He means to promise repayment while recognizing not only that he does not think he should repay, but also that he deems good on the whole an arrangement in which he is not to repay. So besides not willing what he says he wills, he wills the opposite of what he says he wills. The dissemblance figuring in his employment of the capacity to communicate has, therefore, two stages or moments—a positive moment of fabrication and a negative moment of concealment. These will be taken up in the order in which we encounter them as we descend into the maxim.

If we now test the maxim of false promising with the foregoing points in mind, we find that there can be no shared practical knowledge that every person is to follow it. Indeed, this maxim cannot even be conceived as a universal law of our rational nature. Suppose there were a law of false promising that all persons followed out of their shared recognition of its validity and hence knew to be a law without needing to discover it through experience of its effects. Then everyone would practically recognize, in the concrete instance as well as *in abstracto*, that all persons are to conduct themselves according to this law. Hence everyone *could* at least in principle and therefore *would* in the practical-cognitively ideal case recognize (in accordance with the expectation grounded in their shared practical knowledge) any false expression of a practical judgment that might be addressed to them to be the false expression it is and so would not—indeed, could not—believe it (cf. 8:426). Now the possibility of the addressee's believing that a practical judgment is being expressed is, of course, a condition of the possibility of what is willed in the false promising, namely to induce in the addressee the false belief that a certain practical judgment has been expressed, which in turn

is necessary if the latter is to be induced to suppose, erroneously, that a genuine agreement in willing has been reached. The attempt to will the maxim of false promising as a universal law thus leads to a contradiction, in that it involves the attempt to will, or to deem it good, that such false beliefs be induced while recognizing that they cannot be.

This contradiction can also be expressed in other, closely related terms. Since the maxim's universality would entail the impossibility of inducing these false beliefs, we can say that under the condition of universality the promising would be necessarily futile. The promising can also be represented as itself impossible, for in practical cognition, if a certain object is recognized to be inherently impossible, there can be no representation of any action that would produce it. As Kant puts it, "there would properly be no promises at all, since it would be in vain to profess my will with regard to my future actions to others who would not believe this profession" (*G* 403).¹²

This line of reflection is noteworthy in that the contradiction it brings to light—the contradiction arising in the attempt to conceive of persons believing those very expressions of practical judgments that they recognize to be false and hence cannot believe—is exposed through a consideration of the maxim with regard to no more than practical cognition's *subjective* universal validity. Objective universality plays no operative role. In the case of the maxim of assaulting others' freedom, in contrast, the idea of objective universal validity figures essentially in the generation of the contradiction. As we saw, it is in attempting to conceive of all persons enlarging their freedom through limiting the freedom of others that the conflict arises.

An explanation of this difference comes into view when we deepen our reflection on Kant's example. The man described thinks he needs money and cannot see a way out of his predicament except by securing a loan. He also recognizes that obtaining a loan will require a promise of repayment, a promise he knows he will not be able to fulfill. In recognizing that a promise is required, he understands, at least implicitly, that the person from whom he seeks the loan would not agree with his judgment—his

¹²The interpretation just outlined is similar in some respects to the one suggested by Allen Wood in "Kant on False Promises," in *Proceedings of the Third International Kant Congress*, ed. Lewis White Beck (Dordrecht: D. Reidel Publishing Company, 1972), 614–619 (see also *Kant's Ethical Thought* [Cambridge: Cambridge University Press, 1999], 87–89).

“presumption of self-love”—in favor of securing an arrangement whereby the needed funds are to be made available to him, but without any repayment to follow. And his awareness of this judgment’s manifest lack of subjective universal validity is what prompts him to specify the action this judgment represents as good on the whole in such a way that the carrying out of that action will keep the judgment hidden from view, concealed beneath the false profession of a judgment affirming that the requested loan is to be repaid should it be granted. Obviously, then, the contradiction just identified arises immediately once we consider the maxim in the light of the subjective universal validity of practical cognition. For in effect that contradiction exposes, under the condition of universality, the latent contradiction between what the man means to do and what he says he means to do.

This contradiction is discernible merely from a consideration of the positive moment in the dissemblance. But from the points just noted it can be seen that the latter, the fabrication, is a specific way of carrying out a more general plan of exploiting others’ trust, a plan that depends for its success on a negative moment, the duplicitous cover-up, in which a more basic inconsistency is already present. For in the self-disguising of his own practical judgment, the man makes it the object of his will to insulate his practical judging from the communication openness to which is already implicit in cognition’s subjective universal validity (and he does so, moreover, by misusing that cognition’s very capacity to communicate). In this conflict between the judgment’s object and the presupposed subjective universal validity belonging to it as an exercise of the capacity for practical cognition, we find the false promise’s original infringement of the form of practical cognition.

Thus the contradiction initially described rests on a more basic underlying conflict. What we first noticed was the inconsistency in supposing that the promises are to be believed and that they cannot be believed. But the reason why the promises cannot be believed is that everyone is cognizant of the underlying judgment that conflicts with what is professed in the false promise. This cognizance, however, includes the recognition that this judgment—the judgment that the loan is to be acquired but not repaid—has specified itself into the self-disguising judgment that this acquisition without repayment is to be achieved by hiding this very plan from others’ view. Thus, anyone to whom a promise is addressed already recognizes the deeper conflict just

noted, that to agree with such an underlying judgment would be to agree—and hence to judge practically, to recognize as valid—that one’s own awareness of the judgment (through communication) should not be allowed. This, however, cannot be thought, since it is this very awareness on which the possibility of that same agreement rests. This impossibility is evidently part of what Kant has in view when, in revisiting this example in connection with the formula of humanity, he says that those to whom the false promise is addressed “cannot possibly agree” with the way they are being treated (*G* 429–430). The validity of practical cognition entails its communicability across persons. Such cognition cannot have as its object—the thing it works to realize—the exclusion of the very possibility of communication that its own validity presupposes.

Shining this light on the roots of the maxim also exposes, of course, a yet more deeply underlying conflict, prior to the contradictions we have traced to the false promise, prior even to the specification through which the maxim comes to involve a plan of false promising in the first place. This deeper conflict is between the man’s affirmation that it would be good for him to secure the funds without any repayment to follow and the view he knows the prospective lender would take of such a course of action—the very conflict in judgment that would *actually* arise between the two of them were the man candidly to express his will. We shall presently return to this underlying affirmation. Consideration of it will reveal a crucial commonality between this maxim and the maxim of assaulting others’ freedom, thereby completing our comparison of the two cases and our explanation of the difference in the ways they give rise to conflict under the condition of universality.

But first a brief summary of what has come to light so far. Like the reading that relies on the notion of a practice, the interpretation of Kant’s example we have just been considering seems capable of bringing into view a respect in which the action of making a false promise is illicit. Interpreting the universalization test as bringing out the wrongful character of such action by revealing, in each of the two stages we have distinguished, the practical incognizability of its use of the communicative capacity of practical reason reveals a misemployment of this power, an abuse and injury of the capacity for practical knowledge as a communicative power comprising both the capacity to express (with candor) and the capacity to understand (in trust) (cf. *MS* 429). Yet this interpretation—again

like the reading that draws on the idea of a practice—does not reveal how the action of making a false promise violates a perfect duty to others, though it is clear from Kant's classification that it is a duty of this sort that the example is intended to illustrate.

Here it will be helpful to return to Kant's discussion of the case. In focusing on the notion of a promise, we have ignored the fact that Kant directs our attention to more than just the promise when he explains why the maxim "could never be valid and agree with itself as a universal law of nature, but must necessarily contradict itself." He says the universality it would have as a law would make impossible not only the promise, but also "the end one might have in it." Moreover, in explaining these impossibilities, he says not only that "no one would believe that anything was promised him," but also that everyone "would laugh at all such utterance as vain pretense." On first consideration, this suggestion that all utterances through which false promises are attempted would be met with derision might look to be a mere rhetorical flourish, an embellishment of the point that they would not be believed. Yet it clearly introduces further ideas beyond the bare thought of disbelief. The natural interpretation of this laughter and the scorn it manifests refers our attention to the action's end, the end Kant has just said would also be made impossible, along with the promise, by the universality the maxim would have as a law. For while it is of course possible for persons to laugh at an utterance they do not believe without considering its end if they think it expresses an innocent error that betrays a contemptible weakness of understanding, such a thought is obviously not the source Kant has in mind, since the falsehood is recognized to be intentional. Rather, the ridicule expresses both a contempt and a hostility on the part of the person to whom the utterance is addressed that each depend on awareness of the action's end: contempt in the recognition of the hopeless futility of the profession as a means to that end (since it cannot be believed, it is a "vain pretense"), and hostility stemming from a recognition that the end conflicts with this person's own end.

Let us, then, widen our view of the maxim and consider not just the false promise by itself but also the use to which it is to be put. Here it will be necessary, however, in accordance with the approach we have been following, to consider the action's end not in the relation that specifically grounds the contempt the laughter expresses, but rather in the relation that gives rise to the deeper feelings of hostility and opposi-

tion on which the derision ultimately depends. That is to say, we shall consider the end, not with reference to the means, but with reference to the corresponding end the recipient of the false promise must be conceived as having when we attempt to represent the maxim as a universal law. The general reason for focusing on the latter relation has already been noted (§VI.5) and will receive further discussion below (§VII.7).¹³ But it is in any case clear that revealing the maxim as a violation of a perfect duty to others must be a matter of finding, not a feature on account of which action upon the maxim might (at least under the condition of universality) expose one to another's contempt, but rather the feature through which it can provoke feelings of opposition and enmity in others whose wills are determined by the same practical principle.¹⁴

From this broader perspective, it is not difficult to see that the universalization test applies here in a manner that parallels—indeed, instances—its application in the previously considered case of the maxim of assaulting others' freedom. In the maxim of making false promises to obtain loans one cannot repay we find the same general judgment to be operative. The action of limiting or impeding others' freedom—their self-sufficiency, or capacity to realize their ends—is deemed good on the whole in cases where doing so augments, or extends, one's own freedom, one's capacity to realize one's own chosen ends, which one also deems good on the whole. This case differs from the previous one in that here the limiting of others' freedom lies specifically in diminishing their property rather than simply in impeding their innate outer freedom, and also in that this limitation is to be achieved by exploiting the communicability of practical reason in general and the readiness to trust on which its exercise depends, rather than by exploiting a natural capacity or advantageous circumstance, such as physical strength. But the factor in this maxim on account of which it violates a perfect duty to others lies not in these specific features that distinguish it from the general maxim of assaulting others' freedom, but in the more basic judgment

¹³ Considering the end in the other relation—to the means—would lead toward what is often called the "practical" interpretation of the contradiction Kant describes (see §VII.7).

¹⁴ Recognition of the necessary futility of the means under the condition of universality may, of course, enhance the impact the universalization test has on the mind, to the extent that attention to this futility tends to humiliate the illegitimate presumption of self-love. Indeed, as we noted earlier, this very circumstance may partly explain the appeal of interpretations that locate the contradiction in this futility. But while such annihilation of efficacy may be the consequence of a maxim's wrongfulness, it cannot be the ground of it.

that these two maxims share in common. It is at the level of this more basic judgment, then, that in attempting to conceive of the maxim as a universal law we encounter a contradiction that depends essentially on considering it in the light of practical cognition's *objective* universal validity.

Of course, to bring this underlying contradiction clearly into view, it is necessary to prescind from the two-stage contradiction we have already seen arise directly in the false promising itself, the contradiction that results merely from considering the maxim with reference to practical cognition's universal validity in respect of subjects. This is in fact what we find Kant doing in his initial treatment of the case in Section I of the *Groundwork*. After remarking that "it would be in vain to profess my will with regard to my future actions to others who would not believe this profession," he adds: "or, if they over-hastily did so, would pay me in my own coin" (*G* 403). In making this additional point, he momentarily ignores the impossibility of promising, supposing instead, counterfactually, that promises would be believed, in order to direct attention to the essential incompatibility between the ends. He thereby puts us in a position to appreciate the impossibility of conceiving of everyone, through reciprocal dupery, increasing their own property by diminishing that of others.

In sum, the maxim of false promising involves a violation of duty in more than one respect. Like the prince in Kant's example, who misuses one subject in attempting to carry out an action that will injure another, the man who acts on the maxim of false promising misuses the communicability of practical reason—therein violating a duty he has to persons generally—in an action that also violates his duty to the person to whom he makes the false promise. Attempts to understand Kant's example by concentrating on the concept of a promise (and perhaps, indirectly, also those relying on the notion of a practice of promising) seem, accordingly, to be capturing an aspect of the case that figures in his own understanding of it, an aspect that relates specifically to capacities persons have to enter into actual communities of willing. Once such capacities are brought into consideration, duties come into view that concern the employment of humanity in a community, as we might put it, rather than merely the employment of humanity so far as it is immediately present in individual persons. Further investigation of such communities would require consideration of how they are constituted and how they relate to the end of

practical cognition (the highest good)—questions that would take us beyond our immediate purpose here. But insofar as we consider the example just as an instance of a maxim in violation of a perfect duty to others, as Kant's placement of it in his taxonomy suggests we should, we prescind from this aspect of the case, focusing instead on the conduct as it bears on the person to whom the false promise is to be made, and find there the same practical incognizability that we found in the case of the conduct represented in the maxim of assaulting others' freedom.

4. *Imperfect Duties to Others: The Duty of Beneficence*

For a case involving an imperfect duty to others, we may consider the one Kant himself takes up in the last of his four examples, which concerns the duty of beneficence, the duty to make the happiness of others an end of one's own (*G* 423).¹⁵ This case, it will be recalled, has a particular importance for our interpretation, since it is often thought that the formula of universal law merely forbids acting on maxims that cannot be universalized and therefore differs from the formula of humanity in being incapable of generating any positive duties, such as that of beneficence. Much depends, therefore, on whether the formal constraints on willing that the universalization test as interpreted here makes explicit include this most basic of the imperfect duties to others.

In contrast to the perfect duties considered above, which pertain to practical judgments concerning the exercise of outer freedom, the imperfect duty of beneficence pertains to practical judgments concerning happiness, the object of such freedom. So before turning to Kant's example, we should call to mind a few points from our earlier discussion of happiness

¹⁵Kant does not say in so many words that his fourth example concerns this duty, and it is sometimes supposed that it deals only with a duty to help others in need, not with a full duty to make others' happiness an end of one's own. But that he has the latter duty in mind seems clear from what he says when he revisits the example in connection with the formula of humanity (*G* 430), and from the resemblance his treatment of this example bears to his explicit arguments for the duty of beneficence at *MS* 393 and 453. And while the full duty of beneficence may initially seem more demanding than a duty to help others in need, they come to the same thing if, as was argued earlier (§III.5), self-sufficiency is a necessary object of wish and as such belongs to happiness as an essential, formal element. The transition from the formula of universal law to that of humanity is simply the transition from a maxim's rule of action to its end (*G* 436), that is, from a practical conception to its object (which is also its subject), and these are but two aspects of the same thing: the duty to help others in need is just the duty to make others' happiness one's end.

