## DRAWING FROM THE SOURCES OF REASON:

# REFLECTIVE SELF-KNOWLEDGE IN KANT'S FIRST CRITIQUE

by Melissa McBay Merritt B.A., Yale University, 1994

Submitted to the Graduate Faculty of Arts and Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy

University of Pittsburgh

#### UNIVERSITY OF PITTSBURGH

#### FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Melissa McBay Merritt

It was defended on

April 14, 2004

and approved by

James F. Conant, Professor, Philosophy, University of Chicago

Anthony Edwards, Professor, Religious Studies, University of Pittsburgh

John McDowell, University Professor, Philosophy, University of Pittsburgh

Nicholas Rescher, University Professor, Philosophy, University of Pittsburgh

Dissertation Director:

Stephen Engstrom, Associate Professor, Philosophy, University of Pittsburgh

### DRAWING FROM THE SOURCES OF REASON: REFLECTIVE SELF-KNOWLEDGE IN KANT'S FIRST *CRITIQUE*

Melissa McBay Merritt, Ph.D. University of Pittsburgh, 2004

Kant advertises his *Critique of Pure Reason* as fulfilling reason's "most difficult" task: self-knowledge. As it is carried out in the Critique, this investigation is meant to be "scientific and fully illuminating"; for Kant, this means that it must follow a proper *method*. Commentators writing in English have tended to dismiss Kant's claim that the *Critique* is the scientific expression of reason's self-knowledge — either taking it to be sheer rhetoric, or worrying that it pollutes the Critique with an unfortunate residue of rationalism. As a result, there is little sustained treatment of the method of the Critique in the secondary literature. Since Kant holds that the substantive insights of critical philosophy are not separable from the methodological context in which they come to light, this is a serious mistake. My dissertation corrects for this, by approaching the Critique through an examination of its method. In doing so, it yields a reading of the Transcendental Deduction that not only promises to resolve current debates about its "proof structure", but also fully accounts for the Deduction's pivotal role in the work as a whole.

# TABLE OF CONTENTS

Acknowledgements	vi
Notes on Sources and Abbreviations	viii
Remarks on Translation, Terminology and Stylistic Conventions	xi

INTRODUCT	ION1
CHAPTER 1	Kant's Idea of a Critique of Pure Reason11
CHAPTER 2	Reflective Metaphysics: a Preliminary Account55
CHAPTER 3	Spontaneity and Self-Knowledge in the <i>Critique</i> 80
CHAPTER 4	The Role of the Transcendental Deduction in the <i>Critique</i> 's Synthetic Method115
	Coda. The Theme of Self-Knowledge in the Deduction178
APPENDIX A The Connection between the Notion of a 'Noumenon'	
	and the Appeal to Spontaneity
APPENDIX B	Sensibility as a "Faculty"

BIBLIOGRAPHY	189
--------------	-----

# LIST OF FIGURES

Figure 1	Division of Rational Cognition into Formal and Material Sciences	. 52
Figure 2	Kant's Division of the Philosophy of Pure Reason (from A841/B869 ff.)	. 53
Figure 3	Schematic Account of the Critique's Front Matter	181
Figure 4	Taxonomy of "Synthesis"	182

#### Acknowledgements

This dissertation was written with the support of an American Dissertation Fellowship from the American Association of University Women during the academic year 2001-2002. I also gratefully acknowledge the non-teaching fellowship support I received from the University of Pittsburgh Philosophy Department for Spring term 2001, and for the academic year 2003-2004.

My greatest intellectual debt is to my adviser, Steve Engstrom. The writing of this dissertation neatly coincided with his tenure as chair of the philosophy department; despite his numerous administrative obligations, he always found time to meet with me to discuss my work. He patiently thought his way through many a muddy first draft, and offered helpful suggestions about how to express my ideas with greater clarity. (And still I have always his shining example to aspire to: no one, I am sure, understands Kant so profoundly and expresses that understanding so gracefully as Steve.) Without his kindness and encouragement, I would not have seen my way through to the end.

I am also grateful to my other mentors. John McDowell was the unerringly insightful second reader on my committee, who was usually able to appreciate what it was I was trying to say better than I could myself. His generous encouragement and interest in my project has meant a lot to me. Nicholas Rescher threw many challenging questions my way, and offered helpful suggestions about further issues to consider. Tony Edwards responded to my work with great thoughtfulness, and always had helpful suggestions for improvement. Special thanks are due to Jim Conant, who supervised the prospectus for this project, and served as my unofficial adviser in my first few years as a graduate student at Pittsburgh. As often as I struggled with the feeling that I was not entirely at home studying philosophy, Jim reminded me that I might have something worth saying as a philosopher. Without his encouragement, I am quite certain that I would not have made it past my second year in graduate school. I am also grateful to Jim for arranging for my visit to the University of Chicago as a Visiting Lecturer in 2000, and to Michael Forster for seeing to it that I had a nice office on campus in which to work. The dissertation project took shape in those months, in large part owing to the conducive environment I found in Chicago.

Several friends have been wonderful readers of my work. Matt Boyle was kind to have read and commented on parts of the dissertation in their early, and particularly obscure,

incarnations; the discussions we had helped to determine the direction in which I took the project. Thomas Land read and commented on several chapters as they were in progress, and responded with great care to my surely tiresome queries about Kant's and Dieter Henrich's German. I am also grateful to Jessica Moss, Kieran Setiya, Michael Thompson, and Doug Lavin for their helpful editorial comments on Chapter 3 (in its guise as my job market dossier paper).

I am deeply grateful to Bridget Clarke and Maura Tumulty, who went through graduate school a couple steps ahead of me. I am the better for the wisdom of their experience. More importantly, though, I am the happier for their friendship and for the merry hours we spent over tea, shortcake, Belgian beer, and Indian food.

My far-flung friends Sarah Parsons, Julie Püttgen, Randy Hampton, and Yael Goldman have seen me through some rough patches; I am grateful to them for their support, and inspired by their exuberance.

My entire family has been most supportive of my decision to go to graduate school, and good-natured about the protracted childhood that it has entailed. I am especially grateful to my wonderful parents, Jim and Nancy Merritt, for their love and tireless faith that I can do whatever I set my mind to doing.

My debt to Markos Valaris far outruns any possible expression of gratitude. His enthusiasm about this project revived and sustained my own. He read every chapter, several times and in various stages of development, and I always learned an enormous amount from the discussions that followed. He was a sharp editor, and I beg his forgiveness for often receiving his constructive criticism like a stubborn child. Best of all, though, he taught me how to forget my work and have a proper *vacation*. For that, and for everything, I thank him from the bottom of my heart.

#### Notes on Sources and Abbreviations

References to the works of Kant, with the exception of the *Critique of Pure Reason*, are according to German Academy edition pagination: *Gesammelte Schriften*, edited by the Königlich Preußischen Akademie der Wissenschaften, later the Deutschen Akademie der Wissenschaften zu Berlin. 29 vols. (Berlin: Walter de Gruyter, 1900-). References to the *Critique of Pure Reason* are given according to custom by the pagination of the first ("A") and second ("B") editions of 1781 and 1787, respectively. If the cited passage is included in both editions, the citation includes both A and B page references.

Full citations include the title of the work (usually by abbreviation) and the page in the Akademie edition (but not the volume of the Akademie edition, which is given only here in the list of abbreviations). References to the *Critique of Pure Reason* typically give only the page in either or both editions.

The following abbreviations are used in citations. Section numbers (§) are generally included if the passage comes from an explicitly numbered section.

Anth.	Anthropologie in pragmatischer Hinsicht (1798). (Ak. 7) Translated by Mary Gregor, under the title Anthropology from a Pragmatic Point of View. The Hague: Martinus Nijhoff, 1974.
Aufklärung	"Beantwortung der Frage: Was ist Aufklärung?" (1784). (Ak. 8) Translated by Mary Gregor, under the title "An Answer to the Question: What is Enlightenment?". Cambridge Edition of the Works of Immanuel Kant, <i>Practical Philosophy</i> . Cambridge: Cambridge University Press, 1996.
Entdeckung	<ul> <li>"Über eine Entdeckung, nach der alle neue Kritik der reinen Vernunft durch eine ältere entbehrlich gemacht werden soll." (Ak.</li> <li>8) Translated by Henry Allison, under the title "On a Discovery, According to which Any New Critique of Pure Reason Has Been Made Superfluous by an Earlier One," in <i>The Kant-Eberhard</i> <i>Controversy</i>. Baltimore: The Johns Hopkins University Press, 1973.</li> </ul>
Fortschritte	"Welches sind die wirklichen Fortschritte, die die Metaphysik seit Leibnizens und Wolf's Zeiten in Deutschland gemacht hat?" (1791). (Ak. 20) Translated by Ted Humphrey, under the title "What Real Progress Has Metaphysics Made in Germany since the Time of Leibniz and Wolff?". New York: Abaris Books, 1983.

G	<i>Grundlegung zur Metaphysik der Sitten</i> . (Ak. 4) Translated by James Ellington, under the title <i>Grounding for the Metaphysics of Morals</i> . Indianapolis: Hackett, 1981.
Hechsel	<i>Logik Hechsel. Logik-Vorlesung: Unveröffentliche Nachschriften</i> <i>II</i> , ed. Tillman Pinder (Hamburg: Felix Meiner, 1998). <sup>1</sup> Translated by J. Michael Young. Cambridge Edition of the Works of Immanuel Kant, <i>Lectures on Logic</i> . Ed. J. Michael Young. Cambridge: Cambridge University Press, 1992.
Logic	<i>Logik Jäsche</i> . (Ak. 9) Translated by J. Michael Young. Cambridge Edition of the Works of Immanuel Kant, <i>Lectures on Logic</i> . Ed. J. Michael Young. Cambridge: Cambridge University Press, 1992.
KpV	<i>Kritik der praktischen Vernunft</i> (1788). (Ak. 5) Translated by Mary Gregor, under the title <i>Critique of Practical Reason</i> . Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 1997.
KrV	<i>Kritik der reinen Vernunft</i> (first (A) edition, 1781 (Ak. 4); second (B) edition, 1787 (Ak. 3)). Translated by Norman Kemp Smith, under the title <i>Critique of Pure Reason</i> . London: Macmillan, 1929; New York: St. Martin's Press, 1965. Translated by Paul Guyer and Allen W. Wood, also under the title <i>Critique of Pure Reason</i> . Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 1997.
KU	<i>Kritik der Urteilskraft</i> (1790). (Ak. 5) Translated by Paul Guyer and Eric Matthews, under the title <i>Critique of the Power of</i> <i>Judgment</i> . Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press, 2000.
MAN	Metaphysische Anfangsgründe der Naturwissenschaften. (Ak. 4) Translated by James W. Ellington, under the title Metaphysical Foundations of Natural Science, in Kant's Philosophy of Material Nature. Indianapolis: Hackett, 1985.
Met L <sub>2</sub>	<i>Metaphysik L</i> <sub>2</sub> . (Ak. 28) Translated and edited by Karl Ameriks and Steve Naragon. Cambridge Edition of the Works of Immanuel Kant, <i>Lectures on Metaphysics</i> . Cambridge: Cambridge University Press, 1997.
Р	Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können (1783). (Ak. 4) Translated by

 $<sup>\</sup>overline{^{1}Logik}$  Hechsel is not in the Akademie edition.

	Paul Carus, revised by James Ellington, under the title <i>Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward As Science</i> . Indianapolis: Hackett, 1977.
Prize Essay	Untersuchung über die Deutlichkeit der Grundsätze der natürlichen Theologie und der Moral, zur Beantwortung der Frage, welche die Königliche Akademie der Wissenschaften zu Berlin auf das Jahr 1763 aufgegeben hat. (Ak. 2)
R	Reflexionen zur Metaphysik. (Kants handschriftlicher Nachlaß). (Ak. 17-18)

#### **Remarks on Translation, Terminology and Stylistic Conventions**

#### 1.

Translations from Kant's German texts are my own, although they are based on the English translations listed in the Note on Sources and Abbreviations. I have used italics to express phrases of Kant's emphasis (generally achieved by spacing, or *Sperrdruck*, in the Akademie edition), but have ignored the use of bold type (*Fettdruck*) as a higher degree of emphasis, rendering all emphasized phrases with italics. I italicize all Latin phrases except "a priori" and "a posteriori". I have italicized "Kritik" and "Prolegomena" when it is clear that Kant is referring to his own titles, and left them plain when it is ambiguous.

## 2.

Titles of any section of Kant's *Critique* are given as titles in the following way, usually abbreviated and always without quotation marks or italics: (e.g.) Phenomena and Noumena (instead of "Of the Ground of the Distinction of All Objects in General into Phenomena and Noumena"). This typographical convention will be employed to mark distinctions between particular chapters and the arguments or problems which figure in those chapters. Making this sort of distinction will be particularly important in Chapter 4, where I will want to distinguish between the *chapter* entitled the Transcendental Deduction of the Pure Concepts of the Understanding (§§15-27 in the second-edition *Critique*), and the deduction argument in the strict sense, which (as I argue) constitutes only a portion of that chapter (§§21-26). The shorthand for the former will follow the convention for designating titles of sections of the *Critique* (e.g., the Transcendental Deduction), while shorthand for the latter will use the lower-case (e.g., the deduction argument). Also, I refer to the *Critique*'s Prefaces and Introduction taken together as the "front matter".