(§III.4–5). Although self-sufficiency, as a necessary object of wish, belongs to the object of a human person's generic conception of happiness, it does not, of course, exhaust it. The very fact that for human persons practical self-sufficiency is an object of wish rather than a wholly secure and certain possession is itself a reflection of the fact that the object of the generic conception of happiness includes in addition, under the general heading of agreeable activities, further elements (not yet specified) whose realization is recognized to be dependent on the presence of conditions that are external to such persons and hence not absolutely within their control. Such persons, therefore, can never rule out the possibility that they will find themselves in circumstances in which they are not capable of realizing on their own such further constituents essential to their end of happiness. Yet they also recognize that this end is not something they can simply relinquish, for in their own empirically grounded practical judgment the agreeable activities it necessarily contains are recognized to be simply good and hence to be neither objects of arbitrary election nor dependent for their goodness on their practicability.

With these considerations in mind, we may turn to the maxim of the man in Kant's example. Though willing to confine his pursuit of his own happiness to actions that do not involve any assault on others' freedom, this man, for whom "things are going well," calculates that his pursuit of this end will be most effective if others' ends are disregarded, and he accordingly specifies the good through the generic conception of his own happiness yet without including this specification in a general judgment deeming the happiness of others to be likewise good. He may indeed judge that others' ends are good on the whole in circumstances where, by his estimation, the attainment of those ends will ultimately redound to his own benefit. But such a judgment falls short of regarding others' generic end of happiness in the way the duty of beneficence requires, namely as being simply good and hence good regardless of whether their attainment of it contributes to the realization of his own end. In short, the man Kant places before us could be described as a "moral libertarian" whose way of thinking prescribes a practice of "egoism within the limits of justice."

Let us now bring the formula of universal law to bear on this maxim of indifference to others' ends, to determine whether it is in agreement with the form it presupposes. When the man tries to suppose that

everyone adopts his way of thinking and that everyone agrees that everyone should do so, an obvious conflict results. If this way of thinking were universal, others would judge their happiness but not his to be simply good, and if he were to agree with them in this judgment, he would likewise regard their happiness but not his to be simply good. He would thus deem good his happiness but not theirs and also their happiness but not his. So his maxim contradicts its form. And since all persons necessarily wish for their own happiness, his only way of avoiding this inconsistency is by extending his judgment of the goodness of his own happiness so that it includes that of others, that is, by including his judgment within a general judgment that also deems the happiness of others to be good.

This way of understanding how the attempt to will this maxim as a universal law results in a contradiction differs from other familiar accounts, both sympathetic and critical, in that, being guided by the idea of the form of practical knowledge, it does not conceive of the attempt to will a maxim as a universal law as involving any prudential or instrumental forms of reasoning. It does not, for instance, attempt to show that were everyone to follow the maxim of indifference, the result would thwart, hinder, or fail sufficiently to further the pursuit of some end the man actually has, such as happiness, or some end he must have, as a rational agent; nor does it rely on any calculation that individual or collective interests will be better served if persons help one another than if they do not (see §VII.7 below). These other accounts have been encouraged in part by the fact that Kant's own discussion has seemed to many to indicate that he himself relies on means–end reasoning in describing how the contradiction arises, and this fact may therefore equally be a source of doubt concerning the present interpretation. This doubt can be removed, however, by the following considerations, which show that the example need not be read as involving such reasoning.¹⁶

Kant explains how the conflict in the will arises by saying of one who attempts to will the maxim as a natural law that “many cases could occur in which he would need the love and sympathy of others and in which, by such a law of nature sprung from his own will, he would rob

¹⁶ Also to be noted is the clear absence of means–end reasoning in other passages in which Kant derives the duty of beneficence; see esp. *KpV* 34–35, *MS* 393.

himself of all hope of the assistance he wishes for himself" (G 423). We can take Kant to be characterizing here the very contradiction we just now described, but in a vivid way suitable for someone such as the man in question, who, because "things are going well" for him, is particularly liable to have acquired an exaggerated sense of his own self-sufficiency and so to have lost clear sight of what the formula of universal law helps bring back into view by lifting his thoughts from his immediate position to a higher, general outlook, namely that as his own happiness is not a wholly secure possession, circumstances might arise in which he would rightly wish for help. Now on the present interpretation, such a wish is not merely a wish for some perhaps entirely fortuitous occurrence that the man thinks would enable him to achieve his end, like a wish for rain by someone thirsting in a drought, or a debtor's hope for a lucky roll of the dice. It is also an expression, in those circumstances of need, of the subjective universal validity of his own practical judgment of the goodness of his own happiness, a judgment that is contained in the maxim under consideration (cf. MS 393). That is to say, the wish in question is originally for help that springs from another person's practical recognition of the goodness of the man's happiness, and hence for help that has an immediacy and necessity that would be lacking were the assistance to be contingent upon its provider's determination that it will in the end be personally advantageous. But since to will this maxim as a universal law the man must also not will that such help be provided, he must both will and not will to be helped. As is readily apparent, this contradiction is the same as the general contradiction described above, the only difference being that it is represented more concretely, through the man's consideration of possible cases in which he would need assistance.

It is easy to see that the duty of beneficence brought into view by this application of the formula of universal law involves the requirement that others' happiness be deemed *equal* to one's own in basic goodness. Suppose the man in Kant's example had thought, Yes, others' ends are good; but goodness is a matter of degree, and I assign a greater worth, a greater importance, to my own ends than I do to those of others. The universalization test as here interpreted reveals a conflict within this maxim as well, parallel to the one we have just seen in the maxim that recognizes in others' ends no positive worth at all. It is no more possible to suppose that one's own ends have more worth than others' and that

others' have more worth than one's own than it is to suppose that only one's own ends are good and that only others' are good.

Taken by itself, this requirement of equal worth will perhaps seem an excessive demand, not in accordance with our antecedent understanding of the duty of beneficence. If we suppose it to mean that we are to care just as much about everyone else's concerns as we do about our own, it may strike us as not even intelligible, much less a true obligation. But here, as with many other commonplace propositions of ethics, some care is needed to recognize the validity in it and to distinguish its proper meaning from other senses that can make it seem erroneous.¹⁷

In order to see this duty in a proper light, it is important to bear the following points in mind. First, as interpreters often emphasize, beneficence is an imperfect duty, which simply requires the adoption of certain ends, calling on us to further others' happiness so far as we can (cf. *G* 430). Thus, even though a positive obligation, it does not prescribe any determinate action at all, much less action owed to others. To succeed in one's efforts to contribute to other's happiness is meritorious; to fail is not to wrong the intended beneficiary.

Second, the duty of beneficence is not to be conceived as directing us to increase so far as possible others' pleasure or enjoyment, their happiness conceived merely as an ideal of the imagination. The happiness in question is happiness as an object of wish, which includes self-sufficiency as a component of its very form. Hence the duty of beneficence, even though founded on the idea that each person's happiness, generically conceived, is equal in goodness to every other's, cannot require any positive action furthering others' ends beyond that of assisting others where they are in need. And for the same reason, the object of beneficence is always first and foremost to help others secure or recover their self-sufficiency.

Third, the duty of beneficence originally concerns others' happiness in its generic, formal conception, the conception involved in the very notion of a person in general. From the fact that every person's happiness generically conceived has equal worth, it does not follow that this or that specific object of wish that a given person might introduce in

¹⁷ Cf. Kant's treatment of the related biblical injunction to love one's neighbor as oneself (*G* 399, *MS* 402).

arriving at a specific conception of personal happiness (say, to rule Milan) is part of what the duty of beneficence obliges one to deem good (cf. *MS* 388). Obviously this duty cannot possibly require that another's impermissible ends be deemed good, and what it requires of us is sensitive to further distinctions as well. One such is the distinction between elements of happiness that are necessary to it, in that they are grounded in *natural* inclinations—that is, inclinations based in human instincts, such as those directed to the preservation of the individual or the species—and elements that are by comparison arbitrary and contingent. Elements of the former sort fall under the heading of “true natural need, in which our species is in thoroughgoing agreement with itself” (*KU* 430). While they are contingent and belong to the *content* of our happiness so far as we view ourselves merely as rational beings in general, they are necessary and belong to its *form* when we regard ourselves more specifically as human beings.

Many other objects of pursuit, such as those figuring in innocent forms of competition, are not only contingent but not even objects of practical judgment.¹⁸ A wish to excel in a certain activity or field of endeavor might at first glance seem to conflict with the duty to place a worth on others' happiness equal to the worth one places on one's own. For a wish of this sort constitutes a relation of direct (not merely incidental) competition with others, and where persons so compete, the objects they pursue are always intrinsically opposed. But such a wish can be compatible with the form of practical knowledge if it is a practical thought that is consequent upon a joint (though possibly tacit) practical judgment that constitutes a cooperative pursuit or activity, such as a contest or a game in which winning is the object. In such cases, each participant sets itself in opposition to all the others in necessarily wishing to prevail in the contest, while all the participants agree in wishing that each of them wish for this “same” object. But this object—to prevail over the other contestants—is merely an object of free (arbitrary) practical *thought*, not an object determined to be good in a practical *judgment*. In their joint practical judgment the competing participants together

¹⁸The contingency of such objects does not imply that the impulse to compete has no footing in human nature. Kant places considerable emphasis, in fact, on the idea that contained within human sociability itself there is an essential moment of *unsociability*. For discussion of this idea, see Wood, *Kant's Ethical Thought*, Part II, *passim*.

deem it good that each pursue, in accordance with the terms of the competitive activity, the object of prevailing over the others, and they thereby also deem good the acts of practical thought in which they set themselves to pursue these intrinsically opposed objects and thereby constitute themselves as opponents. But none of the contestants deem any of these objects themselves to be good, not even the one that lies in their own prevailing over the others (though nothing would block them, of course, from using, in connection with those objects, the *relational* concept of the good described earlier (§II.7)). That these objects are not deemed good is reflected in the fact that the mutually recognized opposition in practical thought between the objects the contestants pursue does not generate the feelings of opposition and enmity that characteristically arise in cases of conflict in practical judgment. But it is also reflected in the fact that no one is obliged by the law of beneficence to regard any contestant's aim as anything good. (Of course, the very structure of the competitive relation, in which the parties jointly impose on themselves the condition that the object each of them pursues be one of complete competitive self-sufficiency in relation to the other contestants, precludes from the start any possibility that one contestant might ever, in the competitive activity itself, need the generous assistance of another, and in many forms of competition assistance from outsiders is similarly precluded or limited as well.) In sum, the pursuit of an object of innocent competition, notwithstanding the coldness with which it regards whatever failures and defeats may be suffered by the opposition, does not conflict with the duty of beneficence; nor does this duty oblige those opponents or anyone else to regard the object of that pursuit as good. Small wonder that the ungenerous man in Kant's example is so ready to portray our natural situation in terms that invite him to think of it as a contest: "Let everyone be as happy as Heaven wills or as he can make himself" (*G* 423).

Finally, the basic duty of beneficence is grounded merely in a consideration of human persons in general and the practical judgments they make regarding their happiness, generically conceived. Nothing in the foregoing account precludes the possibility that given persons may be under further, more determinate obligations of assistance based on specific forms of human association, such as friendship, the family, or political society—obligations that may limit their ability to assist others to whom they are not thus related.

5. Textual Confirmation

If the foregoing interpretation of Kant's example concerning the duty of beneficence is sound, it follows that the formula of universal law can support positive as well as negative duties. This outcome will perhaps seem surprising or questionable in view of what may appear to be the plain fact that this formula merely forbids acting on maxims that cannot be willed as universal laws. It will therefore be helpful at this point to look more closely at a noteworthy feature of Kant's discussion of the just but ungenerous man's maxim in order to confirm our interpretation.

It is a very significant fact that in his treatment of this example Kant conceives of this man's maxim as one that, though it does not deem others' ends to be good, is not utterly bereft of all thought of others' happiness. One might wonder why he thinks of it in this way. The practical effect, after all, would be essentially the same if, rather than following the maxim Kant describes, which lays down that others' happiness is to be disregarded, the man were instead to follow a maxim of prudence that simply specified that he is to treat others in a way that, without violating their freedom, best furthers his own happiness. And the application of the universalization test to this latter maxim would equally result in a contradiction; the outcome is the same regardless of whether the way of thinking is one in which concern for others' happiness is rejected or one in which their happiness is simply neglected.

An explanation of why Kant thinks of this man's maxim as involving the thought of others' happiness begins to emerge once it is noticed that he presents this way of thinking as arising in this man's will, not in an utterly spontaneous way, but rather as a *response* to another thought, quite different from the one that constitutes his maxim: "a fourth, for whom things are going well, while he sees that others (whom he could very well help) have to struggle with great hardships, thinks: What concern is it of mine? Let everyone be as happy as Heaven wills or as he can make himself . . ." (*G* 423). This man does not ask "What concern is it of mine?" simply out of the blue. The question, along with its hint of irritation, arises in response to a thought that is already present, if only obscurely, in his own mind—namely that others' happiness *is and ought to be* a concern of his. Seeing others who have to struggle with great hardships arouses his own obscure practical understanding that their happiness is a matter of concern, and not just to them, and this stimulation

leads to the further reflections described in the example. His enlivened yet still indistinct understanding is not enough to move him to come to their assistance, of course, but it carries enough weight in his thinking to prompt the recalcitrant and ungenerous response by which he seeks to counter it in the court of his conscience. But where is this obscure thought coming from?

On the account of the universalization test being suggested here, the answer is obvious: This thought is nothing other than the implicit universality of the judgment in which the man determines his own happiness to be good, the objective universal validity figuring in the presupposition of universality. Indeed, the presence of this implicit universality in the obscure thought to which he is responding is reflected in the fact that he frames his response in correspondingly universal terms, propounding not merely a private maxim for himself, but rather—as if he were legislating for all—a universal law: “Let everyone be as happy as Heaven wills or as he can make himself.” Also noteworthy here is that this proposed law obtains its semblance of validity—it is, after all, a commonplace maxim taken from ordinary practical discourse (cf. *MS* 452)—in part from the possibility of recognizing in advance that self-sufficiency is a component of every person’s end of happiness. The universalization test thus brings explicitly into view a universality that is already implicit in the man’s own practical judgment. This universality is not only at work in the thought against which he reacts in framing his maxim of rejecting concern for others’ happiness. It is even reflected in that very maxim, insofar as the latter is itself couched in universal terms, pretentiously mirroring the universality of the thought it opposes, as if a judgment-form of unrestricted quantity were secretly thought to be necessary in order to give the usurping thought the trappings of legitimacy, so that it might, in logical appearance at least, match, or stand up against, the unlimited magnitude of the validity of the original implicit thought it strives to unseat.

It should be also be observed, however, that the universal terms in which this maxim is expressed make it, if not atypical in this regard, at least disingenuous and hypocritical. Kant states that when persons act contrary to duty they do not strictly speaking will their maxims as universal laws, for this is impossible; what rather happens, he says, is that we will that “the opposite of our maxim should instead remain universally a law; only we take the liberty of making an *exception* to it for ourselves (or

even just for this once) to the advantage of our inclination" (G 424). Presumably, then, the possibility of propounding the universal rule "Let everyone be as happy as Heaven wills or as he can make himself" depends on a certain illusion in practical thought, facilitated, as just noted, by the presence of self-sufficiency as a component in the end of happiness, and perhaps also supported in part by a propensity of self-interest to regard itself as in alignment with the will of Heaven, the propensity that makes it possible, as Kant notes, for prosperity to lead to arrogance (G 393). If we suppose, accordingly, that behind this self-deceptive illusion and the explicit espousal of the universal rule it makes possible, the ungenerous man's reflections follow the general pattern just indicated, then he both wills the universal practice of beneficence yet also wills that an exception be made in his own case, thereby degrading in his thought the true universality of the principle of beneficence to mere generality (though even here, Kant argues, the practical reflections are not strictly speaking consistent). What is noteworthy for our purposes, however, is that in this discussion of what standardly happens when a person transgresses duty, Kant again relies on the idea that "the opposite" of the maxim is *already* willed as a universal law. Thus, when we suppose that the ungenerous man engages in what Kant sees as the general pattern of reflection involved in action contrary to duty, we again find that the implicit universality of the judgment in which the man determines his own happiness to be good is still present in his reflection, furnishing the rule to which he wishes to make an exception.

This standard pattern of reflection is also displayed in Kant's treatment of the example involving the maxim of false promising. Kant says the man who is inclined to make the false promise "has enough conscience to ask himself, Is it not impermissible and contrary to duty to help oneself out of need in such a way?" (G 422). Here we can see that the man's contemplation of the prospective course of action, itself prompted by his belief that he is in need, in turn arouses in his own mind this question about the action's permissibility; and we are now in a position also to see this question as itself the manifestation of an implicit universality in the man's own practical judgment of the goodness of the self-sufficiency that he necessarily wishes for in his own case. This obscure recognition that every person necessarily wishes for self-sufficiency on the same grounds that he does prompts him to pose this question to himself and accordingly to scrutinize the proposed action more closely.

The obscure recognition is also manifested in the man's endeavor to represent his action as a permissible exception, which we can see reflected in the modal terms Kant uses when, in applying the universalization test, he speaks of "a law that everyone who believes himself to be in need can promise whatever he pleases with the intention not to keep it."¹⁹ The law is here presented as determining that everyone *can* promise, not that everyone *does* or *must* promise. The presence of "can" seems clearly to express the thought of permission (not mere physical capacity), and in such a thought a general background prohibition or restriction is presupposed, for an action can be regarded as permitted only where a default prohibition is lifted. On the present interpretation, this prohibition lies originally in the man's own antecedent implicit recognition that the self-sufficiency, generically conceived, of each and every person is necessarily good, a recognition with which his proposed plan of obtaining money conflicts. Accordingly, his representation of his action as permitted also reflects an attempt on his part to regard the universality of this implicit judgment as mere generality. It reflects, that is to say, his attempt to rationalize his action: though mere generality is still enough to constitute a default presumption against actions that undermine others' self-sufficiency, it might nevertheless allow exceptions if they can be justified or excused—by urgent need, for example, as the man presumes in the present case.

In sum, positive duties can be brought to light through the application of the formula of universal law in fundamental practical judgments persons necessarily make merely in virtue of being persons at all. The

¹⁹ *G* 422; cf. also *G* 403: "could I truly say to myself that everyone may make a false promise . . . ?" Such passages have encouraged some to suggest that Kant's test might be understood as considering, not whether a given maxim can be willed as a universal law, but whether it is one on which everyone could be permitted to act. (This proposal has been advanced by W. D. Ross in *Kant's Ethical Theory* [Oxford: Oxford University Press, 1954], 30, and by T. M. Scanlon in unpublished lectures on the *Groundwork*; it is elaborated by Thomas Pogge in "The Categorical Imperative," in *Grundlegung zur Metaphysik der Sitten: Ein kooperativer Kommentar*, ed. Otfried Höffe [Frankfurt am Main: Vittorio Klostermann, 1989].) On the reading presented here, the occurrence in these passages of "can" and "may" (as opposed, say, to "does" or "must") reflects no departure from Kant's repeated indications that a maxim is to be tested by considering whether it can be willed as a universal law, but merely the subject's attempt to represent a proposed way of acting as a justified exception to an implicit universal practical judgment with which it conflicts. Independently of the application of the universalization test, the thought of permission resides in the way the maxim comes to be couched in the face of scruples it arouses in moral reflection.

universalization test expresses the requirement that these judgments be made in such a way as to be in agreement with the implicit universality that lies in their form, and when this happens the resulting judgments are universal and self-consistent. We have found such universal judgments to be implicit in the moral reflection Kant describes in his treatment of the maxims of false promising and indifference to others' happiness, and we have also seen that in his account of the pattern of reflection characteristically involved in the transgression of duty, he refers to these judgments explicitly when he says that in such a case we will that "the opposite of our maxim should instead remain universally a law." From what Kant says in these passages, it has emerged that these implicit universal judgments—which represent the *first* and *direct* application of the moral law, or its application in a person's most fundamental practical judgments—constitute a body of background practical knowledge operative in moral reflection. As the voice of the good will within, they play a role in the workings of conscience, prompting moral reflection in particular circumstances where a person may be inclined to act in a manner contrary to duty. Constituting a body of practical knowledge that is universally self-consistent and always implicitly present, these judgments, even if only obscurely thought, can still, to the extent that they are alive in the practical thinking of a person who may nevertheless be inclined to act wrongly, be a stable source of the scruples, the feelings of resistance or hesitation, that put such a person in mind of the need to consider whether a course of action recommended by inclination or self-interest is truly justifiable and in agreement with reason.²⁰ In this body of knowledge constituted by the first application of the form of practical knowledge in persons' fundamental judgments, then, we find morality itself, substantive knowledge determined by a single formal law of reason.

6. *The Formula of Universal Law as a Positive Requirement*

If positive duties can be brought to light through its application, then the formula of universal law must express a fundamentally positive injunc-

²⁰In this respect, these universal practical judgments play a role in moral reflection similar to the one played by the "rules of moral salience" that Barbara Herman discusses in the title essay of her volume *The Practice of Moral Judgment* (Cambridge, Mass.: Harvard University Press, 1993).

tion. This outcome is fully in line with what we were led to anticipate by the considerations on which the present interpretation is based. We saw earlier that the fundamental formal practical law implicated in the presupposition of universality has application in a person's self-constituting act of making happiness an end, thereby giving rise directly not only to a negative law of natural justice, but also to a positive law of beneficence (§V.3). And we drew on similar considerations in arguing in support of Kant's claim that the formula of universal law is equivalent to the formula of humanity, despite the appearance that only the latter can support positive obligations (§VI.7). Yet as we have also noted, the formula of universal law can easily seem—and is often taken—to be a merely negative requirement, prohibiting action on maxims that cannot be universalized. A closer inspection of Kant's formula, however, will reveal not only that there is nothing in his statement of it that constrains us to read it as merely negative, but also that, on the contrary, when it is viewed in the light of the interpretation here proposed, there is a straightforward way of understanding it as articulating what is at root a positive command.

Let us consider again how the formula is stated: "act only according to that maxim through which you can at the same time will that it become a universal law." To regard this formula as expressing a mere prohibition is to take it to be equivalent to the negative injunction: do *not* act according to a maxim *unless* it is one through which, etc. In this reformulation, in which "only" is replaced with "not . . . unless," the injunction turns out to apply only in cases where the condition governed by "unless" is not satisfied: it prohibits acting on maxims that cannot be willed as universal laws, but provides no directive at all concerning maxims that can, leaving us free to act upon them or not, as we please. There is no need, however, to view Kant's formula through the lens of this reformulation. The interpretation outlined above gives us good reason to suppose that the formula is not a mere prohibition, but rather a command that, being issued in the recognition that as persons we act, and indeed act according to maxims, enjoins us *positively* to act according to maxims that can be willed as universal laws. On this interpretation, the directive of pure practical reason that the formula of universal law expresses is indeed positive, though merely formal in character, requiring that the rational activity of determining the content of maxims—the activity of self-determination in which practical judgment itself consists

and which always begins with self-constituting fundamental practical judgments—be in agreement with its own form. In other words, this principle of pure practical reason is not intended simply to be applied to pre-formed, independently given maxims with a view to deciding which should be rejected, with the result that those maxims that survive its test are counted as permissible, but receive no positive endorsement; rather, the principle functions positively to direct persons, in their rational activity of determining the maxims on which they act, to frame maxims that agree with their own form and so can be willed as universal laws. At root, this positivity is just the positivity belonging to the self-sustaining nature of knowledge itself, distinguishing it from the merely negatively self-sustaining character of bare thought (§IV.2).

On this way of reading the formula, “only” does not express any part of the substance of the moral law itself, but merely belongs to the imperatival form it takes in the consciousness of a practical subject in which the will is not perfectly good in its nature (cf. §V.4). The directive that pure practical reason issues to rational beings generally is simply to act according to (maxims that can be willed as) universal laws, or (on the interpretation here proposed) according to maxims whose contents are in agreement with their form. It is only where a particular will is not perfectly good in its nature—where, in other words, inclinations can prompt the consideration, in practical reflection, of possible maxims that can *not* be willed as universal laws—that the directive can manifest itself as the injunction to act *only* according to maxims that can be willed as universal laws (cf. §II.6). In the case of a will perfectly good in its nature, there would be no possibility of its being influenced to act according to a maxim that was not in agreement with its form; since its exercise would in all possible cases be sufficiently determined by reason’s directive to act according to maxims that can be willed as universal laws, the inclusion of “only” in the formulation of this directive would be entirely gratuitous, an affront, as it were, to such a will’s perfection.²¹ Absolutely and originally, then, the moral law lies in the positive directive to act according to maxims that can be willed as universal laws; the prohibition against acting according to maxims that cannot be so willed

²¹ Similarly, if from the formula of humanity—“so act that you employ humanity, in your own person as well as in the person of every other, always at the same time as an end, never merely as a means”—we strip away all that belongs to its imperatival form, we must strike the final clause, “never merely as a means.”

is secondary, deriving from the original positive directive through the latter's relation to imperfect particular wills.

Thus, when in Section I of the *Groundwork* Kant follows the practice of popular philosophy and considers, with the aid of concrete examples, the concept of duty, which contains the idea of the moral law in its relation to an imperfect will, he is led to a statement of the formula of universal law that has most explicitly the appearance of a prohibition: "I ought never to act otherwise than in such a way that I could also will that my maxim should become a universal law" (G 402). In Section II of the *Groundwork*, on the other hand, where he follows a more strictly philosophical method, which makes it possible to represent the moral law purely through concepts of reason rather than also through reference to the feeling of respect for the law, the various statements of the formula reveal more explicitly the moral law's fundamentally positive character. At the culmination of his discussion of the different formulas of the categorical imperative, for example, we find this purely positive statement of the formula of universal law: "act according to the maxim that can at the same time make itself a universal law" (G 436–437). And shortly thereafter, another purely positive version follows (the one that, as we noted earlier, makes explicit the self-related double universal validity of a practical law): "act according to a maxim that at the same time contains in itself its own universal validity for every rational being" (G 437–438).

7. *The Universalization Test and Means–End Reasoning*

Because Kant's universalization test is so often thought to involve prudential or instrumental reasoning, it bears emphasizing that when it is understood along the lines here set forth, as consisting in the comparison of a maxim's content with the form it presupposes, it is completely free of such reasoning, as the examples we have considered serve to illustrate. Interpreters have often supposed, for understandable reasons, that Kant is proposing a test in which one considers whether one could will to follow a certain maxim if doing so meant that this maxim would become a universal law of nature, understood solely as a law of heteronomous natural necessity, and they have supposed further that deciding whether one could will to do this is a matter of determining whether such action under this condition of universality would be prudent, or at

least compatible with the requirements of means–end rationality.²² If we attempt to understand the universalization test in this way, however, we inevitably run into problems, especially when we direct our attention to cases of individuals who occupy positions of advantage over others.

Thus, if we leave out of view that the law in question is a law of rational nature and simply conceive of it as a heteronomous law of natural necessity, then we face no difficulty—certainly no contradiction—when we attempt to conceive of everyone as naturally acting to limit or to impede others' outer freedom whenever they suppose that doing so will augment their own. As Mill remarks, "that a rule even of utter selfishness could not *possibly* be adopted by all rational beings—that there is any insuperable obstacle in the nature of things to its adoption—cannot be even plausibly maintained."²³ Many, indeed, have thought this

²²This is a general characterization, covering a variety of interpretive proposals differing significantly in their details. While it might at first seem that testing whether a maxim, under the condition of universality, is prudentially or instrumentally rational is not in general a matter of determining whether the attempt to will it as a universal law results in a *contradiction*, there is a prominent recent interpretation that takes the test to be precisely for prudential or instrumental consistency. Drawing on Kant's account of hypothetical imperatives, according to which willing to do whatever is indispensably necessary to achieve some end is analytically contained in the willing of that end (*G* 417), and taking this account to imply that willing the latter but failing to will something one recognizes to belong to the former would amount to a contradiction, this interpretation maintains that testing a maxim is a matter of determining whether, under the condition of universality, willing the action specified in the maxim is consistent with willing the maxim's end or with willing other ends necessarily tied to rational agency in general. This is often called (following Korsgaard) the "practical" or "practical contradiction" interpretation. For discussion of it, see O'Neill, "Consistency in Action," *Constructions of Reason: Explorations in Kant's Practical Philosophy* (Cambridge: Cambridge University Press, 1989), and Korsgaard, "Kant's Formula of Universal Law"; see also Herman, *The Practice of Moral Judgment*, 136–143; Wood, *Kant's Ethical Thought*, 87–90; and Andrews Reath, *Agency and Autonomy in Kant's Moral Theory* (Cambridge: Cambridge University Press, 2006), 211ff. Sometimes interpreters have supposed that in applying the test the question to be asked is whether a general pattern of such conduct as is specified in the maxim would be *generally* advantageous. This seems to be the opinion of Mill (*Utilitarianism*, chaps. 1, 5), and similar views can be found in Marcus G. Singer (*Generalization in Ethics* [New York: Knopf, 1961]) and R. M. Hare (*Moral Thinking* [Oxford: Oxford University Press, 1981]). But the remarks in Kant's text that have been thought to suggest that he is relying on means–end reasoning concern for the most part how the universal law would bear on the individual whose maxim is being considered, rather than how it would bear on the interests of society or of human persons generally. And even if the question is taken to concern general advantage, problems will arise similar to those described below.

²³*Utilitarianism*, chap. 5.

Hobbesian vision to be a true description of our actual world. Plato even represents it as the opinion of the majority. And if we proceed to ask whether, under this condition of universality, acting on this maxim would be prudent, we again have no difficulty in supposing that, while many of us would perhaps answer in the negative, certain individuals, cognizant of the advantage they enjoy on account of their natural endowments, good fortune, or industry, might reasonably believe that they would in all likelihood be well served were the world governed by such a law of nature. The same sort of problem arises when the test, so understood, is applied by a similarly advantaged individual reflecting on the maxim of indifference to the happiness of others. As Sidgwick observes, "a strong man, after balancing the chances of life, may easily think that he and such as he have more to gain, on the whole, by the general adoption of the egoistic maxim; benevolence being likely to bring them more trouble than profit."²⁴

Clearly this way of interpreting the test does not square with our recognition that the enjoyment of such advantages does not exempt a person from the duties of natural justice and beneficence. Nor consequently does it square with Kant's assertion that the different formulas are all formulas of the same law, given that the formula of humanity is in complete agreement with this recognition. And because it thus makes the answer to the question whether one can will these maxims as universal laws of nature dependent on contingencies lying outside the will, contingencies unrelated to the formal presuppositions of practical judgment, it does not square with the idea that has been guiding us throughout, the idea that the moral goodness or badness of an action is intrinsic in character.