## 3.

My decision not to work exclusively with existing translations of Kant's texts stems from the observation that many of the terminological distinctions that Kant draws — e.g., among *Kraft, Fähigkeit*, and *Vermögen* — are not always preserved within and across the various translations of Kant's texts. I have rendered "Vermögen" and all of its cognates with "faculty" (in keeping with Kant's Latin gloss for "Vermögen", *facultas* at *Anth.* §7, 140); "Kraft" and its cognates with "power"; "Fähigkeit" with "capability". As Béatrice

Longuenesse notes, the distinction between "Vermögen" and "Kraft" is a kind of potentiality-actuality distinction.<sup>2</sup> For this reason, I have chosen to keep the traditional "faculty" for "Vermögen", and render "Kraft" with "power", since it seems (to my ear) to convey better this sense of "actuality". "Fähigkeit", which Kant uses primarily with regard to sensibility, I have rendered with the weaker "capability".

"Selbsttätigkeit" and "Spontaneität" have both been rendered "spontaneity", as I have not been able to discern any substantial difference in the terms, apart from etymology. "Lehrart" and "Methode" also seem to be synonyms of the this sort; hence, both are translated "method" (see Kant's Latin gloss in *Hechsel Logic*, 114). Kant distinguishes them from "Verfahren", which I translate "procedure". The same goes with "Überlegung" and "Reflexion": the former is simply the German word for the latter (as indicated by Kant's Latin gloss at A260/B316), and hence they have both been rendered "reflection". Also, Kant seems to refer to the activity of "reflection" in both a technical and an ordinary sense: the technical notion is expressed with *überlegen*, or *reflektieren*, and the ordinary notion is expressed with *nachdenken* or *nachsinnen*. I have rendered them all the same way ("reflect"), noting the German in brackets when it is one of the latter variants.

I have not distinguished between the notions of "cognition" and "knowledge" in my account of the *Critique*. However, I have tended to translate Kant's use of "Erkenntnis" (and its cognates) with "cognition" (and its cognates). Nevertheless, I opted for "knowledge" when that made for better English.

## 4.

In the translations, I occasionally employ square brackets to indicate the German term or phrase being translated. I do not italicize the bracketed German term or phrase unless it is italicized or stressed in the original. Square brackets are also used to indicate if I have elided a portion of the text in quotation, and occasionally for interpolations.

# 5.

I take "critical philosophy" and "criticism" to be technical terms. It is not the case that anything Kant wrote in the so-called "critical period" is a work of critical philosophy.

<sup>&</sup>lt;sup>2</sup> Kant and the Capacity to Judge, 7-8. Adelung's 1808 dictionary (*Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart*) draws a similar distinction between *Kraft* and *Vermögen*. *Vermögen*: "eine Art der Fähigkeit. Kraft ist im eigentlichsten Verstande das Bestreben, sein Vermögen zu äußern, das Vermögen in der Anstrengung, in der Thätigkeit betrachtet" (1095). See also Tetens, *Philosophische Versuche*, Twelfth Essay (v. 2), 70; and Schiller, *Über die Ästhetische Erziehung des Menschen*, XI: §8, to see this terminological distinction put into use.

Critical philosophy has rather severe methodological restrictions, as I suggest in this dissertation. In this dissertation, the term "critical philosophy" generally refers to the work carried out in the *Critique of Pure Reason* itself.

# 6.

On the term "capacity" (as in "mental capacity"): Kant uses a variety of terms which fall under this heading — e.g., "power" (*Kraft*), "faculty" (*Vermögen*), "capability" (*Fähigkeit*). These terms are not synonymous in Kant's usage. I deliberately do not render any of Kant's terms for mental capacity with "capacity" so that I can reserve the term for myself. I use it when I wish not to commit myself to any of Kant's terms, but want to invoke the basic idea of a mental capacity.

# 7.

I use "intellect" as a similarly generic term. Kant referred to the *understanding, the power to judge*, and *reason* as the "higher cognitive faculties" (*obere Erkenntnisvermögen*, A130/B169). In this dissertation, "intellect" refers to this higher cognitive faculty generically, without specification. Kant calls this whole shebang the "understanding in general" (A131/B169). The term "intellect" avoids the confusion of maintaining a distinction between broad and narrow senses of "understanding". To complicate matters further, Kant sometimes suggests that the entire higher cognitive power can be called *reason* (e.g., at A835/B863) — a suggestion that is borne out in the very idea that the entire *Critique* can be characterized as reason's self-critique, even though nearly half of the book is concerned with the understanding and the power of judgment.

Have the goodness to cast once again a fleeting glance upon the whole, and to notice that it is not at all metaphysics that the *Critique* is doing but a whole new science, never before attempted, namely, the critique of an a *priori judging* reason. Others have indeed also touched upon this faculty, like Locke and Leibniz, always in combination with other cognitive powers. To no one has it even occurred that this faculty is the subject of a formal and necessary, indeed extremely broad, science, requiring such a manifold of divisions (without deviating from the limitation of the mere consideration of that *uniquely pure faculty of knowing*) and at the same time — which is wonderful — *deriving* out of its own nature *all objects* within its scope, enumerating them, and proving their completeness by means of their coherence in a complete cognitive faculty. Absolutely no other science has been able to do this, that is, to develop a priori out of the mere concept of a cognitive faculty (when it is precisely determined) all the objects, everything that can be known of them, even what one is involuntarily but deceptively compelled to judge about them.

Immanuel Kant to Christian Garve
 7 August 1783 (10:340)

#### INTRODUCTION

My overarching concern in this dissertation is to understand how the *Critique of Pure Reason* could be the project that Kant advertises it to be. At the outset, Kant promises that it should fulfill reason's "most difficult" task: *self-knowledge*. Most generally, it is reason's attempt to legitimate its "pretension" to have knowledge *a priori* — i.e., independently of experience — of the domain of material nature. (It is reason's examination of its *theoretical*, as opposed to its *practical*, capacity.) This "pretension" manifests itself even with the empirical sciences of nature: it manifests itself whenever we take ourselves to make cognitive claims that hold with apodictic necessity. But Kant is particularly interested in the possibility of *metaphysical* claims about nature — claims that pertain to nature *as such*, and not merely to the lawful determination of some particular array of phenomena. For Kant, this is the problem of metaphysics.

In this dissertation, my particular interest lies in assessing the viability of the critical project itself. For the project that Kant advertises is puzzling all on its own, leaving aside the question of whether it fulfills its ultimate task of legitimating future metaphysical inquiry. Some of those puzzles are primarily exegetical. For example, the idea that the project is a "critique" suggests that its task would be normative or corrective, and the idea that it is a critique *of reason* suggests that it is principally *about* reason. But

we do not seem to learn much about reason until the latter half of the book, when the work become conspicuously corrective. In the first half of the book, reason does not seem to be on the scene at all; instead we find an ostensibly *descriptive* account of sensibility and the understanding (which Kant distinguishes from reason). Moreover, the advertised conception of the *Critique* — as reason's *self*-knowledge — has the odd implication that "reason" is the agent of this investigation, leading us from one claim to the next. It is not clear what this means, and where it would leave *us*, the flesh-and-blood readers of Kant's text. A more philosophical puzzle follows from the idea that the *Critique* is reason's *self-knowledge*. According to the *Critique*'s own doctrine, knowledge is limited to objects of possible experience; at the same time, "reason" figures in Kant's text as something that could never *be* an object of possible experience. So what sort of self-*knowledge* could this be?

Ultimately, these puzzles are methodological. To answer them we must understand how the *Critique* works. Commentators as diverse as Jean-Fran $\diamond$ ois Lyotard, Henry Allison, and Nicholas Rescher have suggested that the key to understanding the method of the *Critique* might lie in Kant's conception of *reflection*.<sup>1</sup> For the *Critique* quite obviously belongs to the philosophical tradition that was inaugurated by Descartes' attempt to ground all knowledge — and most importantly, metaphysics — in the certainty

<sup>&</sup>lt;sup>1</sup> This insight, such as it is, does not generally go beyond being merely suggestive or programmatic for further work. Some commentators look to the *Critique of Judgment* as the probable best source to understand the relevant notion of reflection: Lyotard (*Lessons on the Analytic of the Sublime*, 31) and George Agich ("L.W. Beck's Proposal of Meta-Critique") are two examples. Other commentators look to Kant's remarks about reflection in the Amphiboly (the Appendix to the *Critique*'s Transcendental Analytic): e.g., Marcus Willaschek ("Phaenomena/Noumena und die Amphibolie der Reflexionsbegriffe", 341) and Henry Allison ("Things in Themselves, Noumena, and the Transcendental Object", 45). Stephen Palmquist looks to both the third *Critique* and the Amphiboly appendix ("Knowledge and Experience", 55ff.). Finally, Nicholas Rescher looks to the "resource of reflection" as the only conceivable way that Kant could "make room for his own philosophical deliberations within his own cognitive framework" (*Kant and the Reach of Reason*, 141-2).

of the *cogito*. We could call this the tradition of "reflective philosophy", because it takes self-knowledge to be the primary task of philosophy. Continuing this tradition, Locke remarks upon the methodological peculiarities of the reflective endeavor: "The Understanding, like the Eye, whilst it makes us see, and perceive all other Things, takes no notice of it self: And it requires Art and Pains to set it at a distance and make it its own Object" (*Essay*, I.i.1). Hume, with a keener sense of the justificatory nature of the task, speaks of the need to "cultivate true metaphysics" (*Enquiry*, §I, 6) in order to free ourselves from chasing after spurious answers to rotten questions; this true metaphysics, he claims, begins with an inquiry into "the nature of human understanding", and offers "an exact analysis of its powers and capacity".

It is clear, at least, that Kant's *Critique* belongs to this broadly modern tradition of reflection. But in his conception of how reflection is possible, or philosophically meaningful, Kant differs profoundly from his predecessors. Yet it is much less difficult to appreciate what Kant would find philosophically unsatisfying in his predecessors' attempts to pursue reflective philosophy than it is to arrive at any positive account of Kant's own conception of reflection and its proper role in philosophical investigation. The task of my dissertation is to begin that positive account.

My strategies are various, and somewhat unorthodox. I have paid relatively little attention to the one place in the *Critique* where Kant talks about "reflection" explicitly, which is the Appendix to the Transcendental Analytic, known as the Amphiboly of the Concepts of Reflection. For me, it has proven to be the most tantalizing stretch of the *Critique*: in it, Kant remarks that "transcendental reflection is a duty that no one can repudiate if he wants to judge about things a priori" (A263/B319), coyly suggesting that

this "transcendental reflection" must be most profoundly connected to the project of critical philosophy itself. For in the *Critique*'s Prefaces, Kant emphasizes that critical philosophy — reason's examination of its own capacity — is a self-incurring requirement of reason — a "duty", if you will. But Kant only tosses this remark off in preparation for his polemic against Leibnizian ontology, and the rest of what Kant says about reflection in the Amphiboly is difficult to dissociate from that immediate task.<sup>2</sup>

My approach is unorthodox because I pursue what is generally thought to be the least defensible aspect of Kant's own conception of the critical project, and through it I find my exegetical bearings. The *Critique* is supposed to be some kind of *science*. As it draws to a close, Kant looks back over the whole and dubs it reason's "scientific and fully illuminating self-knowledge" (A877/B849). In the *Prolegomena*, Kant refers to the *Critique* as a "whole new science", and to Christian Garve he called it the "science [...] of an a priori judging reason" (10:340). Kant understands the very idea of what a science is in methodological terms; hence, Kant's general conception of proper scientific method should tell us *something* about the method of the *Critique*.

Thus, Chapter 1 focuses partly on Kant's general conception of proper scientific method. At the general level, Kant's conception of scientific method is not particularly innovative: he is drawing on an existing conception of scientific method that can be traced back to the ancients. Scientific knowledge is systematic knowledge "from principles": these principles must be discovered before the knowledge in question can be

<sup>&</sup>lt;sup>2</sup> Only in the very late stages of this project have I begun to make sense of the Amphiboly as a whole. I do anticipate working on the Amphiboly more in the future, particularly with the aim of assessing how Kant's polemic against Leibniz relates to the conspicuously *metaphysical* project advanced in the Analytic of

established as a system. Thus, there are two fundamental aspects of scientific method according to this general conception of it: an "analytic" path *to* these principles, and a "synthetic" laying out of the system on the basis of these principles.

But Kant is not simply pressing an existing conception of scientific method into service. For it would not be unfair to say that Kant was obsessed, for the entire duration of his philosophical career, with the issue of proper scientific methodology: one might even say that it is the substratum on which he exercised his philosophical genius. His "Prize Essay" of 1763 is devoted to the topic of what the proper method of *metaphysics* ought to be. There Kant argues that its proper business is the analysis of concepts, and thus its proper method is, in turn, *analytic*. Decades later, Kant is still working on the same topic: for this is the topic of the *Critique*, which is supposed to be a "treatise on the method" of a future metaphysics (Bxxii). But the position advanced in the 1763 Prize Essay had been abandoned in the interim. Kant's mature account of the proper method of metaphysics takes shape as critical philosophy itself.