If it is assumed that the universalization test must be interpreted along prudential lines, as described above, then it will likely seem that these difficulties are to be avoided by introducing a further condition to insulate the test from the influence of contingent advantages. It might be stipulated, for example, that the prudential or instrumental reasoning is to be carried out without reliance on beliefs one may have about

²⁴ *Methods of Ethics*, 7th ed. (London: Macmillan and Company, 1907), 389n. These difficulties have been raised by a long train of critics. An earlier version of essentially the same objection can be found in Schopenhauer, *On the Basis of Morality*, §7. And Kant himself makes the same point in criticizing the maxim *quod tibi non vis fieri, alteri ne feceris* (do not do to another what you do not want done to you) (G 430n).

how one compares with others in respect of such contingencies.²⁵ Introducing a restriction of this sort would be an effective way of avoiding the difficulties mentioned, and such a restriction seems, moreover, to reflect a genuine aspect of moral reflection. Indeed, in the application of the universalization test, as it has been interpreted here, to the cases we have considered, there has been a similar disregard of the contingent advantages one person may possess over another. For insofar as the practical judgments in which persons deem their own generically conceived happiness to be good are prior to, and so independent of, the specific judgments as to where their happiness lies and hence also independent of the empirical theoretical judgments of practicability on which all pursuit of happiness depends, all cognizance persons may have of the contingent advantages that they may enjoy over others or that others may enjoy over them are irrelevant to these fundamental judgments.

But to respond to these difficulties simply by introducing such a restriction would not touch the fundamental problem, for it would leave in place the assumption it has been our principal concern not to adopt, namely that the universalization test involves, albeit in an indirect or hypothetical way, prudential or instrumental reasoning. As was indicated earlier (§VI.5), our reason for avoiding this assumption is that it is tantamount to the supposition that the outcome of the test's application—the presence or absence of a contradiction in the will—depends on certain empirical theoretical judgments of practicability and the conditions they represent, all of which lie outside the will. This dependence remains even if, in carrying out the test, we ignore the contingent advantages that may be enjoyed by ourselves or by others and so rely only on judgments of practicability that would be made by a representative human person, who is normal or typical in the relevant respects. If we under-

²⁵ Such an approach has been suggested by Rawls, who draws on his idea of a “veil of ignorance” in his interpretation of the universalization test (“Themes in Kant’s Moral Philosophy,” in *Kant’s Transcendental Deductions: The Three Critiques and the “Opus postumum”*, ed. Eckart Förster [Stanford, Calif.: Stanford University Press, 1989], 85–86; cf. *A Theory of Justice* [Cambridge, Mass.: Harvard University Press, 1971], 252); Rawls does not himself introduce this idea in response to the specific difficulty we are considering here, but the attempt to use it for this purpose in connection with the maxim of indifference is discussed by Barbara Herman in her perceptive criticism of the prudential line of interpretation (see §I of her “Mutual Aid and Respect for Persons,” *The Practice of Moral Judgment*).

stand the universalization test's outcome as dependent on such external conditions, then we undermine our ability to see it as suitable for determining an action's intrinsic goodness or badness (see §V.5) and to that extent undermine our ability to understand the test in a way that fits with Kant's idea that what is distinctive about the moral goodness or badness of an action, what marks it off from goodness or badness of other sorts, is that it is intrinsic and so unconditioned in character.

To forestall a possible misunderstanding, however, it should be added that to interpret the universalization test as involving no instrumental reasoning is not to deny that the consequences of action are of any moral significance. The universalization test is carried out for the sole purpose of explicating the intrinsic, formal character of a possible exercise of the will and the action expressing it. All maxims that cannot be willed as universal laws are intrinsically wrong, which is to say that they are unconditionally prohibited by reason. In this regard they are all on a par. There is no more that a contradiction resulting from the application of this test can reveal; since contradiction is not a matter of degree, the universalization test provides no basis for comparative judgments regarding the gravity of wrongdoing (though as will be noted below, the contradictions in willing that the test can reveal are of two types). But this is not to say that in the practice of moral praise and blame we are to regard the gravity of one wrongful action as necessarily equal to that of any other. When we make judgments concerning merit or culpability, we view a person as a free (accountable) cause of an action and make our assessment in the light of the effects that ensue, taking into account the benefit or harm that results from the action.

These judgments of merit and culpability, however, are fundamentally different in character and function from the judgments of right and wrong that are based in the universalization test. The universalization test is to be carried out in advance of choice and action: it applies in the choice of maxims and articulates the moral reflection of the agent, who is deciding how to act. The assessment of merit or culpability, on the other hand, takes place after the decision and action have occurred: it is carried out by an observer or judge who considers the outwardly manifest deed or misdeed and what follows from it. Moreover—and this is of particular importance for our purposes here—in their consideration of an action's consequences, judgments of merit and culpability also *presuppose* judgments assessing whether the actions in question are *intrinsically* good

or bad, judgments that depend on the moral reflection made explicit in the universalization test. (Assessments of merit or culpability thus depend on the judgments of appraisal described earlier (§II.7).) If an action is intrinsically good (morally worthy, an expression of good willing), then only the good effects are to be considered,²⁶ and other things equal the degree of merit—the magnitude of praiseworthiness—is judged by the magnitude of that resulting benefit; similarly, if an action is intrinsically bad, only the bad effects are to come under review, and the degree of culpability—the magnitude of blameworthiness—is judged in the light of the magnitude of the harm (cf. *MS* 227–228). Thus, not only can there be no consideration of an action’s consequences in the determination of its intrinsic goodness or badness; we cannot even specify which consequences it is pertinent to consider in assigning praise or blame until this determination has been made. Consequences play a role in determining the magnitude of praise or blame, and they are often the basis of conjectures about which of these two—praise or blame—is in order; but which of them is truly appropriate is determined by the action’s intrinsic quality, and this does not depend on the consequences at all.

8. *Contradictions in Conception and in the Will*

If the foregoing account of the application of the formula of universal law is sound, it should fit with, and cast light on, Kant’s often discussed distinction between two ways in which contradictions result when we attempt to will the maxims of impermissible actions as universal laws (*G* 424).²⁷ Some actions, Kant tells us, are so constituted that their maxim cannot even be conceived as a universal law of nature; the attempt to do

²⁶ Only the good effects have a form of the sort that enables them to be intelligibly traced to the cause of the action. For this cause is the form of goodness itself, present in the agent, and the effects of this form can never be anything but good. Any bad effects of the action must be traced to misfortune, ignorance, or other impediments to that form’s efficacy. The original understanding of the good can no more be the cause of anything that is not good than can any bit of practical knowledge be the cause of anything other than the existence of its own object, though of course if impeded it may fail to realize that object, or succeed only in part.

²⁷ Prominent recent interpretations of the formula of universal law that include discussion of the two types of contradiction can be found in O’Neill, “Consistency in Action,” *Constructions of Reason*; Korsgaard, “Kant’s Formula of Universal Law,” *Creating the Kingdom of Ends*; Herman, *The Practice of Moral Judgment*, chaps. 6–7; and Wood, *Kant’s Ethical Thought*.

so results in a contradiction (what commentators often call a “contradiction in conception”). In the case of others, the maxim can indeed be conceived as such a law, but a contradiction results when we attempt to will it as such (a “contradiction in the will”). Kant claims that this distinction corresponds to the distinction between strict, narrower duty, or what we owe, and wider duty, whose fulfillment is meritorious—a distinction he seems in the *Groundwork* to regard as coinciding with the one he draws between perfect and imperfect duty. An action whose maxim cannot be conceived as a universal law of nature violates strict, or perfect, duty; one whose maxim can be conceived but not willed as such a law conflicts with wider, or imperfect, duty.

Since we have not yet investigated Kant’s explanation of how contradictions arise in the case of maxims that conflict with duties to oneself, nor examined in detail the difference between perfect and imperfect duties, we are not in a position to make a full assessment of this claim. But since the main examples we examined concern the fundamental duties of natural justice and beneficence, we have reason to expect them to be indicative of a general pattern. And if we compare them with regard to the distinction Kant describes, we find that the contradictions revealed in our discussion do differ in what seems to be the way he suggests. In the case of the maxim of assaulting others’ freedom, the contradiction can be traced to the fact that the practical judgment each person would make in adopting the maxim *conflicts* with the corresponding judgment each of the others would make; in the case of the maxim of indifference, it can be traced to the fact that the practical judgments *fail to agree* with one another.

Because it involves judgments that conflict, the contradiction brought to light in the first case is in the *object*, or *content*, of the will. Here there is no possibility even of conceiving what we are supposed to be attempting to will: what is to be willed (deemed good) is that each person’s outer freedom, or self-sufficiency in relation to others, be both extended and limited. So though there is no conflict in the content of the *maxim*, no incoherence, that is to say, in the practical thought of augmenting one’s own outer freedom through limiting others’, there is no possibility of thinking this content in accordance with the universality—specifically the *objective* universal validity—that belongs to the form of practical knowledge, with which maxims, as practical judgments, at least implicitly purport to be in agreement, and so no universal

object possible for practical judgment, no possible law available to be willed. In this case, the contradiction can be brought to light merely by considering the maxim with regard to practical cognition's *objective* universal validity.

In the case of the maxim of indifference, the contradiction revealed is not in the object willed, for in this instance nothing that is willed is incompatible with anything else that is willed. The contradiction is rather in the *willing* of the object: each person's happiness is both willed and not willed (both deemed good and not deemed good). So while there is no conflict in the content of the maxim, nor even in the universal object that results from thinking the maxim's content in accordance with practical cognition's form of universality, there is nevertheless conflict in the willing of that universal object. What is actually willed in the willing of the maxim's content falls short—indeed is *made* to fall short—of what must also be willed in accordance with practical cognition's form of universality: the maxim of indifference is a refusal to deem good something (others' happiness) that must also be recognized as good if the practical judgment in which this maxim consists is to have the form of practical knowledge. For only through this recognition can the judgment deeming one's own happiness to be good have an objective universal validity that is compatible with the subjective universal validity that it must also have if it is to be practical knowledge. It will be noted that to bring the contradiction to light in this case it is not enough merely to consider the *objective* universal validity of practical knowledge; consideration of its *subjective* universal validity is also necessary.

These observations suggest the following general contrast: Contradictions in conception constitute the violation of the *proscriptions* of practical reason: they occur in practical judgments that “go too far” in that, by regarding as good objects that necessarily cannot be known to be good, they transgress the limits defined by the formal condition of practical knowledge. Contradictions in the will constitute the failure to satisfy practical reason's *prescriptions*: they arise in judgments that “don't go far enough” in that they fail (or refuse) to regard as good objects that would necessarily, merely in virtue of the form of practical knowledge, also be recognized as good in any practical cognition of the objects deemed good in these judgments.

This way of understanding the distinction between the two types of contradiction is confirmed by the fact that it enables us to see a clear

congruence between this distinction and the corresponding distinction between the two ways in which one can exercise the will in a manner that conflicts with the formula of humanity, namely by employing humanity in persons merely as a means and by simply failing to regard it as an end. For in this latter distinction we have what is at bottom the same contrast between practical judgments that “go too far,” making shared practical knowledge impossible (no person can agree to being treated merely as a means), and practical judgments that “don’t go far enough” in that they do not contribute to making such knowledge actual (all persons must regard themselves as ends). And since the distinction between employing humanity merely as a means and merely failing to regard it as an end seems clearly to be in alignment with the distinction between the violation of perfect duty and the failure to fulfill imperfect duty, this congruence gives us good reason to expect that the distinction between the two types of contradiction, when they are understood as outlined above, does in fact, as Kant holds, line up with and reflect the distinction between perfect (owed) and imperfect (meritorious) duties.

The form of practical knowledge is the practical idea of an “omni-lateral agreement” in universal willing among all subjects sharing the capacity for such knowledge, in which “one and the same object” is shared by the will of each and every subject (*KpV* 28). Maxims that conflict with the objective condition under which such agreement is possible generate contradictions in conception; maxims that, being deficient on the subjective side, fail to contribute to its actuality, yield contradictions in the will.

9. Duties to Oneself

So far we have been concentrating on duties to others. But Kant holds that the categorical imperative also grounds duties to oneself. Two examples are offered, one involving a duty against committing suicide when life promises more trouble than enjoyment, the other an obligation to perfect oneself by cultivating one’s talents. In both instances Kant claims that contradictions result when we attempt to will as universal laws of nature the maxims of self-love opposed to these duties. Both examples also involve a certain *type* of duty to oneself, what he later refers to as duty to oneself “considered as an animal as well as a moral being,” in contrast to duty to oneself “considered merely as a moral being.”

Duties of the former sort differ from the latter in that they depend on an empirically determined concept of human nature, in which components of the human animal constitution are specified (*MS* 420, 444–445). Possibly Kant's decision to select as his examples duties that depend on this concept reflects his interest in using the idea of a law of nature in his illustrations of how duties can be derived from the formula of universal law. Whatever the explanation, this concept lies beyond our present concern, as do the associated questions concerning its specification and significance. Although Kant's reliance on it is closely bound up with the parts of his doctrine of duties to oneself that readers today usually find most disputable, the main question that this doctrine poses for our account of the formula of universal law arises at a more fundamental level. Since our focus is on the basic case, we shall concentrate on duties to oneself considered merely as a moral being, that is, as an individual person, a human subject with the capacity for practical cognition. And as in our interpretation of the cases involving duties to others, we shall of course be interested in looking through the notion of a law of nature to the underlying idea of a practical law that it expresses (cf. §VI.4).

But first we must set clearly in view the respect in which, from the standpoint of our concern here, Kant's sample derivations differ most significantly from the explanations offered regarding duties to others. In the latter cases, contradictions resulted because asymmetries in the maxims' representation of relations between oneself and others came into conflict with themselves when subjected to the condition of universality; here, however, the contradictions Kant describes do not appear to exploit the idea of universality at all. In neither instance does the contradiction seem to turn specifically on the impossibility of supposing that *everyone* is to act on the maxim or that *everyone* can agree that everyone is to act on it. In the case of the maxim of suicide, for instance, the contradiction is said to arise in the attempt to conceive of a law "to destroy life through the very sensation whose function it is to impel to the furtherance of life" (*G* 422). To appreciate the conceptual conflict Kant is attempting to describe here, it does not seem necessary to regard the persons subject to this law as standing in some particular relation to other persons, nor even to suppose that they share a practical world, in which mutual recognition and influence are possible. Of course, this difference comes as no surprise; the duties in question are, after all, duties

to oneself. It does, however, raise the question whether such duties can be accommodated by the interpretation set forth above, given the central role the idea of universality plays in its explication of the form of practical knowledge.

At this point, it might be proposed that we can perhaps find the idea of universality to be at work in Kant's examples if we look for it in another place. Earlier on, it was suggested that under the fundamental practical law there is room for a division of a priori laws according as they concern persons in relation to one another or in relation to themselves. The former would lie in conditions of intersubjective validity, or validity whose universality reaches across the wills of different persons, the latter in conditions of intrasubjective validity, where the universality extends across the occasions of exercise for the will of one and the same individual (§V.3). This suggestion might encourage the thought that subjecting the will's exercise to the latter condition is a matter of, say, considering whether one's willing in a given case would be valid on all occasions. We have seen that the validity of cognition extends across occasions as well as across judging subjects; perhaps duties to others are determined by reference to validity of the latter sort and duties to oneself by reference to the former.

Such a proposal does not seem viable, however. It would have the effect of resolving the idea of a duty *persons have to themselves*, if not into the idea of a duty their judgments have to one another, then at best into the idea of a duty persons-on-occasions have to themselves-on-other-occasions. But neither of these ideas seems to capture the idea of a duty or obligation that I, as a particular person, have simply to myself. Just as in the case of duties to others it is the capacity for practical knowledge *in others* that binds me, so in the case of a duty to myself it must be this capacity *in myself* that binds me, not its *exercise* in some other judgment, nor even this-capacity-in-me-on-other-occasions. Nor does the proposal fit comfortably with Kant's thought that what the formula of universal law immediately governs are maxims. As we have noted, the exercise of the will in a particular person begins with wish, a practical judgment in which an end is represented as good, and this representation of an end is a practical principle, or maxim. So willing always begins with judgments that have a universality extending across a particular person's different occasions for judgment. Hence, the presupposition of universality does not in this regard represent a condition with which a maxim is not

already in agreement. To that extent, maxims already bear the form of universal law, in the sense that universality already figures in their content. We should not, therefore, expect that bringing the form of practical law to bear on one's willing in relation to oneself is a matter of, say, considering whether one can will one's maxim on all occasions, for in having a maxim, or principle, at all, one is already supposing that this is what one is up to. These considerations confirm what Kant's two examples appear in any case to suggest, that duties to oneself do not lend themselves to elucidation through the idea of universality. Such duties may therefore seem not to be amenable to the interpretation of the formula of universal law through the idea of the form of practical knowledge.

One way of reacting to this disparity would be to conclude that Kant is mistaken in thinking it possible to derive duties to oneself from the formula of universal law. Certainly the specific arguments he presents in the two examples have attracted their share of criticism. Some would go further, questioning whether the idea of a duty to oneself is ultimately coherent.²⁸ Kant himself later acknowledges that the concept seems at first glance to contain a contradiction (*MS* 417). Doubts about the coherence of the concept arise most easily when duty and obligation are conceived in a narrow legal sense, as they are in the doctrine of right. In this sense, obligation presupposes some contractual agreement, entailing that wherever one person is bound, or obligated, another obligates, or holds a right against the first. Since a contract is voluntary, it is always possible for one party to relinquish a right deriving from it, releasing the other from the corresponding obligation. Lenders, for instance, are free to forgive their borrowers' debts. On this conception of obligation, a duty to oneself would have to be an obligation from which one would always have the freedom to release oneself, a freedom that seems flatly incompatible with being bound. It is clear, however, that Kant is not relying on any such narrow conception of obligation. As we have seen, the practical laws of natural justice and beneficence do not depend on any actual contract or agreement among persons; on the contrary they are themselves presupposed in all such agreements, being

²⁸ M. G. Singer, for example, in "On Duties to Oneself," *Ethics* 69 (1959): 202–205. Cf. Warner Wick, "More about Duties to Oneself," *Ethics* 70 (1960): 158–163.

formal conditions of all validity in willing. The duties under consideration in the examples are not legal, or juridical, but in every case ethical.

Alternatively, we might allow that we have duties to ourselves, but assign these requirements to some other region of morality than the part to which the laws of natural justice and beneficence belong. Of course, Kant does something of the sort himself in deploying the distinction between duties to oneself and duties to others as a principal division in the organization of his doctrine of ethical duties (*MS* 412, 442). But the question would be whether duties on both sides of this divide can be referred to a common principle or formula, as Kant supposes, or whether his attempt to account for morality in its entirety through the idea of universal self-legislation as expressed in the three formulas is better regarded as an account of morality in a narrower sense, comprising the obligations persons have to one another.²⁹ That there is no such common principle underwriting both duties to oneself and obligations to others is often taken for granted, if not outright asserted, in contemporary moral philosophy. In current treatments of Kant's ethics, moreover, obligations to others tend to receive by far the greater bulk of attention, as they do here.³⁰ This tendency, like the common preference for the formula of humanity over the formula of universal law, seems to be related to the considerable recent convergence between interpretive work on Kant's moral philosophy and the development of contractualist accounts of moral obligation.³¹ Such accounts, part of a broader modern trend toward social conceptions of morality, are usually not closely affiliated with classical conceptions of ethics as concerned with how one should live and with the care of the soul, views more congenial to the idea of duty to oneself. For approaches inspired by the idea of a contract, the aspect of practical cognition that has particular salience is its subjective universal validity, whereby different persons are able to reach agreement in practical judgments—an aspect that is highlighted in the formula

²⁹ See T. M. Scanlon's useful discussion of broader and narrower senses of "morality" in *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998), 6–7, 171–177.

³⁰ There are, to be sure, notable exceptions, such as Thomas E. Hill's discussion of self-respect and related topics in *Autonomy and Self-Respect* (Cambridge: Cambridge University Press, 1991).

³¹ A good recent example is Stephen Darwall's *The Second Person Standpoint: Morality, Respect, and Accountability* (Cambridge, Mass.: Harvard University Press, 2006).

of humanity, but plays no apparent constraining role in Kant's derivation of duties to oneself. If the prominence Kant gives to the formula of universal law reflects his connection with older, pre-modern traditions in ethics, so certainly does his attempt to find a systematic relation between duties to self and duties to others. And just as it would be no small misfortune should these older questions and concerns slip from view in philosophical reflection, so it would be a welcome reaffirmation of their significance if some systematic relation between the two sorts of duty could be recovered that would help us to understand them again. For our purposes here, it will be enough if we can trace the outlines of a unified account by describing in general terms how the form of practical knowledge grounds duties to oneself.

We have concluded that it cannot be through a condition of *universality* that the form of practical knowledge and of practical law gives rise to duties to oneself. Only duties to others are derived in this way. It does not follow, however, that this form cannot ground duties to oneself. Such a result may seem to be entailed if we think of the formula of universal law as providing a "universalization test," but the logical idea of universality is not by itself enough to furnish a window on the formula's full significance. The formula of universal law is at bottom a formula of *law*. In expressing the form of practical knowledge, it expresses, as we have seen, the form of practical law, in accordance with which maxims are to be framed. Universality is only one of the marks of law, and not the most primitive. Necessity is another. As we saw at the very beginning of our investigation of judgment, rational cognition is never arbitrary. Being conscious of itself as self-sustaining, it is aware of itself as involving "something of necessity." It was by reflecting on cognition's self-consciously self-sustaining character, its validity, that we were able to bring to light the two respects in which this validity is universal and to incorporate them in our account of the form of practical knowledge and of practical law. We therefore need to consider the possibility that, although on the one hand, as it bears on persons in relation to one another, the form of practical law subjects maxims to the necessity of law through the representation of universality across persons, yet on the other hand, as it bears on persons in relation to themselves, it subjects maxims to the necessity of law in a more immediate way than through a representation of such universality. We should be able to make out such a possibility if exercising one's will necessarily begins with a con-

ception of oneself as a particular person, and if in that conception the self-sustaining form of practical cognition is already presupposed.

In our investigation of judgment, we argued in effect that the form of practical knowledge lies originally in the validity, or self-consciously self-sustaining character, of practical judgment. And by spelling out the self-related double universality of this validity and the possibility it implies of necessarily coinciding relations of subjective as well as objective universal agreement among judgments—relations expressed in the presupposition of universality—we saw how duties to others arise. But we also noted that the validity of practical cognition entails not only agreement *among* judgments, but also an original relation of positive agreement among the elements *within* judgments. In order for diverse judgments to agree with one another, each must agree with itself, and hence must involve a consciousness of the self-sustaining relations among its constituents. In accordance with the form of rational cognition, however, these constituents are related in a certain order, in which the concept of the subject has primacy. As rational cognition, practical knowledge is knowledge “from concepts” in the sense discussed earlier (§IV.5), so a practical judgment must start with the concept of its subject and proceed to determine it by attaching to it a practical predicate. In the cases of interest to us here—those of wish and choice—the subject of the judgment is just the judging subject itself. So here the presupposed concept of the subject of the judgment is the self-conception of the judging subject. And in the judgment this presupposed concept is not changed, but only determined (to make a judgment is not to change one’s mind), so it is necessarily self-sustaining, as indeed it must be if the judgment itself is to be the self-sustaining practical cognition it understands itself to be. The act of attaching the predicate—the act of practical judgment proper—must therefore be in agreement, positively as well as negatively, with the presupposed conception of the subject. It must not conflict with the subject’s original act of self-conception, and beyond that it must further it. In this way, there arise both negative and positive duties to oneself, duties that govern all of a particular person’s acts of practical judgment, even the primitive act of making happiness an end, the act that first constitutes a particular person.

Admittedly, the suggestion that these duties bear even on this primitive act may on first consideration seem puzzling, or even inconsistent. If this act first constitutes a particular person, how can there be any

presupposed self-conception of the subject distinct from the conception reached through the act? As we said, this act is as much a judgment that makes a person as a judgment a person makes. Moreover, if this act is what first constitutes a particular person, then how could it furnish a conception of the subject that could constrain or bind that very act? We appear to face a variant of the problem described earlier, that the concept of a duty to oneself seems at first glance to contain a contradiction. Indeed, it is just this sort of consideration that leads Kant to say that it would be a contradiction to suppose that we have a duty to make our own happiness an end (*MS* 386, 388). The duties to self just described may thus seem to collapse into mere prudence, or the requirement on a particular person that its choices be in agreement with the wish for happiness that constitutes its identity as a particular person (§V.5).

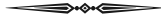
It is crucial, however, to keep in view that even though the primitive wish for happiness first constitutes a person, it is still a practical judgment and hence is structured and conditioned in the manner constitutive of practical rational cognition. This wish cannot take place in a vacuum, but is possible only in a certain material condition of receptivity, namely the subject's feeling of pleasure or enjoyment in the agreeable, the feeling on which sensible desire is founded (§III.4). Thus the act of judgment that first constitutes a person presupposes not only the self-conception of the judging subject, but specifically a self-conception that includes awareness of this sensible condition, distinguishing this particular practical subject from every other. Hence even the primitive act of wishing for happiness that first constitutes a particular person presupposes an indeterminate, or yet to be determined, conception of oneself as a particular practical subject capable of determining oneself in a practical judgment, and this presupposed conception accordingly belongs to the form of every maxim. Since this capacity to determine oneself in a practical judgment is just the capacity to be a person, we can call it *personality* in Kant's technical sense; alternatively, since it is the capacity for practical cognition *in* sensible conditions (in a word, rationality and animality), it could be identified with *humanity* in one's own person (cf. *KpV* 87).

In sum: Because the form of practical cognition is originally grasped in practical judgment's consciousness of itself as a self-sustaining and self-determining activity, the original conception of the particular subject already presupposed even in the fundamental practical judgment first constituting a particular person—the wish for happiness—must be

itself a self-sustaining conception of oneself as a subject with the capacity for such self-sustaining and self-determining practical cognitive activity. The form of practical cognition is therefore already presupposed in this original indeterminate self-conception, which itself belongs to the form of every maxim. Thus to agree with this subject negatively and positively is, respectively, to practically judge, or to will, in a way that preserves and furthers its self-sustaining activity of practical cognizing. In this way there arise obligations to oneself, both a negative duty to preserve one's own existence as a particular person by not infringing one's self-determining practical cognitive activity, and also a positive duty to further one's own existence by strengthening and extending that activity.

Since one's self-conception as a particular person already contains the idea of the form of practical knowledge, there is a sense in which duties to oneself enfold the other duties. The negative duty to oneself prohibits acting against oneself as a particular person, but to respect this prohibition is to respect the general prohibition against infringing the outer freedom of others. Similarly, the positive duty to oneself includes the injunction to further one's own practical cognitive power through the cultivation of virtue, but to do this is to observe the positive duties to others contained under the heading of beneficence. Thus obligations to others are encompassed in the duties to oneself, and just as a particular person grasps the form of practical knowledge only in its own self-conception, so it is only through observing duties to oneself that it is possible to observe duties to others (cf. *MS* 417–418). We saw earlier that the fundamental form of the will is a principle of self-agreement, in that it is the source of a particular will's agreement with the form it shares with all such wills. We now see that through duties to oneself it is also the source of a particular will's agreement with itself as a particular will. Just as the form of practical knowledge, as the form of practical law, is a law for particular wills in general, so this form *in* a particular subject is a law just for that subject. The point can also be expressed in the terms of the formula of humanity by saying that duty in general requires agreement with humanity, or rational nature, while duty to oneself requires agreement with humanity in one's own person. And we can draw the parallel in terms of the third formula's notion of autonomy by observing that the proposition that persons are a law to themselves is valid in a distributive as well as a collective sense.

VIII



Conclusion

1. The principal aim of this study has been to develop an account of the categorical imperative that elucidates its basis in practical reason in a way that also clarifies how it constitutes a substantive constraint on the will, the same constraint in each of its three formulations. In our pursuit of this aim we have stepped back from the details of Kant's argument in order to take a broader interpretive approach, in which we have elaborated his idea that practical reason is the capacity for practical knowledge, or the capacity to know the good, and his related thought that morality lies in a specific type of practical knowledge, the knowledge of *intrinsic* goodness. Drawing on these ideas, we have explicated Kant's conception of the will as practical reason by elaborating an account of practical judgment, spelling out such judgments' formal presuppositions as well as their involvement in choice and wish. This account has guided both our explanation of how the categorical imperative, and in particular the formula of universal law, is based in practical reason, and our illustration of how, through this formula's application, substantive constraints in the form of familiar duties come to light.

In following this broader approach, we have outlined a positive interpretation of the formula of universal law, one that aims to remove the false appearance it may initially have of being merely an external imposition of a logical operation with no discernible license to serve as an unconditional constraint upon our will. In working out this interpretation, we have developed the idea that the function of the universalization test is to make manifest a common form that already lies present in the will's maxims, and we have traced this form to a presupposition maxims in-

volve on account of being practical judgments. This interpretation was initially proposed because it identifies a specific conception of internal inconsistency—a conception of conflict between a maxim's form and its content, rather than within its content alone—that enables us to make sense of Kant's talk of intrinsic badness or wrongness and thereby affords us a prospect of understanding his idea that such badness is distinctive of morally impermissible action and his related suggestion that it is the intrinsic character of this badness that distinguishes it from the sort of badness or wrongness characteristic of action that is imprudent or contrary to a hypothetical imperative. With the outlines of this interpretation now in place, we return to the idea on which it is based—the idea that intrinsic wrongness lies in internal conflict between form and content—for a few concluding comments regarding the implications it has for our understanding of the relation in which persons, as self-constituting practical subjects, stand to morality.