Although Kant's general conception of scientific method may not be particularly innovative, critical philosophy, as Kant insists in the *Prolegomena*, *is*. It is a "whole new science" that claims to owe nothing crucial to any existing science, or to the philosophical tradition that precedes it. Regardless of whether this claim can really be defended, it does suggest that the general conception of scientific method that informs the *Critique*'s own project will not, on its own, indicate what is truly innovative about critical philosophy. While I urge in this dissertation that Kant's general conception of scientific investigation is an indispensable guide to the text of the *Critique*, it alone is clearly not

Principles — and to see how all of this bears on whether and how, precisely, the work of the *Critique* is "therapeutic".

decisive. It brings the signposts of the text into clearer focus, which helps us to recognize better what the *Critique*'s presuppositions are, and how we are driven to the work's conclusions.

Thus, my initial account of Kant's general conception of scientific method is only a point of departure. The core task of Chapter 1 is, rather, to clarify the very *idea* of a "critique of pure reason". In order to move beyond Kant's general conception of scientific method, I ask that we consider what makes the *Critique* the particular science that it is. It is a reflective project, a science of self-knowledge, which "takes nothing as given except reason itself", as Kant claims in one of the few passages where he remarks explicitly on the method of the *Critique* (*P* §4, 274). This suggests that the work would begin with some preliminary account of reason. Where is this preliminary account, and what does it tell us? We know what is at stake in the *Critique*: namely, reason's *theoretical* cognitive capacity — more specifically, its capacity for *scientific* knowledge of material nature, and principally *metaphysical* scientific knowledge of material nature. Given this, where should we look for the preliminary conception of reason that the *Critique* itself presupposes?

To answer this, I turn to the account of scientific — i.e., rational — cognition that is the express topic of the *Critique*'s Preface. There Kant considers the paradigm cases of logic, mathematics, and physics, and raises his well-known question about the status of metaphysics as a science. By considering Kant's examples, I suggest that the point of the Preface is to introduce some preliminary conception of reason that critical philosophy will take as given. As Kant describes his examples, the cognition in question seems to turn on some kind of self-consciousness that could be attributed to our cognitive capacity.

The precise nature of this self-consciousness is not yet understood. But we are meant, I suggest, to suppose that this self-consciousness might be the essential, or fundamental, characteristic of reason.

The work of the first two chapters is largely programmatic: in both, my aim is to clarify what the critical project is *supposed* to be, so as to outline constraints for a faithful reading of the text. The aim of Chapter 2 is to provide a preliminary account of the metaphysical project of the *Critique*, focusing on how a reflective project of self-knowledge could ultimately yield metaphysical conclusions about material nature. To account for this, I draw on Kant's remarks about the role of reflection in the development of conceptual capacities from the *Jäsche Logic*. With this, our appreciation of Kant's conception of reflection becomes somewhat less generic and preliminary, and more attuned to the specific parameters of Kant's critical philosophy.

Chapter 3 is devoted to clarifying Kant's appeal to the *spontaneity* of the mind. I begin by examining the idea that spontaneity is proper to the intellect (broadly understood), as opposed to sensibility. Kant's conception of the intellect is multi-faceted: reason and the understanding are the two principal capacities that fall under this broad heading. Therefore, we must come to terms with Kant's distinction between reason and understanding in order to grasp his appeal to spontaneity properly. Continuing the theme from Chapter 1, I argue that the fundamental activity of reason is *reflection*, while the fundamental activity of the understanding is *judging*.<sup>3</sup>

<sup>&</sup>lt;sup>3</sup> This approach to Kant's distinction need not be viewed as standing at odds with the more typical reading which focuses on Kant's idea that reason plays a "regulative" role in theoretical cognition, while the

Reflection thus emerges as one aspect of Kant's appeal to spontaneity. In this chapter, I distinguish Kant's conception of reflection from that of his rationalist and empiricist predecessors. I show that Kant's conception of reflection is distinctive because it is *non-introspective*. Reflection is attributed to the *spontaneity*, not the receptivity, of the mind. Hence Kantian reflection is not to be conceived according to any kind of perceptive model. Reflection, in Kant's account, is not a response to activities or operations of the mind that are themselves independent of any possible reflective awareness of them. Rather, the "operations" in question — exercises of judging, and even perception, I argue — are reflective through and through.

The dissertation culminates in Chapter 4, with a point-by-point account of the Critique's synthetic method and the central role that the Transcendental Deduction plays in carrying it out. In this chapter, I draw on the suggestion from Chapter 1 that the preliminary conception of reason is laid out in the Critique's front matter (i.e., the Prefaces and Introduction). In the front matter, the problem of metaphysics is reformulated as the problem about the possibility of synthetic a priori judging; this is then dubbed the "general problem of pure reason". Kant's initial account of the problem, I argue, dictates a preliminary thesis about sensibility and understanding as heterogeneous elements of pure theoretical reason. These preliminary considerations underwrite what I call a "strategy of isolation". The first parts of the Critique operate according to this

understanding plays a "constitutive" role. This standard reading of Kant's distinction is not one that we could possibly have at the outset, or in the beginning stages, of the work: it is, rather, the result of the work at the end of the day. (The account of the proper, "regulative" role of reason only comes on the scene towards the end of the Transcendental Dialectic, and in the Appendix to the Transcendental Dialectic.) Since my concern is with the *method* of critical philosophy, I focus on how the project gets under way in the preliminary stages.

strategy, yielding separate accounts of the constitution of sensibility (in the Transcendental Aesthetic) and of the constitution of the understanding (in the presentation of the table of categories). But given that sensibility and understanding must *cooperate* in order to yield knowledge, the strategy of isolation can only be provisional. We need an argument that will bring the disparate elements together: without it, the status of sensibility and understanding as cognitive capacities hangs in the balance. Hence, the synthetic method.

I should conclude these introductory remarks with an apology. My reading of the *Critique* focuses almost entirely on the second edition version. To a large extent, the two versions of the *Critique* — of 1781 and 1787, respectively — are identical. The two parts of the book that Kant completely rewrote were the Preface and the Transcendental Deduction, both of which figure prominently in this dissertation. In each, I have found something crucial for my story. Since the second edition Preface is expressly devoted to the topic of rational cognition, it naturally serves as an initial illustration or sketch of the preliminary conception of reason that Kant's scientific methodology (as applied to the task at hand) tells us that we should expect. Moreover, the general methodological strategy of distinguishing the paths to and from the first principle of the science has a particularly neat application to the second edition Deduction, in which Kant announces half-way through that we have just come to the "beginning" — i.e., the principle, or *arche* — of the deduction (and presumably the science itself).

To some extent I can only explain, and cannot fully justify, my preoccupation with the second edition *Critique*. The explanation is simply the limitation of my own

energy and skill: the second edition Deduction had a natural place in my story, and my hands were quite full trying to come to terms with it. (They still are.) I recognize that I should examine the first edition Deduction more closely — to see whether, and possibly how, it corroborates my overall thesis.

But perhaps I can indeed justify my preoccupation with the second edition version. As I argue in this dissertation, Kant's critical philosophy marks an innovation in philosophical method of the highest order. It is not until the 1783 *Prolegomena* that Kant discusses explicitly, and in published print, the methodological conception of the *Critique*. More helpful remarks about methodology follow in the 1785 *Groundwork for the Metaphysics of Morals*, and again the Preface to the 1786 Metaphysical *Foundations of Natural Science*. Quite simply, it is my belief that Kant was becoming clearer and clearer about the methodological framework of critical philosophy as this decade wore on. For this reason we should focus our attention on the second edition *Critique* if we want to understand how that methodological framework makes the substantive insights of critical philosophy possible.

### CHAPTER 1 KANT'S IDEA OF A CRITIQUE OF PURE REASON

This sort of investigation will always remain difficult, since it contains the *metaphysics of metaphysics* [...]. — Kant to Marcus Herz, May 1781

He appeals to the *whole* in judging the world. For this we need knowledge that is no longer *made of pieces*. To argue from the whole to a fragment is like arguing from the unknown to the known.

—Johann Georg Hamann on Kant, to Johann Lindner, October 1759

1.

In the Preface to his 1781 *Critique of Pure Reason*, Immanuel Kant remarked that self-knowledge is reason's "most difficult" task (Axi). At the time that he wrote, Kant thought that it was also reason's most pressing task. "Ours is the proper age of *criticism*," he claimed, registering what may be the high-water mark of Enlightenment philosophy. This philosophical tradition is characteristically concerned with self-knowledge: Descartes' *Meditations*, Locke's *Essay*, and Hume's *Treatise* — to pick three notable examples — all advertise themselves, in one way or another, as such. Of all his predecessors, Kant thought that Hume alone understood that this self-knowledge must be *evaluative*, or "critical". But Hume's attempt at the project was limited: he may have brought the dogmatic metaphysician to self-knowledge, but not reason itself (A763/B791).

He dealt with particular claims of the dogmatic metaphysical tradition, more or less as they occurred to him. He submitted the "*facta*", or particular deeds, of reason "to evaluation, and when necessary, to blame" (A760/B788).

The proper era of criticism only arrived with the recognition that the critical examination of reason must be absolutely universal. We must submit the "entire faculty" of reason to evaluation (A761/B789). It is conceived as a project of self-knowledge: "scientific and fully illuminating self-knowledge [Selbsterkenntnis]" (A849/B877).<sup>1</sup> The *Critique* is reason's complete account of its theoretical capacity, particularly with respect to its "suitability" for a priori knowledge.<sup>2</sup>

Moreover, critical philosophy is supposed to establish the viability of metaphysics as a science; it is the science that promises to make metaphysics a science. Yet the *Critique* sets out with the recognition that — as a matter of historical fact, as it were there is no viable metaphysics at all; what goes by the name of 'metaphysics' is thoroughly undercut by the fact that everything said under its banner is surrounded by bitter, and apparently unresolvable, controversy. For Kant, there is *no extant metaphysics* 

<sup>&</sup>lt;sup>1</sup> See also *Metaphysik Mrongovius*, on transcendental philosophy: "It is a form of self-knowledge [sie ist eine Art von Selbst Erkenntniß]" (29:756). And further on: "transcendental philosophy [...] does not say anything about objects a priori, but rather investigates the faculty of understanding and of reason a priori [...]; thus it is a self-knowledge of understanding and of reason" (784). See also *Metaphysik Volckmann* (28: 392).

<sup>&</sup>lt;sup>2</sup> In the *Critique of Pure Reason*, reason examines its theoretical capacity (i.e., its capacity to judge about objects that can only be given in experience), while in the *Critique of Practical Reason*, reason examines its practical capacity (i.e., its capacity to determine the will). This suggests that critical philosophy as such — of which Kant offers three 'episodes' — is rational self-knowledge. The first *Critique* is concerned to account for the rational source of the supreme law of morality. So what about the third episode, the *Critique of Judgment*? In what sense is it a project of rational self-knowledge? This is a difficult question. In the Introduction to the third *Critique*, Kant suggests that its function is to unify the first two *Critiques*. Perhaps we can understand this in the following way: the contribution of the third *Critique* is to demonstrate how the spontaneity of the understanding (which is the central topic of the first *Critique*) stem from one and the same faculty of reason.

*at all* prior to the critical investigation. The thought behind this possibly outrageous claim is simply that we cannot say anything at all in metaphysics unless we also appreciate the basis on which we speak. Bitter dispute among the self-proclaimed metaphysicians stood as proof positive for Kant that this reflective validation was still missing. Critical philosophy cannot draw from the tradition, but only from the "sources of reason" itself (*P*, 255). The recognition that metaphysics must be "critical" leads directly to the idea that it must rest on reason's clear appreciation of itself as the source of metaphysical knowledge. Critical metaphysics is reason's self-knowledge.

My aim in this dissertation is to present an account of the "scientific and fully illuminating self-knowledge" that Kant claims is manifest in the *Critique*. The work begins here, with a precise explication of Kant's *idea* of the project, and the method that Kant supposes is required in order to carry it out. Unfortunately, Kant says remarkably little about what exactly critical philosophy is, and how it works. However, the *Critique* is supposed to be a member of a general class: scientific (or rational) cognition. Granted, it is a very special case. But the difficulties of this special case may be approached slowly, and with surer steps, if we start with the idea that it is one science among many.

The goal of this chapter is to come up with a working grasp of the idea that the *Critique* is a project of rational self-knowledge. I will begin (in §2) with a notorious problem about the viability of reason's self-knowledge in the *Critique*, which has been raised by a variety of commentators — but answered, I think, by almost none.<sup>3</sup> This problem is raised by P.F. Strawson in his influential work on the *Critique*. Although it

<sup>&</sup>lt;sup>3</sup> One of the better guides to conceiving of the *Critique* as reason's self-knowledge can be found in some of Tyler Burge's work on self-knowledge — even though it is not offered as a direct commentary on Kant. Still, it seems to me that Burge's insights in these papers ("Reason and the First Person", "Our Entitlement

remains essentially unanswered by Strawson, it nevertheless determines much about his approach to Kant's text: for it seems that Strawson allows the apparent intractability of this problem to justify his complete disregard of Kant's own conception of the *Critique* as the scientific expression of reason's self-knowledge. I will briefly consider Strawson's work — not so much with the aim of mounting a sustained attack on it, but rather to clarify what we are missing if we follow Strawson and refuse to take the *Critique* on its own terms. The answer, in short, is that we forfeit any chance of having an adequate understanding of the methodological framework of the *Critique*. For Kant, the distinguishing mark of scientific cognition is that it follows a proper *method*. So if we want to understand the method of the *Critique*, we might begin by coming to terms with Kant's idea of it as some kind of science.

Thus, I continue by examining Kant's conception of proper scientific method (in §3). Specifically, I am concerned with the role of analytic and synthetic arguments, as well as with the role of *principles*, in scientific cognition as Kant understands it. While this will leave us with a preliminary grasp of what scientific method involves according to Kant, it alone does not shed much light on whatever might be peculiar about the science of critical philosophy. Given that critical philosophy is the science in which reason is expressly concerned with its own capacity, we must set out with some preliminary idea of what the capacity of reason *is*. I address this issue (in §4) by considering Kant's conception of rational cognition. The context of this discussion will also allow me to explain, briefly, Kant's idea of critical philosophy as "formal philosophy".

to Self-Knowledge", and even his earlier "Individualism and Self-Knowledge") must stem from close study of Kant.

This introductory chapter outlines constraints on our reading of the *Critique*. What is critical philosophy? Or what, at any rate, does Kant *take* it to be? What does critical philosophy share with other modes of scientific cognition? And how does it differ from other modes of scientific cognition? How, at the end of the day, can we come to terms with the idea that the *Critique* is the scientific expression of reason's self-knowledge? I hope that by pursuing answers to these questions we will not only see how to take the *Critique* on its own terms, but also recognize the importance of our doing so.

#### 2.

Let us begin with the most obvious difficulty contained in the idea that the *Critique* is reason's self-knowledge. According to the *Critique*'s own doctrine, knowledge is limited to objects of possible experience. At the same time, reason figures in the *Critique* as something that could never be an object of possible experience.<sup>4</sup> *So what sort of self-knowledge is this?* Since critical philosophy is advertised as reason's self-knowledge, this puzzle — as long as it remains unresolved — threatens to pull the plug on the entire enterprise.

Commentators have generally responded by looking the other way. Rather than give serious attention to the idea that the *Critique* is a project of rational self-knowledge, it is easier to focus on aspects of Kant's argument and disregard the idea of the whole.<sup>5</sup> In the Anglo-American tradition, this has proved to be a convenient way of dealing with

<sup>&</sup>lt;sup>4</sup> See, e.g., A556/B584; for more on this issue, see Chapter 3.

<sup>&</sup>lt;sup>5</sup> Susan Neiman (*The Unity of Reason*, Chapter 5) and Richard Velkley (*Freedom and the End of Reason*) address the idea at some length, however. Their commentaries, which both lie somewhat outside of the Anglo-American mainstream, focus on Kant's conception of the place of critical philosophy in the "history" of pure reason, so that the *Critique* is viewed as a kind of completion of human reason. The initial reading that I offer of this conception of Kant's project (at the end of this chapter) is in this one respect akin to theirs.

certain unwelcome consequences of the project as Kant advertises it. Kant introduces the project of self-knowledge as reason's "*critique of its own faculty*" (Bxxxv): the idea that the *Critique* is a project of self-knowledge naturally makes appeal to the notion of a mental "faculty". As a result, this overarching conception of the project is a source of discomfort to many philosophers in the Anglo-American tradition, who are — rightly, I think — concerned to avoid a "psychologistic" reading of the *Critique*.<sup>6</sup>

The discomfort with the Kant's talk of faculties is felt most acutely, perhaps, in Strawson's seminal work, *The Bounds of Sense*. I turn to Strawson's work now, in an attempt to illustrate how our view of the *Critique* may be distorted if we fail to come to terms with Kant's idea that it is reason's scientific self-knowledge.

Strawson thought that the good work of the *Critique* is undermined by its putatively "psychological" idiom. The provenance of this way of speaking is the fraudulent enterprise of "transcendental psychology": an unsightly and cumbersome — but fortunately benign — tumor that should be excised from the work.<sup>7</sup> For Strawson, the problem of critical self-knowledge looks something like this:

Kant says that we can only have knowledge of objects of possible experience. He also says that the *Critique* is a project of self-knowledge: this can only mean that reason investigates itself as a cognitive power. Kant understands this in terms of a search for the subjective "sources" of cognition. But if the investigation of these "sources" is also supposed to

<sup>&</sup>lt;sup>6</sup> Patricia Kitcher has pursued an avowedly "psychological" reading of the *Critique* that aims to recognize that Kant's talk of faculties has to be understood in its proper methodological context. (See *Kant's Transcendental Psychology*, Chapter 1.) I certainly agree with her general point about how to approach Kant's talk of faculties; but I disagree with her account of the methodological framework of the *Critique* (see Chapter 3, note 2).

<sup>&</sup>lt;sup>7</sup> The term "transcendental psychology" is used in the A-edition Paralogisms to refer to *rational psychology*'s status as one of the "apparent sciences", along with cosmology and theology (A397; see also A350, 351, 361, 367). That is, the term "transcendental psychology", as used by Kant, singles out the very thing which is submitted to criticism in the Paralogisms, and not Kant's own project. The term, as far as I can tell, is omitted from the B edition.

be self-*knowledge*, then its talk of the "sources" of cognition will be tainted by the explanatory models of empirical psychology, even though Kant insists that there is nothing empirical about his project, and emphatically denies that it is a work of empirical psychology.<sup>8</sup>

In short, Kant's confusion about the nature of his own project led him to the talk of mental faculties, as if it were the only way to give the semblance of carrying out the impossible mission he set for himself. The good work of the *Critique* is concerned to make explicit the "limiting or necessary general features of experience" (15). But when Kant comes across such limits, he unfortunately goes on to "declar[e] their source to lie in our own cognitive constitution". In other words: with every recognition of a limiting or necessary feature of experience, a faculty of some kind is arbitrarily declared to be its source.<sup>9</sup> The stipulation of these cognitive sources or faculties is meant to be *explanatory*.<sup>10</sup>

Strawson is, in effect, reprimanding Kant for asking after the "sources" of knowledge — the very question which, as Kant announces in the first paragraph of the *Critique*'s Introduction, eluded his empiricist predecessors. This question about the sources of knowledge, Strawson supposes, is "incoherent in itself" and "masks, rather than explains, the real character of his inquiry" (16). To understand the *Critique* properly, we must cut out the loathsome growth: the true insights of the work are all

<sup>&</sup>lt;sup>8</sup> "Transcendental psychology," Strawson writes, is a "strained" and "misleading" analogue of empirical psychology (*The Bounds of Sense*, 15).

<sup>&</sup>lt;sup>9</sup> "It is true that Kant thought of himself as investigating the general structure of ideas and principles which is presupposed in all our empirical knowledge; but he thought of this investigation as possible only because he conceived of it also, and primarily, as an investigation into the structure and workings of cognitive capacities of beings such as ourselves. The idiom of the work is throughout a psychological idiom. Whatever necessities Kant found in our conception of experience he ascribed to the nature of our faculties" (19).

<sup>&</sup>lt;sup>10</sup> "[T]his doctrine [of the faculties] he considered indispensable as an explanation of the possibility of knowledge of the necessary structure of experience" (15-6).

found in an "analytical argument which is in fact independent of [the doctrine of the faculties]" (16).

The upshot, for Strawson, is that the real work of the *Critique* can be represented as an analytical argument. The general form of the analytical argument which Strawson finds in the *Critique* involves asking after the conditions of the possibility of our having a viable conception of some x. On Strawson's reading, the *Critique* is about articulating a "possible general structure of experience", which we arrive at through analyzing our given conception of x in terms of the conditions of its possibility. Strawson admits as given some conception of experience in general, and aims to arrive at its most general features; Kant, Strawson claims, is an "analyst of the conception of experience in general" (52).

Strawson is not wrong to say that Kant is engaged in an analysis of experience in general.<sup>11</sup> Where Strawson goes wrong is in his understanding of how this analysis is carried out and what it is responsible to. To see this, let us consider Strawson's conception of the critical procedure in more detail. According to Strawson, we arrive at these most general features of experience when we discover that some candidate cannot be severed from our conception of experience without obliterating that conception altogether. Here is a particular instance of his procedure in play:

We are confronted not merely with the thought of an intimate link between the idea of particular items capable of being encountered in experience and the idea of their being temporally and spatially ordered items. We are confronted with the thought of this link being so vital that it cannot be

<sup>&</sup>lt;sup>11</sup> In a letter to J.S. Beck (20 January 1792), Kant himself referred to an unspecified stretch of the *Critique* — possibly the Transcendental Aesthetic and the Transcendental Analytic — in very similar terms, as an "analysis of experience in general" (11:313); he refers to "this […] analysis" again in the same letter (315). I discuss this remark in Chapter 4.

broken without nullifying the whole conception of experience [...]. To abstract altogether from the idea of time, of temporal sequence, while preserving that of experience in general we may admit at once to be a task beyond our powers. (50)

The sort of reflection Strawson finds in the *Critique* involves taking as given some conception of experience; we then find some candidate limiting feature and ask if our initial conception of experience would survive the excision of that feature. This certainly seems to be an analysis of our concept of experience: when a candidate for a constituent concept comes on to the scene, it is tested by seeing if an excision of this candidate constituent concept would undermine the viability of our given concept. We are responsible, in such an analysis, to the "limits to what we can conceive of, or make intelligible to ourselves, as a possible general structure of experience" (15).

The sort of procedure Strawson presents as the good analytical work of the *Critique* Kant would probably condemn as a mere "groping about", a *Herumtappen*: we stumble upon certain candidates for limiting features of experience, and we test them against the "limits to what we can conceive of". It is, perhaps, a "mechanical procedure", which Kant distinguishes from a properly *systematic* procedure: the former discovers its concepts (its "limiting features") "as the opportunity arises" (*bei Gelegenheit*), and not according to a principle (*nach einem Prinzip*) (A67/B92).

According to Kant, rational cognition begins with an idea of "the whole" — that is, an idea of the entire domain of the subject matter at hand. An idea of the "form of a whole of cognition" is a preliminary requirement of systematic cognition in general, which Kant attributes to reason (A645/B673). This idea of the whole "precedes the determinate cognition of the parts". The "parts" or "elements" of a science are determined according to a principle; this principle (or set of principles), Kant implies, is contained in the preliminary idea of the whole of the science (A645/B673). Kant contrasts scientific rational cognition (which is necessarily systematic) with what he characterizes as a mere "aggregate" of knowledge that is "heaped up" without the guidance of a unifying principle.

The scientific rational concept contains [...] the end and the form of the whole that is congruent with it. The unity of the end, to which all parts are related and in the idea of which they are also related to one another, allows that the absence of any part can be noticed from our acquaintance with the rest; and there can be no contingent addition or indeterminate magnitude of perfection that does not have its boundaries determined a priori. The whole is thus articulated [gegliedert] (*articulatio*) and not heaped up [gehäuft] (*coacervatio*); it can grow internally [...] but not externally, like an animal body whose growth does not add limbs but rather makes each limb stronger and fitter for its end without altering the proportion. (A833/B861)

So it turns out that Kant would associate the idea that the *Critique* is a science (articulated with the aid of a unifying principle) with the idea that its results are *complete*. If the *Critique* is a science, then it must begin with an idea of the whole, and articulate the parts of this whole according to a principle, or perhaps a set of principles. This "whole", presumably, is pure reason itself in its theoretical capacity. Its elements or parts would be determined by a principle that would be uncovered through the examination of the idea of the whole with which we begin.<sup>12</sup> In Chapter 4, I will present an account of this;

<sup>&</sup>lt;sup>12</sup> Remarks in the *Prolegomena* underscore the relevance of this general conception of scientific cognition to the particular case of the *Critique*, as reason's investigation of its own capacity. Reason, he says, "is like an organized body, the end of each member can be derived only from the complete concept of the whole. Thus one can say of such a critique [of pure reason] that it would never be reliable if it is not *entirely complete*, up to the smallest elements of pure reason, and that one must make out and determine either *everything* or *nothing* of the sphere of this faculty" (*P* 262-3).

for the moment, I am more concerned to explain why Strawson's approach must fail to illuminate Kant's text in an adequate way.

Strawson gives no indication of how we would know when the analysis of experience in general is complete. According to Kant, arriving at a complete articulation requires that we have a clear idea of the whole and the principle according to which its elements can be articulated. When Strawson overlooks the idea that the *Critique* is reason's self-knowledge, he forfeits this required conception of the whole, and its basic principle. We are supposed to be starting with the idea of pure theoretical reason as a capacity to judge synthetically and a priori.<sup>13</sup> Strawson, however, urges that we set this issue of the possibility of synthetic a priori judging aside, since there is no "satisfactory theoretical account of the dichotomy between analytic and synthetic a priori propositions" (43).