2. As we have seen (§V.2), if we understand the intrinsic character of the wrongness of morally impermissible action through the idea of conflict between form and content, and if we also recognize that the form of willing is necessarily constant and unchangeable across all exercise of the will and so something that belongs to the nature of the will itself, or to what, again on account of this necessity, Kant calls “our true self” (*G* 458, 461), then we can see the categorical imperative as lying in the requirement that the will's content be in (nonaccidental) agreement with its constant form, and we can describe this imperative, accordingly, as a principle of self-agreement, unity, or identity. Therefore, to the extent that morality is based in such a principle, we have a way of understanding how morality, rather than being a threat to a person's integrity, as some have recently suggested,¹ is in fact, as was traditionally thought, its necessary condition. Since this self-agreement is requisite for the unity of a person's self-constituted practical self-conception, a unity that is implied by the idea of character as well as by that of integrity, the same point can be seen in Kant's assertion that morality is the condition, the *sine qua non*, of having a character at all (*KpV* 152; cf. *Anth* 292, 329).

¹This concern has been raised perhaps most prominently by Bernard Williams. See his “Persons, Character, and Morality,” in *Moral Luck* (Cambridge: Cambridge University Press, 1981).

Kant also maintains, as we have seen (§III.7), that we are not in a position to rule out the possibility of prudentially rational beings without the least inkling of the possibility of such a thing as the unconditionally commanding moral law. Such beings would not make practical judgments determining the good, so the presuppositions of such judgments would not figure in their practical thinking, nor accordingly would the form of willing that we have traced to such presuppositions. For such beings, morality would not be a condition of integrity, or of unity in their practical self-conceptions, and the only way in which the form expressed in the universalization test could constrain their conduct would be through being imposed from without, never through self-constraint. But for beings such as ourselves, in whom the same moral law is inwardly manifested in the feeling of respect and outwardly expressed in the real relations of agreement in practical judgment that constitute and shape our practical lives, there is no way of maintaining unity of practical self-conception and so no way of having such a thing as integrity or character except through maxims that are in agreement with this law, which constitutes our rational nature, or humanity.

3. The understanding of our relation to morality that flows from a conception of the categorical imperative as a principle of self-agreement is profoundly different from the characterizations of this relation that are commonly implied or taken for granted in much of our discourse about morality. For a variety of reasons, which are difficult to identify with precision, but which evidently trace in part to our natural constitution and in part to the development of human culture, the attempts we make to articulate the relation in which we stand to morality frequently involve modes of expression that suggest that morality is somehow external to us. Even to speak of our "relation" to morality can intimate as much. Often this suggestion of externality is conveyed by our speaking of a difference in "point of view," or by our drawing some similar contrast, usually one that introduces the thought of different interests. Thus, "the point of view of the individual" is often distinguished from "the moral point of view"; the "standpoint" of this or that person or group is contrasted with morality's "impartial standpoint"; and the private, the subjective, and the inner are set in opposition to the public, the objective, and the outer. Implicit throughout is the thought of a contrast or opposition between a person's own interests and the interests of morality—or rather, since there can be no interests where there is no subject, the interests of the community, of

society, or perhaps of some dominant group within it (“another’s good,” as Thrasymachus said).

These contrasts obviously have their place and their point. Indeed, they are ultimately expressions of the awareness of conflict that, according to Kant, figures in our consciousness of obligation and reveals our will not to be constituted in such a way that its own nature guarantees a necessary agreement between its exercise and the moral law (cf. §V.4). But they can have the cumulative effect of steering philosophical reflection toward a misconception of the nature of our relation to morality. To speak of the moral point of view, for example, is inevitably, on account of the very idea of a point of view, to suggest that this is but one among many different points of view and therefore to that extent optional, something we can occupy only if we “adopt” it,² rather than something that we naturally inhabit (and that may with greater justice be said in a certain sense to have already long since adopted *us*, as particular persons). It may also reflect the tacit presumption that, like different positions in space, these different points of view are all fundamentally on a par in the sense that none of them is naturally prior to the others (and in particular that none of them is uniquely presupposed by all the others), even if it is supposed that special arguments may be offered to show that the adoption of the moral point of view, or of a point of view that resembles it, may be advantageous to persons who in the first instance occupy another. It thus invites the question Why should I, as an individual with a point of view of my own, which I normally and in the first instance occupy, adopt the point of view of morality?

4. No philosopher appreciates more clearly than Kant does that if we lapse into a position from which this is the question we are posing for ourselves, there is no hope of providing a satisfactory answer. But characterizing the application of the moral law as involving a “universalization test” can easily reinforce the familiar tendency to view morality as external in the way this question assumes, for it can suggest that Kant, too, takes moral reflection to involve a stepping outside of one’s

²Hume, for instance, says that to engage in moral thought and discourse a man must “depart from his private and particular situation, and must choose a point of view, common to him with others” (*An Enquiry Concerning the Principles of Morals* § IX, Pt. I). Though in speaking of a “common point of view” Hume insightfully avoids depicting morality as external, and though he rightly regards this point of view as naturally available, he does not see the choosing of it as a retaining or a recovery of a position originally occupied.

individual standpoint and a movement from the private and the particular (one's maxim) to the public and the universal (law). It is therefore important to emphasize, in conclusion, that, according to the interpretation developed here, this view of the moral reflection involved in the moral law's application, though perhaps true in a certain superficial sense, is at a deeper level just the opposite of Kant's fundamental understanding of it. If there is any "movement" involved in this moral reflection, it is of the sort that the term "reflection" itself suggests—that of *return*, the turning back in thought to the original position of practical knowledge in an acknowledgment of what is already implicitly present in all maxims and practical judgments. This reflection, Kant holds, is already at work in an obscure and possibly on occasion confused way in ordinary moral consciousness, as the latter is exercised in the particular cases of practical life, but independently of the abstract articulation of its workings and of its principle (in "formulas") that it is the business of philosophical reflection to furnish (see *G* 403, *KpV* 8n).³ Once such articulation is provided, it is possible to speak of a "test," a "method," or a "procedure" whereby this principle is applied in moral judgment, and to elaborate this application in explicit detail, thus reenacting and recovering in abstract philosophical thought ordinary moral reflection's return to the original position of practical knowledge, but now with a clear and explicit recognition of the form that constitutes that position—the form that is already implicitly present in all maxims and practical judgments and that always to some degree attracts the attention of human practical knowledge in virtue of the latter's inherent self-consciousness. Here we may recall the notice Aristotle takes of the value of Plato's question "Are we on the way from or to the first principles?"⁴ In carrying out the universalization test we are *explicitly* on the way back *to* the first principles of practical judgment (that is, the form of such judgment), but in practical judgment itself we are always *implicitly* on our way *from* those same principles. The good will's reflective act of *willing a universal law through its maxim* retraces, explicitly and analytically, the steps of its original implicit and synthetic act of *willing its maxim through a universal law*.

³This is not to say that this consciousness cannot in such concrete cases find articulation in the familiar expressions of everyday moral discourse ("What if everyone did that?", "Put yourself in her shoes," etc.).

⁴*Nicomachean Ethics* 1095a30–b3.

Yet at the same time there is a sense in which in practical judgment we never *depart* from those principles. Practical judgment itself, essentially, or in respect of its form, its understanding of itself as self-sustaining, constitutes an absolute position, originally and necessarily shared by all persons capable of practical judgment in all their exercise of this capacity, and it is this that we seek to recover in an explicit and consistent way when, as we say, we “adopt the moral point of view.” **Anything that can be described as the particular point of view of this or that individual person is always reached from the original, absolute position of judgment, by restriction or limitation of one sort or another—a limitation that is either legitimate, through the introduction of content in judgments determined by (or at least in agreement with) the form of knowledge, or else intrinsically wrong or deficient, when the content conflicts with the form or at least fails to agree with it.** But insofar as we engage in practical judgment at all, we never *abandon* this original position, or wholly depart from it, for it is only by occupying it that practical judgments themselves and hence agreement or even conflict among them are possible. Morality, as the knowledge of intrinsic goodness, which constitutes the original form of all of our knowledge of the good, is external to us only in the way that space is external to us; it would be better to say that we inhabit it from the beginning, as the space of practical reason in which, as persons, we together carry out our pursuit of the good.

To the extent that we lose sight of this internal relation to morality, as the original form of our knowledge of the good, we face the question mentioned earlier: Why should I depart from my own point of view, which I normally and in the first instance occupy, in order to adopt the quite different point of view of morality? Since this question implicitly associates a different “good” or interest with each of the distinguished points of view, the very posing of it betrays that the capacity to gain reflective access to our implicit original understanding of practical reason as the capacity to know the good is being impeded by other factors, such as the feelings of conflict that can figure in the consciousness of obligation owing to the presence of opposing inclinations, or our experience of seemingly intractable practical disagreements, or perhaps certain theoretical convictions that in one way or another seem to us to preclude the possibility of knowledge of the good. But if reflective access to this original understanding of practical reason is thus obstructed, then it will

almost inevitably appear that reason itself, under the name of prudence, endorses the opinion of the many who, according to Plato, prefer the reputation for justice and morality over the genuine possession of these as virtues, even though, like everyone else, they prize reality over appearance when it comes to the good itself. Impediments to our recovery of the original understanding of practical reason can accordingly draw philosophical reflection toward a conception of morality as standing in an external relation to us and to our wills, and toward a corresponding conception of reason as inevitably favoring the appearance of morality and virtue over the reality. If, therefore, moral philosophy is to contribute in any way at all to the aim to which Kant assigns it, namely “to secure acceptance and durability” for morality’s principle—if it is to be, in Plato’s image, a “guardian” of justice and morality themselves—then its principal business must be the difficult one of articulating the internal relation we have here been attempting to identify and to describe by investigating the form of our rational knowledge of the good. It is only when morality is so understood that our natural interest in knowing the good itself can recognize itself as having morality as its original form.



Epilogue: Kant's Idea of the Practical Purpose of Moral Philosophy

PERHAPS it will seem surprising to some readers to find Kant's ethics portrayed as it has been in these pages. There are so many often-cited pronouncements in his writings, as well as aspects of tone and manner of expression, that seem to suggest that Kant sees morality as quite unconnected with, or even opposed to, happiness and the good, that it may still be wondered how the interpretation outlined here could emerge from those texts. This is not the place to take up for individual consideration the particular passages that have often been thought to reflect this or that Kantian dualism. It may, however, be appropriate to add a few general remarks on Kant's understanding of the practical purpose of moral philosophy, and also on his view of the audience for which he was writing. For there are some general facts about his purpose and the conditions in which he endeavored to carry it out that are easy to overlook, and without a due appreciation of them, the manner in which he presents his ethical doctrines and the rhetoric and tone he employs in expressing them are likely to give us a severely distorted impression of their substance.

As Kant understands it, the practical need for moral philosophy arises out of a "natural dialectic" in which ordinary human reason in its practical use is liable to become entangled. But this need does not so much stem from the fact that motives of self-interest can directly lead persons to ways of *acting* that infringe one another's outer freedom, for the primary and most proper remedy for such action lies not in philosophy but in a public system of coercion subjecting the exercise of outer freedom to law. The need arises rather because these same motives can

by themselves leave us naturally susceptible to fall into ways of practical *thinking* that cast doubt on the legitimacy, or at least on the strictness and purity, of moral imperatives and can further, when supported by certain overweening theoretical philosophical speculations about nature and our relation to it, give rise to certain ways of practical *philosophical thinking* that call into question the idea of the unconditioned goodness of a good will or the possibility of human freedom and thereby the very reality of the imperatives of morality (*G* 394–396, 404–405, *MS* 378). It is when self-interest and the ambitions of theoretical philosophical speculation thus converge or even join in alliance, threatening to weaken or to hinder the acceptance of the moral principle, that the practical need for moral philosophy properly and most pressingly arises. For even where, in common (pre-philosophical) moral consciousness, the motives of self-interest are, as in a virtuous character, properly subordinate to the motive of morality in a way that would be sufficient for everyday conduct in a generally virtuous though not culturally or philosophically sophisticated society, such an order of motive is present in a way that does not involve an abstract, philosophically articulate self-understanding, and therefore by itself it is neither prepared for the threat just described nor consequently capable of warding it off when it arises.

What is needed in the face of such a threat is for the rational knowledge in which ordinary morality consists to become philosophically reflective and to provide for itself a philosophical account of morality in which the proper order of motives that is present only obscurely and unreflectively in philosophically innocent ordinary virtue is articulated in philosophical reflection. This account must involve an analytic and a synthetic stage: First, the moral principle must be clearly distinguished from all other motives—which can be assembled under the general heading of happiness—and displayed in its separateness. But this must be done in a way that does not make it impossible to carry out the task of the second stage, which is to show that the moral principle is nevertheless coherently related to the ends that make up happiness, rather than simply separate from or even fundamentally opposed to them. If successful, these stages will together contribute to moral philosophy's aim of securing "acceptance and durability" for the moral principle, and they will do so in the two ways that moral philosophers have always had available and have always employed to foster the moral motive—namely by expounding the true nobility, or greatness of soul, that virtue and virtue

alone involves, and by tracing the connection this motive of virtue has with the human good. Thus, the first stage will contribute to this aim in that it will help enliven a person's interest in freedom and autonomy by confirming the presumption of ordinary morality that the moral principle is the source of all genuine greatness of soul. The second will contribute by identifying the place of morality and virtue in relation to the good (and ultimately to the highest good), thus forestalling doubts about whether the practice of morality and virtue, being so separate from all other motives, has a coherent relation to happiness, an end a person necessarily deems good. Through being displayed in this way in their essential, constitutive relation to the good of happiness as well as in their separateness from happiness, morality and virtue can be seen as attractive and good as well as admirable and sublime, just as the good will was depicted, at the outset of Kant's ethics, as not only the sole unconditioned and intrinsic good but also the only thing whose effects are necessarily beneficial.

Kant seems to have thought that what needed special emphasis was the first stage—the display of the moral law and the motive of duty in their purity. Noting the prevalence in the popular philosophy of his own day of a certain eudaimonistic outlook and also a certain tendency to enthusiasm and moral fanaticism, both of which contributed to confusion at the level of reflection regarding the proper order of motivation in a virtuous character, Kant seems to have thought, reasonably enough, that in view of these tendencies it was his analysis that should be given the most prominent elaboration and emphasis, and so he endeavored repeatedly to bring clearly to light the heterogeneity of the motives of morality and happiness. It can hardly be denied, after all, that a sound understanding of the relation between these elements is impossible if the difference between them is not properly grasped. But this considerable emphasis he placed on the analysis entailed that his synthesis, though in fact elaborated in extensive detail and with great insight in his doctrine of the highest good, was given less prominence, if not in his presentation of his ethical doctrines as a whole, at least in certain parts of his writings on ethics, including those parts, such as Section I of the *Groundwork*, that have in recent times chiefly attracted philosophical interest, particularly in anglophone scholarship. As a result, his repeated insistence on the contrast between duty and inclination, on the heterogeneity of morality and happiness, and so forth, though entirely sound in itself, has had the

effect of making it seem to many today as though his treatment of morality is too austere and hostile to happiness and the human good.