Obviously we will need to consider Kant's conception of systematic rational cognition in more detail. Nevertheless, it should be clear that if the *Critique* is supposed to be a science, or systematic rational cognition, then Strawson does not acknowledge this aspiration in the least. He is correct to see analytic arguments as playing a crucial role in carrying out the work of the *Critique*, but he fails to appreciate them in their proper methodological context. In order to understand what that context is, we should draw on the idea that the *Critique* is supposed to be a special kind of science — the science in which reason is expressly concerned with its own capacity. Now, Kant understands the very idea of what a science is in methodological terms. Once we understand Kant's conception of proper scientific method, we might be able to draw

<sup>&</sup>lt;sup>13</sup> I will discuss the starting point of the *Critique* in more detail in Chapter 4.

inferences about the special case of the *Critique*. Thus, I turn now to Kant's general conception of scientific method.

#### 3.

For Kant, anything which is to count as a proper *science* — which is distinguished as a system of knowledge both from what is mere "art", and also from mere aggregates of knowledge — "must be established according to a method" (*Jäsche* §95, 139). From at least as early as the 1763 Prize Essay, Kant was concerned to identify the method proper to philosophical investigation, and most importantly the method proper to its "highest part", metaphysics.<sup>14</sup> There he argued that philosophy ought to be principally engaged in the analysis of concepts, and claimed that the proper method of philosophical inquiry was "analytic" for this reason. We cannot confuse the method proper to philosophical inquiry with the "synthetic" method that is proper to mathematical inquiry. Philosophy does not proceed by the construction of concepts; it is not, by the lights of the 1763 Prize Essay, "synthetic".

Obviously there is some shift in Kant's thinking about philosophical method between 1763 and 1781 — or at any rate, by 1783, when he claimed in the *Prolegomena* that the method of the *Critique* is "synthetic" (263; 274). The method of the

<sup>&</sup>lt;sup>14</sup> According to Ernst Cassirer, Kant's concern with scientific methodology can be traced as far back as to Kant's first published work: "What is noteworthy in this maiden paper is that the first step Kant takes into the realm of natural philosophy immediately turns into an inquiry into its method. His entire critique of the Leibnizian conception is subordinated to this point of view; at one point he expressly explains that he is not so much combating Leibniz's result as its foundation and derivations, 'not actually the facts themselves, but the *modus cognoscendi*" (*Kant's Life and Thought*, 27). See also Kant's unpublished remark from the 1760s: "Alle Betrachtung über die Methode ist das Wichtigste einer Wissenschaft. — Es ist wenig daran gelegen, ob einige Sätze der reinen Philosophie über das Objekt wahr oder falsch sind; es ist wichtiger, ob sie in der gehörigen Methode gedacht sind" (Erdmann, *Reflexionen Kants zur Kritik der reinen Vernunft*, 183-4).
*Prolegomena* itself, he said, is "analytic" (263). Kant does not conceive of the *Prolegomena* as an independent inquiry; it is only a presentation of the results of the *Critique* in an analytic form. As we shall see, the *Prolegomena* is "analytic" because it takes as given certain sciences, and gradually uncovers their "sources" in certain a priori representations.

It is reasonably clear that Kant could not suppose that the method of the *Critique* is "analytic" in the terms of the Prize Essay: the *Critique* is not principally engaged in the analysis of concepts. For conceptual analysis, according to Kant, never draws to a definitive close. The *Critique* cannot be principally concerned with the definitional analysis of non-arbitrary (given) concepts,<sup>15</sup> since "one cannot become certain through any test whether one has exhausted all the marks of a given concept through a complete analysis", and thus "all analytic definitions are held to be uncertain" (*Logic* §104, 9:142). The same point is made in the *Critique* itself: "philosophical definitions [...] come about only analytically through analysis [analytisch durch Zergliederung] (whose completeness is never apodictically certain)" (A730/B758). The *Critique*, therefore, is not engaged in a definitional analysis of given concepts. For if the aim of the first *Critique* is to establish a *foundation* for "any future metaphysics", then the project could not rest on such inherently ongoing work.<sup>16</sup> The *Critique* "must lay before us a complete enumeration of

<sup>&</sup>lt;sup>15</sup> "All given concepts, be they given a priori or a posteriori, can be defined only through *analysis*" (*Logic*, §104).

<sup>&</sup>lt;sup>16</sup> Susan Neiman thinks of this as the "constitutive" conception of the *Critique* (critical philosophy as the foundation for a future metaphysics), and supposes that it is in some tension with what she calls the "regulative" conception of the *Critique*, according to which metaphysics is *replaced* by critical philosophy (*The Unity of Reason*, Chapter 5). She herself plumps for the "regulative" conception of the *Critique* as the preferred interpretation. However, it seems to me that the "constitutive" conception predominates. The *Critique* is supposed to be a reflective project of self-knowledge that leads to a metaphysical claim about the possibility of material nature: the reflective work culminates (at the end of the Transcendental Deduction) with the justification of a conception of nature as a totality — nature "regarded formally" ("as *natura formaliter spectata*", B164) — the account of which is spelled out in the Analytic of Principles. It

all the ancestral concepts that constitute the pure cognition under consideration", in order to fulfill its foundational aspirations. It does not contain "the complete analysis of these concepts"; any analysis in the *Critique* would have to be "purposeful", which Kant says means that it would be carried out for the sake of making a certain synthesis possible (A13-4/B27-8).<sup>17</sup> These remarks, which shall become clearer once we come to a better understanding of the method of the *Critique*, already suggest that if there is analysis in the *Critique*, Kant is not inclined to think of it as conceptual analysis.<sup>18</sup>

So what exactly changes for Kant between 1763 and 1781? In the Prize Essay, Kant's insistence that philosophical inquiry must be limited to the analysis of concepts is keyed to his desire to distinguish the nature of philosophical inquiry from that of mathematical inquiry. In the *Critique*, he is just as keen on making this distinction (A712-38/B740-66). Kant's injunction not to confuse the method of philosophy with that of mathematics is unshaken by his later idea that critical philosophy proper, or at any rate the first *Critique*, is carried out according to a "synthetic" method. The *Critique* is not concerned with "arbitrary" concepts and their construction, as mathematics is, but rather with "intellectual" concepts — concepts drawn "from the nature of the understanding" and with the possibility of their application to material nature.<sup>19</sup>

strikes me that such a project, if successful, would have to be conceived as "constitutive". However, this is not to deny that the *Critique* may be meant to serve some "regulative" function for future metaphysicians as well.

<sup>&</sup>lt;sup>17</sup> Kant says that an analysis of these concepts would not be "purposeful [zweckmäßig], since it would not contain the difficulty which is encountered in the synthesis for the sake of which the entire *Critique* actually exists" (A14/B28). The significance of this remark will only become evident in Chapter 4.

<sup>&</sup>lt;sup>18</sup> See also Kant's introductory remarks to the Analytic of Concepts, where he stresses that he is not engaged in an analysis (*Analysis*) of *concepts* (which he associates with the "procedure commonly found in philosophical investigations"). Rather, he is engaged in the "still rarely attempted articulation [Zergliederung] of the faculty of understanding itself" (A65/B90).

<sup>&</sup>lt;sup>19</sup> For this taxonomy of "concepts" into arbitrary, intellectual, and empirical, see *Logic* §5, 94.

Kant may be drawing on ancient sources to arrive at his idea of the *Critique*'s method. At any rate, the idea that analysis and synthesis figure as complements in certain methods of investigation has a pedigree tracing back to Plato and Aristotle. Aristotle refers to Plato's concern to distinguish the argument which takes us to first principles (the analysis) from the argument that takes us *from* first principles (the synthesis), "just as we distinguish on a race course between the way from the judge to the turning-point, and the way back again".<sup>20</sup> Somewhat like Kant centuries later, Plato attributes this method to "reason itself" — reason ascends first to "the unhypothetical first principle of everything", before it "reverses itself and [...] comes down to a conclusion without making use of anything visible at all".<sup>21</sup> And even Aristotle, who is not so keen on the idea of a "first principle of everything", still supposes that proper method requires that we discover the principles that constitute each distinct domain of inquiry. This means, for Aristotle, that there are two different sorts of "beginnings" to investigation. One is a conditioned beginning, in which we are concerned with the "things that are less clear by nature, but clearer to us". From here we "move on to things that are by nature clearer and more knowable": analysis takes us from our knowledge of composites to the "elements and principles" of a science. Systematic knowledge arises from our grasp of the relevant principles.<sup>22</sup> According to Aristotle, the method capable of producing systematic knowledge of nature involves complementary movements of analysis and synthesis: analysis takes us to the principles which are employed in the systematic expression of the cognition in question. When we "return" from the principle, we return to the domain of

<sup>&</sup>lt;sup>20</sup> Nicomachean Ethics A, 1095a31-b1 (Ross/Barnes translation).

<sup>&</sup>lt;sup>21</sup> *Republic* VI, 511B (Grube/Reeve translation).

<sup>&</sup>lt;sup>22</sup> *Physics* A, 184a10-23.

nature, to the realm of matter-form composites, of which we now have systematic knowledge. As we will see, this ancient account of scientific method informs Kant's *Critique*.<sup>23</sup>

The general idea that scientific investigation involves complementary analytic and synthetic arguments is widespread: it accounts for certain kinds of demonstrations of the ancient Greek geometers,<sup>24</sup> and survives through the scholastic era finding its way into the logic textbooks of the modern period.<sup>25</sup> It even appears to account for some of Descartes' solutions to particular problems in natural philosophy.<sup>26</sup> Analysis and synthesis figure in Kant's lectures on logic largely in terms of the ancient metaphor of ascent and descent: *Logik Blomberg* claims "*Analysis* proceeds *ascendendo*, but *synthesis* proceeds *descendendo*" (§116, 24:110). The idea of analysis in Kant's logic texts is in keeping with the original idea of analysis that Aristotle attributed to Plato: analysis takes us to principles. The *Jäsche Logic* tells us that the analytic method "begins with the conditioned and grounded and proceeds to principles" (§117).<sup>27</sup>

<sup>&</sup>lt;sup>23</sup> This will be at issue throughout the dissertation. In particular, the idea that a principle is a starting-point or a "beginning" plays a crucial role in my reading of the Transcendental Deduction in Chapter 4.

<sup>&</sup>lt;sup>24</sup> There are several accounts of the role of analysis and synthesis in the method of the ancient Greek geometers, most focusing on Pappus's text, "The Treasury of Analysis": Jaakko Hintikka and Unto Remes, *The Method of Analysis: Its Geometrical Origin and Its General Significance*; Norman Gulley, "Greek Geometrical Analysis"; Michael S. Mahoney, "Another Look at Greek Geometrical Analysis"; Ali Behboud, "Greek Geometrical Analysis". (Although all the works listed refer to a "method of analysis", the method in question involves analysis and synthesis as complements.) Unfortunately, most of these authors fail to account adequately for how the notions of analysis and synthesis play out in the proofs themselves. I thank Ken Manders for sharing his deep understanding of the practice of ancient geometry with me, allowing me some rudimentary grasp of the distinction between analysis and synthesis in ancient geometry.

<sup>&</sup>lt;sup>25</sup> For an account of the history of this conception of scientific method into the modern period, see Peter Dear, "Method and the Study of Nature".

<sup>&</sup>lt;sup>26</sup> See Daniel Garber, "Descartes' Method and the Role of Experiment".

<sup>&</sup>lt;sup>27</sup> See also *Logik Hechsel*: "[I]n the analytic method one proceeds from things which rest on principles and proceeds towards principles" (115).

What are principles? Kant claims that the term is "ambiguous", distinguishing cognitions that can be used as principles even if they do not have the "origin" that is proper to principles. Since he seems satisfied enough to call reason the "faculty of principles" (A299/B356), then this "origin", apparently, is reason. But what does this really mean? Kant also associates reason with the drive for complete determination, for complete knowledge of the conditions of things (see, e.g., the Antinomies). The highest principle of morality is the categorical imperative; it is supposed to be that without which moral life would not be intelligible at all. In this way, a principle in this robust sense expresses the totality of a certain domain; it tells us what it is to be a person, or a moral agent. The highest principle of theoretical cognition is what Kant calls the "principle of the synthetic unity of apperception"; as I will argue in Chapter 4, its role is to unify independent accounts that Kant gives of sensibility and understanding. In doing so, it yields the principles of the pure understanding, which are principles determining the complete domain of material nature, with respect to its "possibility" or "form". They tell us what it is to be an object (which is understood broadly) in this domain. The point of critical philosophy — or at any rate the *Critique of Pure Reason* and the *Critique of Practical Reason* — is to show that these principles come from reason. And this means that certain conditions of the possibility of moral life, and likewise of theoretical cognition, are determined independently of experience.

But I mean to keep our attention fixed on a very general conception of "science", and not yet critical philosophy in particular. According to this general conception of science, we arrive at our principles analytically. What follows is the synthesis. Synthesis, on a rather ordinary conception of the term, involves putting elements together. This would suggest that the elements of a science are discovered in the analysis somehow, and combined in the synthesis. This combination, presumably, establishes the systematic unity of the elements; at least this is what Kant's logic lectures suggest throughout. "The true method of exposition is synthetic [...] for even if I have thought the thing analytically, the synthetic method first makes it a system" (*Hechsel*, 116).<sup>28</sup> *Dohna-Wundlacken* suggests, along lines similar to the *Logik Hechsel*, that scientific (*wissenschaftliche*) method just is, in the primary instance, synthetic because it involves combination, not separation: "{*Methodus* — the way a cognition can attain scientific form.} Method is combination of thoughts" (24:779). A science, according to Kant, is systematic knowledge from principles. Therefore if it is the synthesis that demonstrates the *systematic* character of the knowledge at issue, then proper scientific method would have to be "synthetic" at the end of the day.

These are highly general remarks on the respective roles of analysis and synthesis in scientific method. How does the distinction figure in Kant's explicit remarks about his own work? Kant says that the *Prolegomena* is carried out according to an analytic method. Specifically, this means that in addressing the question of the possibility of synthetic a priori cognition, the *Prolegomena* admits as given certain actual a priori sciences (pure mathematics and pure natural science), and exposes, through an analysis, the representations in virtue of which these sciences are capable of a priori claims. It has

<sup>&</sup>lt;sup>28</sup> A method of exposition, I take it, is concerned to determine clearly the particular domain of a science. When we present some domain of knowledge as a science, we demarcate in terms of the "distinguishing feature that it has in common with no other" (*Prolegomena* §1, 265): and *this*, I take it, would involve clearly demarcating its domain.

The contrast here is with a "method of discovery". The typical method of the empirical natural scientist, I take it, would count as a method of discovery. Only with a certain amount of progress, and reflective distance, would it be possible to lay out exhaustively the architectonic of the principles that are involved; this would require a method of exposition. Given that the first *Critique* is concerned to determine what it is

the structure of an iterated search for the conditions of the possibility of some given pure mathematics and pure natural science, as actual bodies of synthetic a priori cognition — which continues until we arrive at the principles on which this given rests. The analytic procedure of the *Prolegomena*, Kant says, is an ascent to the sources of knowledge (§4, 275).<sup>29</sup> As we would expect on the basis of Kant's logic lectures, the two main analyses of the *Prolegomena* lead to the "principle [Prinzip] of the possibility of what is given" in each case (275).

[Prolegomena] must support themselves on something which one already knows to be reliable, from which one can start off with confidence and ascend to the sources [Quellen] which are as yet unknown, the discovery of which not only explains to us what we already knew, but also at the same time presents to us a range of many cognitions which all spring from the named source. The methodological procedure of prolegomena, above all those which should prepare for a future metaphysics, will therefore be *analytic*. (§4, 274-5)

The argument of the Prolegomena draws to some extent on facts about the actual

practices of mathematics and physics. In the first analysis (of pure mathematics), it is

to figure in the domain of material nature, its synthetic method would count as a method of exposition. For more on this distinction, see Chapter 4, §2.

<sup>&</sup>lt;sup>29</sup> I must note that Kant mixes his metaphors here. Principles are conceived as "sources" of knowledge, and Kant sometimes employs the image of a "source" as something lying below the surface. For an example of this, see the metaphor in the *Prolegomena*'s Preface and Introduction, where reason is a "source" gushing the foam of metaphysical fancy that we, with our "natural" disposition for metaphysics, greedily scoop off the surface. There is always more on the way; and through critical philosophy, we are supposed to plumb the depth of this inexhaustible font. Critical philosophy is offered as a corrective or normative account of the capacity of reason; and the "principles" that it determines will be principles that express the nature of reason as a theoretical cognitive capacity. This metaphor in play of reason as a "source" of theoretical cognition is certainly at odds with the idea that critical philosophy should involve some kind of analysis that is conceived as an *ascent* to the principles of knowledge. Perhaps this conception of analysis as an ascent stems from the Aristotelian conception of scientific method, as aired at the outset of *Physics* A. The ground-level for this "ascent" is presumably experience; in other words, we ascend to the principles that make empirical cognition possible. But again, these principles are attributed to our theoretical cognitive capacity (a *source*) in Kant's account. The idea of analysis as some kind of ascent is simply another metaphor, and it does not seem that any real confusion should result from the conflicting imagery.

through the recognition that we must *construct* concepts in intuition a priori that we are led, eventually, to certain pure forms of intuition as the "sources" of mathematical knowledge. In the second analysis (of pure natural science), it is by close consideration of the concept of nature and the force of "judgments of experience" that we are led to the principles of the pure understanding as the "sources" of our knowledge of physics.<sup>30</sup> The method of the *Critique*, on the other hand, is "synthetic". The *Critique* does not admit any "facts" as its starting point; it is "a whole science, robbed of all help from other sciences and thus is necessarily in itself entirely new" (§5, 279).<sup>31</sup> It "lays down nothing as given [...] except reason itself" (§4, 274).

The most general characterization we can give of the difference between analysis and synthesis in scientific method — one that, I think, captures the essential elements of the distinction from the ancient Greeks to Kant's own *Critique* — is to say that analysis is an ascent to principles, and synthesis is some sort of descent from these principles, generally described as a return to the starting point. It is in just these terms that Kant invokes the distinction at the end of the Preface to the *Groundwork for the Metaphysics of Morals*:

I have adopted my method in this work as I believe it to be most suitable: if one proceeds analytically, from ordinary cognition to the determination of its highest principle, and then back again synthetically, from the

<sup>&</sup>lt;sup>30</sup> Remarking on the analyses of the *Prolegomena*, Kant remarks: "One sees that [...] the solution to these problems [...] has something peculiar, which is worthy of attention on its own account: namely to search for the sources of the given sciences in reason itself, in order that its faculty of knowing something a priori be explored and measured by means of the very act [...]" (P §5, 280).

<sup>&</sup>lt;sup>31</sup> And hence its great difficulty; for without admitting "help" from any other science (one has to wonder about the role of pure general logic, however), the *Critique* will only arrive at a solution to the problem about synthetic a priori judgment "with trouble and difficulty, and indeed bound up with a certain obscurity" (279).

examination of this principle and its sources to ordinary knowledge in which its employment is found. (G, 392)

The first stretch of the *Groundwork* is an analysis that takes us from "ordinary rational moral cognition" — i.e., folk wisdom about the "good will" — to its principle or "source". This principle is known as the categorical imperative. It counts as a first principle, because without it the ordinary moral life that was invoked at the outset with the idea of a "good will" would not be intelligible at all. The synthesis is supposed to take us from this principle back to our starting point, giving us an appreciation of the employment of the principle.

But the idea that we return to our starting point need not imply that we are supposed to return to folk wisdom about the good will at the close of the *Groundwork*, but rather that we return from a purely formal expression of the moral law to some appreciation of that law as the very foundation, or condition for the possibility, of moral life at all. (We need not suppose that the "descent" is a project of applied ethics.<sup>32</sup>) Once again, in this passage from the *Groundwork* we find the suggestion that analysis and synthesis are complements to one another; the "method" Kant describes here involves

<sup>&</sup>lt;sup>32</sup> At the end of *Groundwork* §II, Kant tells his reader that the first two sections of the work were "merely analytic [bloß analytisch]", and implies that the third proceeds synthetically. Still, this is a matter of some controversy. Dieter Schönecker ("Zur Analytizität der *Grundlegung*") argues that *all three* sections of the *Groundwork* comprise an analytic argument that is a preparation for the allegedly synthetic argument of the *Metaphysics of Morals*. H.J. Paton argues that the third section of the *Groundwork* is a synthetic argument which begins with an examination of the principle arrived at through the analysis of *G* §§I-II "*and its sources* […] in practical reason itself" and winds up indeed back with "the common knowledge in which it is employed". I find it hard to understand the third section of the *Groundwork* as getting back to "common knowledge" to quite this degree — but then again, Paton seems not to want to push the matter, either: "Such at least is his [sc. Kant's] own account of the matter, and it is substantially correct, although he does not in fact pay any attention to the lower stages of the descent" (*Categorical Imperative*, 29).

*both* analysis and synthesis.<sup>33</sup> This conception of a method involving complementary movements of analysis and synthesis is quite in keeping with the conception of method that Aristotle attributes to Plato. Presumably, then, the method of the *Critique* might involve complementary strands of analysis and synthesis. The analysis would lead to some first principle of the critical science, and the synthesis would establish the system of the whole according to this principle. I have presented here only the bare preliminaries of Kant's distinction between analysis and synthesis in scientific method; I offer an account of how it illuminates the text of the *Critique* itself in Chapter 4.

4.

We now have a preliminary grasp of Kant's conception of scientific method. It may help to recapitulate, briefly, some of the basic points that we have covered. For Kant, a proper science is nothing if not *systematic*, in precisely the way that Strawson ignored in his presentation of the *Critique*. Whatever is to count as a proper science proceeds according to a proper *method* (not a "manner", or a "procedure", which cannot

<sup>&</sup>lt;sup>33</sup> No doubt it is difficult to account for the "synthesis" in the *Groundwork*; by most accounts, it would be found in the third part, which is viewed by many commentators as a kind of Critique of Practical Reason in germ form. Perhaps, then, the best strategy is to consider the role of analysis and synthesis in the Critique of Practical Reason itself. According to Stephen Engstrom, in his Introduction to Werner Pluhar's recent translation of the second *Critique*, its analytic and synthetic movements map (respectively) onto its two main "books", the Analytic of Pure Practical Reason, and the Dialectic of Pure Practical Reason. The Analytic is an "analysis of the faculty of reason in its practical employment, in which the a priori and empirical sources of its principles and of the motives for acting on them are distinguished" (liii). This allows Kant to identify the source of morality in the autonomy of the will. Nevertheless, the analysis alone could "give us the impression that the two elements it separates — morality and happiness — have no relation to one another", saddling us with "the image of a fragmented practical life" (liv). According to Engstrom, this problem is not entirely academic: "such an image can easily become the source of doubts about morality that can weaken the moral motive, or [...] it can become the source of doubts about Kant's own analysis". So the synthetic argument of the Dialectic demonstrates how the a priori and empirical elements that were clearly separated in the analysis are "necessarily combined in the highest good". The intended result is a "systematically unified conception of practical life, to which both virtue and happiness are integral", which is itself supposed to help "secure 'acceptance and durability' for the moral law".

yield knowledge as a *system*).<sup>34</sup> Scientific method involves complementary analytic and synthetic arguments. The analysis is an ascent to the principles of the science, while the synthesis establishes a systematic body of knowledge on the basis of the principles that were uncovered in the analysis. Any analysis that may be found in the *Critique*, Kant tells us, takes place for the sake of the synthesis that follows; the analysis is "purposeful" (*zweckmäßig*). Some of Kant's remarks suggest that the synthesis establishes the relation among the parts or elements of the science on the basis of the principle or set of principles that is uncovered in the analysis. At the same time, the passage from the *Groundwork* suggests that the synthetic argument may be characteristically concerned to demonstrate the *employment* of the principles uncovered in the analysis.<sup>35</sup>

Obviously we have a long way to go before this general account of scientific method might helpfully guide us through the pages of the *Critique of Pure Reason*. We must first arrive at a clearer understanding of the particular concern of the work, which is apparently reason itself. A quick glance at the *Critique*'s table of contents might give us the impression that the we learn nothing about reason until the second half of the book, the Transcendental Dialectic. But in fact what we learn about reason in the Dialectic is mostly negative: it is meant to be a systematic account of reason's "incapacity", or

<sup>&</sup>lt;sup>34</sup> Kant distinguishes "method" proper (*Methode*, *Lehrart*) from mere "procedure" (*Verfahren*). For example, Hume's quasi-critical "censorship" of reason is not a science; his haphazard inquiry is deemed a mere "procedure" (*Verfahren*, A760/B788), which Kant consistently distinguishes from method proper (*Methode*). Kant also distinguishes between "method" and "manner": a manner, like a procedure, yields an aggregate of knowledge and not a system. This comes out particularly clearly in the critical-period *Logik Hechsel.* "*Method* is the unity of a whole of cognition according to principles. A unity of cognition can be empirical, [...] in so far as it is in accordance with rules that can be drawn from experience. But there is also unity in accordance with universal principles of reason, where we can produce a thoroughgoing coherence [Zusammenhang], and can produce a system, in which we discover the nature [Art] of the whole through the connection [Verknüpfung] of the manifold. In so far as the unity of cognition rests on empirical rules, it is called manner, in Latin *modus*. But in so far as the unity of the manifold rests on principles of reason, it is called *methodus*, method [Lehrart]" (*Hechsel*, 114).

*Unvermögen*, and not its "capacity", or *Vermögen*. As I will demonstrate in the course of this dissertation, the positive account of reason is found in the first part of the *Critique*, and the Transcendental Deduction should be recognized as the central text of reason's scientific self-knowledge.

One thing that Kant's general conception of scientific investigation tells us is that the *Critique* should begin with some "idea of the whole". This "whole", it seems, must be reason. For what else could Kant have in mind when he claims that the *Critique* takes "nothing as given except reason itself"? We are presumably beginning with some very general, preliminary conception of reason, and working towards a "scientific" account of it. This preliminary account of reason can be found in the *Critique*'s Preface. The official topic of the Preface is the nature of rational cognition, specifically that which has found the "sure path" of a science.<sup>36</sup> This is where Kant considers paradigm cases of scientific rational cognition — Euclidean geometry and Newtonian physics — and asks what it would take for metaphysics to join these as their rightful peer. This story about scientific rational cognition has much to tell us about how Kant thinks about the nature of reason. My aim in this section (§4) is to spell out what this preliminary conception of reason is supposed to be. In the following section (§5), I will unite this with what we have just covered about Kant's conception of scientific method, in order to draw some preliminary conclusions about how to come to terms with Kant's idea of the *Critique* as reason's "scientific self-knowledge".