The principal difficulty, then, that we face today when we attempt to appreciate Kant's moral philosophy is to understand how the elements his analysis so radically separates from one another can also stand in a proper synthetic unity—how the “anatomy” (even if the primary element it reveals proves to be awe-inspiring rather than hideous) can be compatible with an understanding of the elements in a systematic whole that we can recognize to be good and so regard with satisfaction. The concern in this study has been to find a way of conceiving of the analysis that will make intelligible how the synthesis is nevertheless possible, and even necessary. And the guiding idea has been that of viewing morality as the knowledge of intrinsic goodness, or—what comes to the same—as the form of knowledge of the good in general. Through this idea, an understanding of both the difference and the relation among the elements is simultaneously achieved. Morality's formal character separates it out radically from all the content in our knowledge of the good, but precisely because it is the form of that knowledge, it is also related to our interest in knowing and achieving the good in such a way that morality, far from being unconnected with or opposed to that good, is rather its original form.



Index

- Action, 38, 47–48, 114; means and ends in, 41, 74n8, 85, 199–200n11; types of, 48
- Agreeable, the, 38, 75–77, 82–83; sensible desire and, 37–38, 70–71; the good and, 69–70, 70–78, 80, 82–84; judgments of, 70–71, 98; pleasure in, 70–71, 76–77, 78, 82
- Agreement: in philosophy, 1–2; concept of good and concept of, 12, 59; between content and form, 19, 127–128, 131–134, 144, 159, 220, 221–222, 229–230; rationality and, 35; knowledge and, 39, 59, 77, 104–105, 107–108, 111; communicability and, 58, 204–205; practical, 60–61, 65, 198–199; in an end, 75, 77–78, 172; judgments and, 104–105, 107–108, 111; subjective and objective forms of, 111–115; universality and, 112–115; subjective and objective universal validity and, 115–116; among practically cognizing subjects or persons, 154, 231; humanity and, 171–172, 173; as joint act (in a contract), 198–199, 200–201; capacity to communicate and, 200–201; among elements within judgments, 237. *See also* Self-agreement
- Akrasia. *See* Weakness of will
- Allison, Henry E., 5n3, 33n9, 141n4, 174n9
- Anscombe, G. E. M., 11n5, 55n27
- Appraisal, judgments of, 56–58, 228.
See also Attitudes, practical
- Approval, judgments of, 56–58, 62, 76.
See also Appraisal, judgments of
- Aquinas, Thomas, vii, ix, 25n1, 26n2, 28n4, 117n11
- Aristotle, vii, 25n1, 26, 117n11, 141n4, 162, 245; on cleverness, 34, 40, 55
- Assaulting others' freedom, 189; maxim of, 190–196, 203, 205, 207–208, 209, 229–230. *See also* Contradiction
- Attitudes, practical, 62–63. *See also* Appraisal, judgments of
- Aune, Bruce, 5n3
- Austin, J. L., 45n18
- Autonomy, ix, 136, 150, 151, 162, 163, 178; practical knowledge and, 134–136, 149–150, 151; practical law and, 134–136; duty and, 154n1. *See also* Formula of Autonomy; Heteronomy
- Bad(ness), 54; intrinsic, 10, 15–16, 164–165, 166–167, 226–228, 242; concept of, 12, 44, 113, 178–179; as content of negative practical judgment, 51–53; practical knowledge and, 53–54. *See also* Good(ness); Wrongness
- Baumgarten, Alexander, ix
- Belief, 67n1
- Beneficence: law of, 138–140, 234–235; duty of, 184, 209–215, 225. *See also* Indifference to others' happiness, maxim of
- Blame. *See* Praise
- Capacity to communicate, 115, 123, 200–201; discursive knowledge and, 115,

- Capacity to communicate (*continued*)
 201; joint self-consciousness and, 201;
 misemployment of in a false promise,
 201–202, 205. *See also* Communication
- Capacity to feel (pleasure and displeasure):
 as receptivity of capacity for practical
 knowledge, 82, 137; aesthetic judgments
 and, 97–98. *See also* Feeling; Pleasure
- Capacity to judge: capacity to think and,
 100–101; as capacity to know, 105; as
 dependent on external conditions,
 105–106; as self-determining, 106–107,
 131–132; error and, 108–111. *See also*
 Capacity to know
- Capacity to know: as capacity to judge,
 105; error and, 108–111; subjective
 universal validity and, 112–116, 122–
 123; capacity to communicate and, 115,
 122–123, 201; formula of humanity
 and, 170, 172. *See also* Capacity to judge;
 Knowledge (discursive)
- Categorical imperative, ix–x, 2, 5, 140–141,
 155; intrinsic goodness/badness and,
 15–16; as principle of will's self-
 agreement, 133, 242–243; as principle
 of self-constraint, 140–141, 159, 243.
See also Imperatives; Moral law
- Character, 242–243; weakness of will and,
 45–46; maxims and, 46; concept of
 person and, 85n19; virtuous, 163–164,
 250, 251
- Choice, 25n1, 45–49; practical judgment
 and, 49–51, 61–64; intention and, 61,
 64; will and, 65; wish and, 66–69,
 79–81, 86–87, 124, 195; judgments of
 practicability and, 67, 142–144; outer
 freedom and, 190–193; contract and,
 198n9. *See also* Free power of choice;
 Power of choice
- Cleverness, 34, 40, 55
- Cognition. *See* Knowledge (discursive)
- Communicability, 58–59, 70–71, 98, 205;
 agreement and, 58; knowledge and,
 58–59. *See also* Capacity to communicate;
 Communication
- Communication, 1–2, 200–201, 204–205;
 discursive knowledge and, 201; as
 shared self-consciousness, 201. *See also*
 Capacity to communicate;
 Communicability
- Competition, 65, 214–215
- Conduct. *See* Practice
- Conflict: impermissible willing and, 18–19,
 242; unity of thought and, 105, 131;
 error and, 108–109; normativity and,
 131–132; in object willed and in willing
 of object, 229–230. *See also* Agreement;
 Contradiction
- Constraint: on judgment, 106–107, 110,
 128; on willing, 128, 140–141, 145, 159,
 241–243
- Contract, x, 130n1, 188–189, 235; social, x,
 130n1, 154–155; concept of, 198–201;
 juridical duty (legal obligation) and, 234
- Contractualist accounts of moral
 obligation, 235. *See also* Social contract
- Contradiction: wrongdoing and, 16;
 inner, or between a maxim's content
 and form, 16–19, 167, 227; in maxim
 of assaulting others' freedom, 194–195;
 in maxim of false promising, 202–209;
 in maxim of indifference to others'
 happiness, 210–213, 216; in conception
 and in the will, 228–231; in maxims
 opposed to duties to oneself, 232
 “Copernican” proposition, 13, 14, 183
- Culpability, judgments of, 227–228
- Darwall, Stephen, 235n31
- Davidson, Donald, 31n8, 45n19
- Deliberation, 38–40, 86–88; joint, 198n9
- Desiderative reason. *See* Practical reason
- Desire, 26–27, 29; rational, 25, 33–44;
 efficacy of, 27, 29; deliberate, 28;
 practical thought and, 28–30, 34;
 sensible, 30, 37–39, 71, 76. *See also*
 Faculty of Desire; Inclination
- Duty (*Pflicht*), 154n1, 184; to oneself,
 173n8, 231–239; action contrary to,
 217–218, 220; perfect (negative) and
 imperfect (positive), 228–231. *See also*
 Beneficence; Justice, natural; Obligation
- End in itself, 154; humanity as, 172, 177;
 rational being's existence as, 176–177
- Ends, 72–78, 171–172; means and, 40–41,
 74–75; practical principles and, 47–48,
 169
- Enjoyment, 38, 76, 143; as material
 condition of wish, 69, 73, 124, 238;
 receptivity and, 71; instinct-based,
 78n10; happiness and, 82–83, 142–143;

- security and, 87. *See also* Gratification; Pleasure
- Epicurus, vii, 72, 88
- Error: knowledge and, 105–106; conflict and, 108–109; judgment and, 108–111; as misuse of capacity to know, 110
- Fact of reason, 121
- Faculty of desire, 26–27, 29, 37; higher, 25; as expression of lifepower, 27; lower, 38, 76, 82. *See also* Desire
- Faculty of desire in accordance with concepts, 28–33, 34, 47. *See also* Practical thought
- False promising, 201–202; maxim of, 188, 196–209, 218–220. *See also* Contradiction
- Feeling: sensible desire and, 30, 38; aesthetic judgments and, 57n30; in practical attitudes, 62–63. *See also* Capacity to feel; Pleasure
- Fichte, J. G., 18n10
- Form of practical knowledge, 122–124; presupposition of universality and, 124–127, 129; capacity for practical knowledge and, 126, 173; as first principle of practical knowledge, 127; as the form of willing, 129; as reflected in the categorical imperative's three formulas, 149–154
- Form of willing, 18–19, 129–141; as constitution or essence, 129–131; as norm, 131–133; as principle of self-agreement, 133–134, 242; as practical law, 134–135; as self-legislated, 135–136; as categorical imperative, 140–141, 155–159. *See also* Law of the will; Will
- Formula of autonomy (or of autonomous membership in a possible realm of ends), 150–151, 167, 170; as an expression of the form of practical knowledge, 149–150; in relation to the other formulas, 150–151, 153–154, 170. *See also* Autonomy
- Formula of humanity, 3, 21, 150–151, 153–154; in relation to the other formulas, 150–151, 153–154, 170; as an expression of the form of practical knowledge, 167–172; formula of universal law and, 167–183, 185–186, 209, 230–231, 235–236. *See also* Humanity
- Formula of universal law, 2, 155; as standard of moral judgment, 2, 185–186; criticisms of, 2–6; form of practical knowledge and, 7–8, 149–154, 155–159; goodness and, 9–11; in relation to the other formulas, 149–154; concept of a categorical imperative and, 155; laws of nature and, 159–164; formula of humanity and, 167–183, 185–186, 209, 230–231, 235–236; as a positive requirement, 174–178, 209–220, 220–223; application of, 184–239; duties to oneself and, 231–239. *See also* Universalization test
- Frankfurt, Harry G., 84n18
- Freedom: of the power of choice, 107, 190; modes of, 87–88n20; end-setting and, 169. *See also* Outer freedom
- Free power of choice, 44–49, 65, 81, 107, 169, 184; freedom of, 33n9, 87–88n20, 107, 190; inner and outer employment of, 68n2, 81, 191; capacity to wish and, 81; fundamental act of, 81–86, 129–130, 137, 138. *See also* Choice; Maxims; Power of choice; Wish
- Frege, Gottlob, 103
- Geach, Peter, 103n4
- General will, 129–130n1
- Good(ness): as object of practical knowledge, 7–8, 11–14, 44, 49–54, 90–91, 118–120, 127–128, 182; concept of, 8, 12–13, 14, 59, 76–77, 98, 113, 178–180, 183; formula of universal law and, 9–11; intrinsic (unconditioned), 10, 15, 79–81, 164, 166–167, 226–228; relational concept of, 59–60; on the whole, 66–68, 79–81; simple, 67–69, 79–81; the agreeable and, 70–78, 82; categories of, 74–75, 79–81. *See also* Good will; Highest good; Morality
- Good will: intrinsic goodness of, 15, 182; as *Groundwork's* starting point, 181–183; humanity and, 182–183
- Gratification, 76, 83–84. *See also* Enjoyment
- Guyer, Paul, 21
- Happiness: as matter of highest good, 75n9; as fundamental object of wish (and end for every person), 81–86; as

- Happiness: (*continued*)
of imagination (and object of sensible desire), 82–83, 84n17, 86, 88; form and matter of, 86–90, 213–214; generic (formal) and specific (substantive) conceptions of, 86–91; prudence and, 141–143; morality and, 250–252
- Hare, R. M., 224n22
- Hegel, G. W. F., 3, 141n4, 197n8
- Herman, Barbara, 9n4, 21n13, 175n10, 220n20, 224n22, 226n25, 228n27
- Heteronomy: in theories of the will, 136, 180; of natural necessity, 163, 164
- Highest good, 65, 75n9, 163–164, 166, 181; relations among components of, 75n9; in a person, 79
- Hill, Thomas E., Jr., 5n3, 21n13, 235n30
- Hobbes, Thomas, vii, 26, 28n4, 58–61, 71, 72; on “good” and “bad,” 58
- Human nature, empirical concept of, 22, 167, 186–187, 232; idea of humanity and, 22, 167; system of practical law and, 140
- Humanity, 22, 46n20, 93n23, 141n4, 167–169; as practical reason, 167–169; as rational nature, 167–168; as capacity to set an end, 168–169; employment of, 170–171, 231; as an end (in itself), 171–172, 177; in a person, 173n8, 238; the good will and, 183n16; in a community, 208. *See also* Formula of humanity
- Humanity, formula of. *See* Formula of humanity
- Hume, David, vii, 1–2, 12, 14, 25n1, 27, 43, 55, 56, 72, 80n14, 244n2
- Hutcheson, Francis, 56
- Imperatives, 16, 42; hypothetical, 15, 16, 41, 224n22, 242; prudential, 42n15, 141–144. *See also* Categorical imperative
- Inclination, 29–30, 70–71, 102–103n3; natural, 214. *See also* Sensible desire
- Incontinence. *See* Weakness of will
- Indifference to others’ happiness, maxim of, 210–213, 216–218, 220, 225, 229, 230. *See also* Contradiction
- Instincts: as basis of natural system of animal life, 78n10; natural inclinations and, 214
- Instrumental rationality. *See* Means–end rationality
- Integrity, 242–243
- Intention, 28, 30–33; in comparison with will, 26, 34; rationality of, 33–44; sensible desire and, 33, 42–43, 102–103n3; in comparison with (free) choice and practical judgment, 45, 49, 61, 64; practical knowledge and, 55–56. *See also* Practical thought
- Irwin, T. H., 25n1
- Judgment, 97–108, 111; determining, 52, 101; relative, 58–60, 70–71; “logical” (cognitive) and “aesthetic,” 97–98, 117; thought and, 98–104; self-sustaining character of, 101–104; validity of, 104–108. *See also* Appraisal, judgments of; Error; Knowledge (discursive); Practical judgment
- Justice, natural: law of, 138–140, 234–235; duties of, 187, 225; primitive duty of, 189–196
- Kemp, John, 197n8
- Knowledge (discursive), 97–98; Kant’s “Copernican” understanding of, 13–14, 119, 183; communicability of, 98, 115, 201; validity of, 104; unity of, 104–105; two-sided universal validity of, 111–118; as knowledge from concepts, 113–115, 117–118; as rational, 117–118. *See also* Capacity to know; Practical knowledge; Rational knowledge; Theoretical knowledge
- Korsgaard, Christine M., 9n4, 21n13, 79n13, 174n9, 189n2, 224n22, 228n27
- Law of the will, 134–136; as practical law, 134–135; as self-legislated, 135–136; as formal, 136; as practically limiting (constraining), 140–141; in comparison with the principle of prudence, 144. *See also* Form of willing; Moral law; Will
- Laws, 134, 236; theoretical (of nature) and practical (of freedom), 7, 134–135, 159–164; reason and, 94, 113–118; knowledge from concepts and, 113–115; universality of, 113–115, 134, 156, 236; represented through principles, 134, 157–158; necessity of, 236. *See also* Law of the will; Moral law; Practical law