<sup>&</sup>lt;sup>35</sup> Although there seem to be two distinct roles for the synthetic argument, they need not be seen as incompatible. More discussion of the synthetic argument of the *Critique* follows in Chapter 4.

<sup>&</sup>lt;sup>36</sup> This remark pertains primarily to the second edition Preface. For an account of my reasons for focusing almost entirely on the second edition version of the *Critique*, see the Introduction to this dissertation.

In the Preface to the second edition of the *Critique*, Kant considers the nature of theoretical rational science in general, as a platform on which to diagnose the failure of metaphysics in the face of the great success of logic, mathematics, and physics.<sup>37</sup> It is a historical presentation of the problem: metaphysics has failed while the other rational sciences have flourished. Kant's idea is that metaphysics could be a genuine rational science if only it would find its foundation and have its revolution. The implication, of course, is that the *Critique* itself will bring about that internal revolution and establish metaphysics among its rightful peers.

When Kant turns to physics, he is talking about an empirically driven project. Yet it is one that owes its success as a science to its recognition that reason is "taught by nature not as a pupil, who recites everything the teacher wants him to say, but like an appointed judge who requires witnesses to answer the questions he puts to them" (Bxiii).<sup>38</sup> In this paradigmatic expression of its capacity, reason is like a judge approaching nature with its questions. The capacity of reason, presumably, is expressed in the very formulation of these questions; and it awaits answers that it will recognize as answers to *its questions*. Kant then points to certain episodes in the history of scientific experimentation that he takes to illustrate this idea: "When Galileo rolled balls of a weight chosen by himself down an inclined plane, or when Torricelli made the air bear a weight that he had previously thought to be equal to that of a known column of water [...] a light dawned on all those who study nature. They comprehended that reason has insight only into what it brings forth according to a plan of its own" (Bxiii). Kant's

<sup>&</sup>lt;sup>37</sup> The entire discussion is inaugurated by invoking "the treatment of the modes of cognition [Erkenntnisse] which belong to the concern of reason", and which "take the sure path of a science" (Bvii).

<sup>&</sup>lt;sup>38</sup> Kant often remarks derisively about mere "learning" — the passive, and uncritical, acceptance of information. See, e.g., A835-7/B863-5, as well as Kant's critical-period essay "What is Enlightenment?".

remarks here fall well short of any real account of experimentation in empirical natural science; at best, they merely gesture towards a general conception of it. Nevertheless, Kant seems to suppose that there are *some* general characteristics of the investigative practices of individual physicists (Galileo, Torricelli, and the rest) that make them the investigative practices of a *science*. Regardless of what this should turn out to involve, the overarching idea of these practices is that they somehow allow — and require — the experimenter to track clearly what belongs to the framework of the experiment and distinguish it from whatever does not. This means that nature's "response" would be recognized loud and clear —provided, at least, that the experimental practices in question are properly executed. And this would mean that no mere artefact would masquerade as a genuine result.

What does this tell us about the preliminary conception of reason that Kant means to have in place at the outset of the *Critique*? It suggests that even in empirical modes of inquiry, when reason is occupied with particular objects in the domain of material nature, it must at the same time recognize how its own principles — guidelines for what it can admit as reasons and explanations — are expressed in setting up particular experiments. In other words, reason must be tacitly concerned with itself even while it is expressly concerned with the particular phenomena in question. Perhaps Kant means to suggest that reason is fundamentally self-concerned. It is not clear what it would mean to say this; but let us admit it as an exegetical hypothesis, and test it against Kant's treatment of the other forms of rational cognition mentioned in the *Critique*'s Preface.

The development of practices of reasoning with diagrams in Euclidean geometry was another great "light" to dawn in the history of human reason. This "revolution in the way of thinking", Kant proclaims, "was far more important than the discovery of the way around the famous Cape" (Bxi). Let us then look at a simple example of such reasoning: the first proposition of Euclid's *Elements*, demonstrating the construction of an equilateral triangle on a given finite straight line. Each end-point of the given line serves as a point from which to construct a circle, with the given line as the radius. The two circles intersect in two places. We pick one of the two points of intersection, and draw a straight line to each end-point of the given line. We have constructed an equilateral triangle with the given line segment as its base.

Reasoning with diagrams in Euclidean geometry turns on our capacity to distinguish between contingent and non-contingent features of the diagram. The two points of intersection are non-contingent features: in other words, they are features of the diagram that are salient for us as geometers carrying out this particular demonstration. Notice, however, that the two points of intersection are not explicitly or directly referred to in the prior discursive part of the proof; they only "pop up" in the course of carrying out the construction itself. Nevertheless, they are *not* contingent features of the diagram: they are *not* features of the diagram that we are obliged to ignore. Our ability to know the difference rests on our tacit appreciation of the *connection* between the discursive part of the demonstration and what becomes available to us through the construction. Our capacity to distinguish the non-contingent features of the diagram from the contingent features is, at bottom, an awareness of ourselves; it is an awareness of what we put into the figure, versus what we did not. In dealing with a particular drawn figure, Kant remarks that we attend "always only to the action of the construction of the concept" (A714/B742). There are "many determinations" of the particular drawn figure to which

we must be indifferent, in so far as we attend only to the action of the construction. These are the contingent features. In one demonstration, we must be indifferent to the magnitude of the angles of a triangle, and yet in another we may not be (perhaps in certain demonstrations involving equilateral triangles).

Now, reason itself is not the topic or express concern of either physics or mathematics. Yet these are the paradigm cases of rational cognition that Kant considers in the *Critique*'s Preface, where they make a proud appearance for having found the "sure path" of a science. Both of these examples highlight a certain kind of self-concern, or self-consciousness, that gives the practice of investigation its distinctive character.

The other modes of rational cognition that Kant considers in the Preface are metaphysics and logic. With these examples, the theme of self-consciousness is rather more explicit. In the *Critique*'s Preface, metaphysics is not so much examined, but bemoaned; but in the *Prolegomena*, Kant says that metaphysics is "the occupation [Beschäftigung] of reason merely [bloß] with itself" (*P* §40, 327). And Kant says nearly the same thing about *pure general logic* in the Critique's Preface: in logic, "reason has to do only [nur] with itself" (Bx).

Let us review what we have on the table, and see if we can refine our working hypothesis about the nature of reason. We can take it that "rational cognition" is simply a generic way to refer to the activity of reason; thus, if we can discern some common feature of all modes of "rational cognition" (at least among those that Kant considers in the Preface), then this common feature might belong to the nature of reason. Indeed, we have discovered a common feature: for it seems that in all modes of rational cognition reason is — in some sense that we have yet to understand properly — concerned with itself.

In order to understand this better, we might invoke Kant's distinction between "material" and "formal" rational cognition. In the *Groundwork*, Kant divides all rational knowledge into formal and material sciences (see Figure 1). Material rational knowledge is "concerned with some object". Formal rational knowledge is "concerned only with the form of understanding and of reason themselves and with the universal rules of thought in general without distinguishing its objects" (387). On the basis of this distinction, we can refine our working hypothesis. In the material sciences, reason is *expressly* "concerned with some object", but *tacitly* concerned with itself. In the formal sciences, reason is expressly concerned with itself, or with the intellectual cognitive capacity in general. Is it concerned with anything other than itself, even tacitly? This turns out to be a fruitful question, the answer to which should tell us something crucial about critical philosophy.

According to this distinction between material and formal rational cognition, critical philosophy would clearly have to count as a formal science. As a project of rational self-knowledge, the *Critique* is obviously concerned with "the form of understanding and of reason". But should we stop in our tracks when Kant remarks that "[f]ormal philosophy is called logic" (*G*, 387)? Does this mean that critical philosophy is some kind of *logic*? Yes: if logic is construed broadly, as when Kant says that logic is "a consciousness of the understanding" (*Anthr.*, 134n.). In this broad sense, which obviously would include the *Critique*'s own Transcendental Logic, logic is concerned with "the I as *subject* of thinking [...] (the merely reflecting I)". This conception of logic is distinguished from *psychology*, which is concerned with "the I as the *object* of

perception, thus of inner sense, which contains a manifold of determinations which make possible an inner *experience*" (cf. *Fortschritte*, 270-1). In the broad sense, logic is an inquiry into the intellect itself. It is not a mode of material cognition because it does not require reason to direct its express concern to phenomenal objects (of any kind, whether of inner or outer sense). Critical philosophy is a form of logic in the broad sense; this means that critical philosophy is *formal philosophy*. But so, apparently, is pure general logic, a science of the intellect that "abstracts from all content of cognition" (A55/B79). So how do we distinguish between pure general logic and critical philosophy? Presumably, the *Critique* — or, at any rate, the great bulk of it known as "transcendental logic" — is a formal science that does *not* "abstract from all content of cognition". Although this allows us to distinguish between pure general logic and critical philosophy, it presents us with a new problem. For one way to understand *this* distinction between these two modes of "formal philosophy" is to recognize that critical philosophy is a kind of *metaphysics*, and pure general logic is not.

In the Architectonic of Pure Reason, Kant clearly indicates that critical philosophy is part of the "philosophy of pure reason" (A841/B869). Here Kant distinguishes between broad and narrow senses of "metaphysics". Critical philosophy is a "propaedeutic [...] which investigates the faculty of reason with respect to all pure a priori cognition", and is distinct from the "system of pure reason (science [Wissenschaft])".<sup>39</sup> The latter is metaphysics in the narrow sense, which breaks down into the metaphysics of nature and the metaphysics of morals (see also *G* 387-8). Critical philosophy is part of metaphysics in the broad sense, which includes "everything which

<sup>&</sup>lt;sup>39</sup> This suggests that critical philosophy is, in effect, the 'science of science'.

can ever be known a priori, as well as the exhibition [Darstellung] of that which constitutes a system of pure philosophical cognitions, but is distinguished nevertheless from all empirical as well as mathematical employments of reason".<sup>40</sup> So critical philosophy is a kind of metaphysics. (See Figure 2.)

The *Critique of Pure Reason* would have to be conceived as *critical* theoretical metaphysics; it is a formal science. Assuming that the sort of project that Kant pursues in the *Metaphysical Foundations of Natural Science* is the best candidate for the anticipated "future metaphysics" of material nature, then this post-critical theoretical metaphysics would count as a material science, since it is concerned with the fundamental laws determining objects in the domain of nature. Critical philosophy, as a formal science, is not directly concerned with objects in the domain of nature, but rather with the "possibility" of such objects in so far as this has its source in reason. (I will explain this idea further in Chapter 2.) In critical philosophy, reason is *indirectly* concerned with objects. This is what distinguishes it as a metaphysical project. By contrast, in pure general logic, reason is not concerned with objects in the domain of nature *at all.* 

Critical philosophy is a formal science, whereas metaphysics in the narrow sense is a *material* science. This material science divides into the metaphysics of nature and the metaphysics of morals; in either case, it is a "doctrine" (*Naturlehre* or *Sittenlehre*; *G*, 387).<sup>41</sup> Critical philosophy is not: "We are not to call this investigation a doctrine

<sup>&</sup>lt;sup>40</sup> What are we to make of Kant's remark, in his August 1783 letter to Christian Garve, that "it is not at all metaphysics that the *Critique* is doing"? We can interpret this in light of Kant's distinction between the "broad" and "narrow" senses of metaphysics: the *Critique* is not engaged in metaphysics in the narrow sense. At bottom, Kant is trying to impress upon Garve that the *Critique* does not draw upon any prior metaphysics; on that account, it is not a work of dogmatic metaphysics.

[Doktrin], but only a transcendental critique, since it has as its aim not the furthering [Erweiterung] of cognitions themselves, but only their correction [Berichtigung], and is to supply the touchstone of the worth or lack of worth [Unwert] of all a priori cognitions" (A12/B26).<sup>42</sup> The difference between a "doctrine" and a "critique" turns on the fact that the former makes claims about objects, and the latter does not. This point is made again in the *Critique of Judgment*: "The critique of the faculties of cognition with regard to what they can accomplish a priori has, strictly speaking, no domain [Gebiet] with regard to objects, because it is not a doctrine [Doktrin], but has only to investigate whether and how, in accordance with the relation which it [sc. the domain<sup>43</sup>] has to our faculties, a doctrine is possible by means of it [sc. a critique]" (*KU* §III, 176).

This distinction between formal and material sciences allows us to track Kant's distinction between critical philosophy and the rest of metaphysics. In the process of following Kant's distinction between critical philosophy and the rest of metaphysics, we should not lose sight of the idea that critical philosophy is itself supposed to be a *kind of metaphysics*. The importance of this point goes beyond merely distinguishing critical philosophy from pure general logic: for it actually tells us something about how to conceive of the *Critique* as reason's self-knowledge. But in order to have this insight, it will help to become more puzzled about Kant's placement of critical philosophy under the broad heading of "metaphysics". For if critical philosophy is indeed a kind of

<sup>&</sup>lt;sup>41</sup> Empirical psychology is also a doctrine, a doctrine of soul (*Seelenlehre*), considering the object of inner sense (*MAN*, 467).