- Legality. *See* Morality
 Lifepower, 27, 37, 143
 Locke, John, 85n19, 98
- Mackie, J. L., 11n6
 Maxims, 46, 130; form and matter of, 18–20, 233–234, 236, 238–239; universality of, 18–20, 130; universalization test and, 18–19, 241–242, 245; free power of choice and, 46, 47, 48, 129–130; individual persons and, 129–130, 156n3; practical laws and, 136–137. *See also* Contradiction
 Means. *See* Action; Ends; Means–end rationality
 Means–end rationality, 33–44; universalization test and, 164–165, 223–227. *See also* Prudential rationality; Rationality
 Merit, judgments of, 227–228
 Mill, J. S., 3n1, 5n3, 165, 224
 Moral law, 134; as valid for all rational beings, 91–94; as first principle of action, 94; concept of good and, 178–183; as first principle of practical knowledge, 183; first and direct application of, 219–220; as originally positive, 222–223. *See also* Categorical imperative; Law of the will
 Moral point of view, 243–247
 Moral worth, 15, 80, 133, 154n1, 183n16; judgments of merit and, 228
 Morality: practical reason and, vii, ix–x, 8; as a type of knowledge, 7–8; as knowledge of intrinsic goodness, 10, 164, 252; as nonaccidental self-agreement of will, 133; legality and, 133, 180; prudence and, 141–145; as form of knowledge of the good, 220, 246–247, 252; broader and narrower senses of, 235; persons' relation to, 242–247; as condition of integrity and character, 242–243
 Murdoch, Iris, 11n5
- Natural law tradition, x, 154
 Nature, 162–164, 168; objective universal validity of cognition and, 113–115; sensible and supersensible, 162–163
 Necessitation: by practical law, 141, 155; by another's power of choice, 190–191
 Necessity: as mode of practicality, 87; as mode of freedom, 87–88n20; objective practical, 88n20, 236; in rational cognition, 101, 161–162, 236; as necessitation, 141, 155; universality and, 154n1, 236. *See also* Validity
 Normativity, 131–133; conceivability of conflict and, 131–132
- Obligation, 141, 154n1, 184, 233; ground of, 93, 154–155n2, 235; consciousness of, 163, 244; in narrow legal sense, 234; contractualist accounts of, 235–236. *See also* Constraint; Duty
 O'Neill, Onora, 21n13, 41n14, 175n9, 224n22, 228n27
 Outer freedom, 190–193; in comparison with general practical freedom and with inner freedom, 190; self-sufficiency and, 190–193
- Paradox of method, 178–183
 Patzig, Günther, 50n25
 Person: fundamental judgment of, 81–86, 129–130, 137, 177, 237–239; as self-constituting practical subject, 81–82, 92; as subject with the capacity for practical knowledge, 81–82, 84–85, 129; as free (accountable) cause of action, 227; personality and, 238. *See also* Practical subject
 Plato, vii, 1, 9, 68, 171n6, 225, 245, 247
 Pleasure, 76, 82–83; efficacy of sensible desire and, 30; in the agreeable, 70, 76–77, 78, 82. *See also* Capacity to feel; Enjoyment; Feeling; Gratification
 Pogge, Thomas W., 21n13, 219n19
 Point of view. *See* Moral point of view
 Postulates of pure practical reason, 55n26
 Power of choice (*Willkür*), 28, 41, 46n20. *See also* Free power of choice
 Practicability, judgments of, 40–41, 67
 Practical judgment, 49–54; positive and negative, 51–54; in comparison with judgments of appraisal, 56–58; in contrast to relative judgments, 58–60; in comparison with choice and intention, 61–64; as basis of practical attitudes, 62–63; the will and, 64–65; types of, 79–81; fundamental, 81–86. *See also* Practical knowledge

- Practical knowledge, 50, 54–56, 118–122; first principle of, ix–x, 126–127, 183; capacity for, 7–8, 20–21, 64, 126, 171, 173; as knowledge of the good, 7–14, 50–54; theoretical knowledge and, 7, 118–124, 134–135, 162; as efficacious, 50–53, 118–119; know-how and, 55–56; communicability of, 58–61; objective modalities of, 88n20; order of, 90–91, 117–118, 137, 140, 186, 245; as self-determining, 119–122, 237; as self-knowledge, 120–122, 144; practical thought and, 122–124; necessarily coinciding double universal validity of, 122–126; practical law and, 134–136. *See also* Form of Practical knowledge; Rationality; Rational Knowledge
- Practical law, 7, 134–140, 162, 168; as law of freedom, 7, 154, 163; practical knowledge and, 134–136; theoretical law and, 134–135; form of, 135; as self-legislated (law of autonomy), 134, 136; formal and material, 136; system of, 137–140; of natural justice and of beneficence, 138–140, 234–235; as constraint, 140–141, 184; laws of nature and, 159–164. *See also* Laws; Moral law
- Practical reason, 6–8, 118–121, 168–169; as the source of morality, vii–xi, 2–4, 7–8, 20–21, 241; not a power of apprehension, ix–xi, 12; as autonomous, ix–xi; as capacity to know the good (what ought to be), xi, 6–8, 9–14, 44, 53–54, 70, 80; attenuated conception of, 6–7, 34, 165; as capacity for practical knowledge, 7–8, 20–21; as humanity, 167–170; two ways of regarding, 169–170; as original idea of good will, 182. *See also* Rationality; Reason; Will
- Practical subject, 31–33, 35–36; merely prudential, 92–93; plural, 198–199. *See also* Person
- Practical thought, 28–44; self-conscious efficacy of, 29–30; rationality of, 33–44; as rational desire, 34; in a subject whose practical thought of ends is determined by sensible desire, 42–44, 92, 102–103n3. *See also* Faculty of desire in accordance with concepts; Intention; Thought
- Practicality, 29–30; rationality of, 42; perfection in, 55–56, 87–88; modes of, 79, 86–87
- Practices: individual, 48; universalization test and general (social), 188–189, 197, 208
- Praise, 57, 227–228
- Presupposition of universality, 124–128, 129–130, 133; the formula of universal law and, 155–159; the formula of humanity and, 167–172. *See also* Form of practical knowledge
- Presuppositions of practical judgment, 19–21, 108, 124–128; formal and material, 126–127. *See also* Presupposition of universality
- Principle of morality. *See* Categorical imperative; Moral law
- Principles: in philosophy, 1–2, 244–247; as representations of laws, 5, 134, 157–158; as primary in rational cognition, 35, 117–118; subjective practical (maxims), 46, 130; ends and practical, 47–48, 169; practical, 48; broader and narrower conceptions of, 168n4
- Promise, concept of, 198–201
- Promising. *See* False promising
- Prudence (*Klugheit*), 34n10, 87; imperatives of, 42n15, 141–144; morality and, 93n22, 141–145, 166–167. *See also* Prudential rationality
- Prudential rationality, 92–93; universalization test and, 164–167, 223–227, 243. *See also* Means–end rationality; Prudence; Rationality
- Rational being: concept of, 5–7, 8, 59–60, 84; moral law and concept of, 91–94, 156–157, 157–158, 168, 176; sensible and supersensible nature of, 162–163
- Rational desire, 25; traditional conception of will as, 25–26; practical thought as a form of, 34
- Rational knowledge: Kant's taxonomy of, 7; as knowledge from principles, 35, 117–118; as self-determining, 106–107, 119, 121–122; two-sided universal validity of, 111–118, 123. *See also* Knowledge (discursive); Practical knowledge; Reason; Theoretical knowledge

- Rational nature. *See* Humanity
- Rationality, 35, 46n20; goodness and, 13n7; two forms of, 33–34, 141–144. *See also* Means–end rationality; Practical knowledge; Practical reason; Prudential rationality
- Rawls, John, 85n19, 178, 226n25
- Reason: as the cognitive capacity, 7–8, 13–14; as the capacity for knowledge from principles, 35, 94, 117–118; as ultimately one, differentiated only in application, 98, 118, 121. *See also* Knowledge (discursive); Practical Reason; Rational knowledge
- Reason for action. *See* Sufficient reason
- Reath, Andrews, 224n22
- Reflection: philosophical, 1–2, 245; moral, 17n9, 216–220, 244–245; relation between moral and philosophical, 244–247, 249–252
- Ross, W. D., 20n12, 219n19
- Rousseau, Jean-Jacques, 129–130n1
- Royce, Josiah, 85n19
- Scanlon, T. M., 219n19, 235n29
- Schiller, Friedrich, 141n4
- Schneewind, J. B., 9n4
- Schopenhauer, Arthur, 225n24
- Self-agreement: of ends, 75; of knowledge, 105, 131–132; of the will (morality), 132–134; relative (prudential) and absolute (moral), 141–144; in a particular will, 239; as condition of integrity and character, 242–243
- Self-conception, practical, 31–33, 43; self-specification of, 33; ideal, 79; of a merely prudentially rational being, 92; as made up of maxims, 156n3; constituting a person, 177; as presupposed in practical judgment, 237, 238–239; unity of (as condition of character and integrity), 242–243
- Self-consciousness, 98–100; of practical thinking, 29–30; of intention, 30–31; of the practical subject, 31–33, 35–36; rationality and, 35; of choice, 45–46; of practical judgment, 50–51, 124, 125–126, 237–239; of thought, 98–100; as unity, 99; of judgment, 100–101, 131–132; judgment's validity and, 104; of knowledge, 104–105; universality of, 113; of pure practical reason, 120–121; shared, 200–201. *See also* Self-conception; Unity
- Self-determination: judgments of, 63; of knowledge, 107, 121; of practical knowledge, 121–122; formal, 121–122, 132–133; material, 122; practical, 122–123; of a particular will, 129–130, 156n3
- Self-knowledge: practical knowledge as, 120–122, 144; practical, 121; theoretical, 121
- Self-legislation. *See* Autonomy
- Self-relation: of bare practical thought, 122–123; of practical knowledge, 122–124, 125–126, 135
- Self-sufficiency, practical, 88–90, 190; as component of the form of happiness generically conceived, 88–89; as necessary object of wish, 88; law of natural justice and, 138–139; outer freedom and, 190–193
- Sensible desire, 30, 37–39, 71, 76; efficacy of, 30; as material condition of practical thought, 36–39; reproductive character of, 37–38, 76; judgments of the agreeable and, 71; pleasure and, 76. *See also* Desire; Inclination
- Sidgwick, Henry, 225
- Singer, Marcus G., 224n22, 234n28
- Social contract, x, 130n1, 154–155. *See also* Contract
- Spontaneity: of cognition and reason, 14, 107; in contrast to receptivity, 29; of practical thinking's efficacy, 29–30, 103n3; self-consciousness and, 100; of thinking, 100
- Stoics, viii, x, 88
- Sufficient reason, idea of: in judgments grounding choice, 49–51, 64; in judgments grounding wish, 69
- Theoretical judgment, 67. *See also* Theoretical knowledge
- Theoretical knowledge, 7, 13, 14, 118–119, 121, 143; in comparison with practical knowledge, 7, 118–121; practical thought and, 35–36, 38–41; prudence and, 142–144. *See also* Practical knowledge; Rational knowledge

- Thought, 98–102; self-consciousness and, 98–100; unity of, 99, 101, 105; in comparison with judgment, 100–102, 104–105; in comparison with knowledge, 104–105. *See also* Judgment; Knowledge (discursive); Practical thought
- Typic of pure practical judgment, 162–163
- Unity: of agency, 31–33, 42–43, 77–78, 84–85; of consciousness, 99; of thought, 99, 101, 105; of judgment and knowledge, 104–105, 113; of practical knowledge, 171; of self-conception, 242–243. *See also* Self-agreement; Self-consciousness
- Universal law, formula of. *See* Formula of universal law
- Universality: as form of maxim, 18–20, 130; in discursive cognition's self-consciousness, 113; in discursive cognition's representation of its object, 113–115; of discursive cognition's validity, 115–118; subjective, 115–116; objective, 116; comparison of subjective and objective, 116–117; rationality and objective, 117–118; of practical cognition's self-relation, 122–124; of a law, 134, 156, 236; necessity and, 154n1, 236. *See also* Presupposition of universality; Universalization test
- Universalization test, 17–20, 241–243, 244–245; moral reflection and, 17n9, 216–220, 226, 244–245; means–end reasoning and, 164–167, 223–228; idea of a general practice and, 188–189, 197; imperfect (positive) duties and, 209, 216–217, 219–220; universal permission and, 219; assessment of merit or culpability and, 227–228; duties to oneself and, 236. *See also* Formula of universal law
- Validity, 104–108, 236–237; in comparison with error, 108–109; capacity to communicate and subjective universal, 115–116, 201; subjective universal, 115–116; objective universal, 116; comparison of subjective and objective universal, 116–117; necessary coincidence of practical cognition's subjective and objective universal, 122–126; form as standard of, 131–132; necessity and, 236–237
- Vice, 166
- Virtue, vii–viii, 141n4, 145, 163–164, 166, 239, 250–251; as component of highest good, 75n9; happiness and, 75n9, 166; as mode of freedom, 87–88n20; good will and, 183n16
- Voluntary action: in broad sense (as desired), 25n1, 26–27; in narrow sense (as intended), 28–33
- Weakness of will, 44–46
- Wick, Warner, 234n28
- Will (*Wille*), 25–26; as practical (“desiderative”) reason (the capacity for practical knowledge), 7, 20–21, 64–65, 129; as rational desire (practical thought), 28–33; as (free) power of choice, 44–49; in an individual person, 129; autonomy of, 134–136. *See also* Form of willing; Law of the will; Practical Reason
- Williams, Bernard, 4n2, 18n10, 20n12, 109n6, 242n1
- Willkür*. *See* Free power of choice; Power of choice
- Wish, 67–69, 79–81, 86–88; choice and, 66–69, 79–81, 86–87, 124, 195; as judgment of the simply good, 67–68, 79; broader and narrower conceptions of, 68n2; material conditions of, 69, 75–77, 124, 238; for happiness, 81–86; for practical self-sufficiency, 88–90
- Wolff, Christian, ix
- Wolff, Robert Paul, 188n1, 197n8
- Wood, Allen W., 3n1, 5n3, 21n13, 174–175n9, 175n10, 203n12, 214n18, 224n22, 228n27
- Wrongness: intrinsic, 15–20, 144, 166–167, 227, 242, 246; contradiction and, 16; in judgment, 110; futility and, 166, 207n14

"Stephen Engstrom provides a highly original interpretation of the categorical imperative as the criterion of rational knowledge and action. In focusing on the general formula and the law-of-nature variant, and trying to make them work, the book presents a welcome counterweight to currently fashionable interpretations that dismiss these formulations in favour of the second variation, commonly called the 'formula of humanity.'"

—JENS TIMMERMANN, UNIVERSITY OF ST ANDREWS