<sup>&</sup>lt;sup>42</sup> A material "philosophy of nature" makes claims about objects in the domain of the laws of nature, and a material moral philosophy makes claims about persons, i.e., objects in the domain of the law of freedom; both a philosophy of nature and a moral philosophy may (indeed, must) be subject to a corresponding critical philosophy which addresses the "worth" or validity of the philosophical claims contained in the "material" philosophy.

<sup>&</sup>lt;sup>43</sup> es: the neuter pronoun can only refer back to "domain" (*das Gebiet*).

metaphysics, and if the *Critique* is supposed to be "grounding" or "legitimating" metaphysics, then a new problem stares us in the face. The *Critique* is evidently engaged in the very sort of project (metaphysics) that it is also, at the same time, supposed to be "grounding" or "legitimating". Apparently critical philosophy — if it is viably the project that it aspires to be — would have to be self-grounding.<sup>44</sup> Is this another foundering point for critical philosophy?

Heidegger conceives of critical philosophy as a "philosophical grounding of philosophy".<sup>45</sup> He says that Kant's critical grounding of metaphysics must both arise out of the tradition of metaphysical inquiry *and* at the same time transform that tradition.<sup>46</sup> Unfortunately, this otherwise attractive interpretation stands at odds with Kant's own conception of his work as "drawing from the sources of reason" rather than the history of dogmatic metaphysics. Admittedly, Kant's dismissal of the tradition is hard to take

<sup>&</sup>lt;sup>44</sup> One way to understand the idea that critical philosophy is supposed to be self-grounding is to compare it with other works of the critical period that are not works of critical philosophy in sensu strictu. The substantive insights of critical philosophy pertain to the identification of principles that determine 'what it is' to figure in either the domain of material nature (theoretical philosophy), or the "kingdom of ends" (practical philosophy). In the truly "critical" works, Kant is concerned to demonstrate the origin of these principles in reason itself — and to do so without relying on any given body of *actual* rational cognition. Other works of the critical period (that are not works of critical philosophy in the strict sense) take as given that certain modes of rational cognition are "actual": the Prolegomena sets out with the presupposition that pure mathematics and pure natural science are no "phantoms of the brain", and the Groundwork treats "ordinary moral rational cognition" in the same way. Critical philosophy in the strict sense is supposed to identify and establish these principles without relying on anything other than "reason itself" — that is, it relies only on some general idea of the capacity of reason, and not any particular expression of its capacity. Kant holds that only reason can adjudicate the claims of reason (A752/B780; P, 263); critical philosophy is supposed to give an exhaustive account of the bases for the adjudication of these claims. Hence, it aspires to give a complete or exhaustive account of the capacity of reason — it goes to the source, not merely the expression, of these claims. This means that the work cannot draw on any particular science, but apparently just "reason itself".

<sup>&</sup>lt;sup>45</sup> Kant und das Problem der Metaphysik, 37. This self-grounding of philosophy, Heidegger claims, leads Kant to point "consciously to the unknown", which is the "possibility of a priori synthesis". The possibility of such synthesis involves the imagination as the "unknown root" of sensibility and understanding. Dieter Henrich's review of Heidegger's 'Kantbuch' ("Über die Einheit der Subjektivität") convincingly argues against Heidegger's reading of the imagination as such a fundamental power (*Grundkraft*), by showing that if Kant were consciously pointing to the unknown, he would be pointing to something intrinsically unknown, and intrinsically unknowable.

<sup>&</sup>lt;sup>46</sup> Kant und das Problem der Metaphysik, 2-3.

seriously: it smacks of the fantasy of cutting all ties with the past and making a fresh start — a well-known attribute of the self-conception of modern philosophy, and perhaps modernist thought in general. It is rhetoric, at best.

Yet Heidegger's gloss on critical philosophy might actually lead us to the correct interpretation of Kant's dismissive remarks about the history of philosophy. Let us begin with a minor reformulation of Heidegger's gloss: the *Critique* is a metaphysical grounding of metaphysics. The point of Kant's dismissive remarks about the tradition might tell us the following about critical philosophy: namely, that the "grounding" or "legitimation" of metaphysics cannot be carried out from an external perspective. The work that legitimates metaphysics must itself be metaphysics.

Critical philosophy is supposed to give us the *form* of metaphysics; this is what Kant has in mind when he claims that it is a "treatise on the method" of metaphysics, but not the doctrine itself (Bxxii). This point is made again in the First Introduction to the *Critique of Judgment*, where Kant claims that critical philosophy does not belong to philosophy "as a part". At first blush, this may seem to tell against the idea that the *Critique of Pure Reason* is itself a metaphysical project, legitimating metaphysics from within metaphysics.

If philosophy is the system of rational cognition through concepts, it is thereby already sufficiently distinguished from a critique of pure reason, which, though it contains a philosophical investigation of the possibility of such cognition, *does not belong to such a system as part* — but rather sketches [entwirft] and examines [prüfet] the very idea of it in the first place. (*EE* 20:195, my emphasis; see also A13/B27)

As the *form* of metaphysics, critical philosophy is not a "part" but rather the determinate idea of the *whole* of metaphysics.<sup>47</sup> "It [...] specifies the entire outline of metaphysics, both in respect of its boundaries and also its entire internal structure" (Bxxii-xxiii). Nothing loose or imprecise is suggested by the idea that the *Critique* "sketches" (*entwirft*) or draws up a plan for a future metaphysics of nature: the *Critique* is "formal" in something like the sense in which architectural drawings are "formal", while the building itself is "material".

Perhaps this metaphor is misleading: for the value of a blueprint lies largely in the soundness of its prescription for the construction of a building. Certainly it is hard to determine, on the basis of critical philosophy alone, what exactly the "future metaphysics" is that we are meant to anticipate. Still, we can draw some tentative conclusions about the nature of critical philosophy from this picture of it as the "form" of theoretical metaphysics. It is a formal rational science, which means that it involves the explicit self-concern of reason. This suggests that critical philosophy should be conceived as a project of making something explicit that is already implicit in scientific theoretical cognition in general. So even though reason is directly concerned with itself in critical philosophy, it would be indirectly concerned with the domain of material nature. This is what makes it "metaphysics" — we might even call it "reflective metaphysics", since reason's examination of its *theoretical* capacity is supposed to lead to an account of what it is to figure in the domain of material nature.

The *Critique*'s Preface draws our attention to some kind of self-concern or selfconsciousness that is apparently the mark of rational cognition. This is where the work

<sup>&</sup>lt;sup>47</sup> Critical philosophy, Kant says, purifies metaphysics, brings it into a "permanent state [beharrlichen Zustand]" (Bxxiv; see also *P* 366 for a very similar passage). This also suggests a connection with the idea

begins: as we continue on into the work proper, this conception of reason should become clearer and more determinate. The apex of the ensuing account is the determination of a principle that Kant calls the "principle of the synthetic unity of apperception". This principle, I shall argue in later chapters, lies at the heart of the positive account of reason's theoretical capacity. We are meant to recognize it as the principle that tells us *what reason is.* But we are barred from doing so if we fail to appreciate that the *Critique* presupposes a conception of reason that is merely illustrated, and not yet articulated, in its Preface.<sup>48</sup>

5.

The story of this chapter so far has had two basic themes: first, Kant's conception of proper scientific method, and second, the *Critique*'s preliminary conception of reason. What we are still wanting, however, is some rough-and-ready grasp of the idea that the *Critique* is the *scientific* expression of reason's self-knowledge. So I will now try to bring the two themes together.

The idea that the *Critique* is reason's scientific self-knowledge leaves us with an exegetical puzzle. Although the *entire* work is supposed to be reason's self-knowledge, cursory acquaintance with the text might suggest that this is true only of the part of the

of "form", though in a somewhat Platonic vein.

book in which reason becomes thematic — the Transcendental Dialectic. The official topic of the other half of the book — roughly speaking, the Transcendental Aesthetic and the Transcendental Analytic — is sensibility and understanding, not reason. We could be left thinking that the truly "critical" work of the *Critique* is relegated to the Transcendental Dialectic, the complete catalogue of the errors of reason in its speculative employment. The Aesthetic and the Analytic, in this story, would be exempted from the corrective or normative enterprise.<sup>49</sup>

If we accept this story, it becomes difficult to account for the title of the book: the entire thing is supposed to be a *critique of pure reason*. The title suggests that the work is critical or evaluative through and through, and that it is concerned with the capacity of reason from beginning to end. Reason is supposed to be investigating itself in the *Critique*. Reason is, in effect, the "agent" of the investigation. Its presence is not thematic in the first part of the book, but implicit in the work is carried out. In this arena, reason's demand for completeness is supposed to result in an *exhaustive* account of its theoretical cognitive capacity. Reason's investigation of its *theoretical* capacity — its capacity to judge about objects in the domain of nature — will be an account of its relation to sensibility and understanding.

At various points, Kant hints that the role of reason in his *Critique* is akin to the role of reason in Plato's *Republic*. For example, error is characterized as the "unnoticed influence of sensibility on the understanding"; sensibility "meddles" in the proper business of the understanding (A294-5/B351). We cannot isolate the capacity of reason;

47

<sup>&</sup>lt;sup>49</sup> See (e.g.) Brook, *Kant and the Mind*, who says that "when we turn to what the mind must be like to reason and know, we are turning to description, to how the mind does and must work (in some sense of 'must'). *We are not seeking a normative account of how it should work*" (5, my emphasis). The Aesthetic and the Analytic are merely descriptive, and not "critical", on this reading.

it is properly conceived as a "governing" or "legislative" capacity with respect to sensibility and understanding.<sup>50</sup> Thus, the positive story about the theoretical capacity of reason is found in the first half of the book, where the official or thematic concern is sensibility and understanding. This explains Kant's embarrassment about having to give a definition of reason in the Introduction to the Transcendental Dialectic (A299/B355): the reader may not realize it, but in fact the core positive account of reason has already been given with nary a definition in sight.

But the Dialectic is not entirely a negative project, concerned only to determine the "incapacity" or *Unvermögen* of reason in a systematic way. The positive account of reason continues into the Introduction to the Dialectic, where the reflective or governing role of reason — particularly with respect to the understanding — is underscored as the overarching lesson that a careful reader might have taken from the first half of the book. The main text of the Dialectic will drive home the point that reason cannot judge directly about objects. The proper concern of reason is with our cognitive capacity itself. This is simply a development of the same thought that the *Critique*'s Preface made vivid: that scientific cognition turns on some kind of awareness of what our own cognitive capacity contributes to cognition. Reason, Kant says,

never applies [geht ... auf] directly to experience, or to any object, but rather to the understanding, in order to give the manifold cognitions of it a

<sup>&</sup>lt;sup>50</sup> This is implied in Part C of the Introduction to the Transcendental Dialectic (A305/B362 ff.); the governing or legislative role of reason becomes most thematic in the third chapter of the Doctrine of Method (the Architectonic of Pure Reason). Kant seems to understand the "governing" role of reason in terms of "transcendental reflection" (see A295/B351), which apparently has to do with determining the proper bounds of sensibility and understanding, and their proper relation to one another. As I will argue in Chapter 4, the relation of sensibility and understanding — the "elements" of critical philosophy — is determined on the basis of the principle of the synthetic unity of apperception in the Transcendental Deduction. Thus, the positive account of the theoretical capacity of reason — which can be conceived as a "reflective" or "governing" capacity — would be found in the first half of the *Critique*.

priori unity through concepts, which may be called "unity of reason" [Vernunfteinheit] and is of an entirely different kind than that which the understanding achieves. (A302/B359)

Thus if pure reason also applies [geht ... auf] to objects, then it has no immediate relation [Beziehung] *to these* and their intuition, but rather only to the understanding and its judgments, which apply [wenden] directly to the senses and their intuition, in determining their object. The unity of reason is thus not a unity of possible experience, but is essentially distinguished from this, as the unity of understanding. (A306-7/B363; see also A643/B671)

Reason "applies" itself to the understanding, and the understanding judges about what the senses make available. The understanding is a capacity for empirical judgment. Since these passages say that reason is concerned with the understanding, they appear to stand at odds with the idea that some kind of self-concern is the fundamental characteristic of reason. However, the real message — clearest perhaps in the first passage — is that the understanding is not an *object* of reason: Kant says that reason does not apply directly to any object *but rather* to the understanding, suggesting that the understanding *is not an object.*<sup>51</sup> The self-conscious nature of reason expresses itself in its concern with the judging activity of the understanding.

I have been arguing that, for Kant, reason is a self-conscious or reflective capacity. But what about a more obvious claim that one might make about the fundamental activity of reason — namely that reason is a capacity for "reasoning"? In the first *Critique*, Kant is concerned with the higher cognitive faculty at large, which we can think of (broadly speaking) as the *understanding*. The end of the higher cognitive

<sup>&</sup>lt;sup>51</sup> There is to my knowledge just one passage in the *Critique* where Kant speaks of the understanding as reason's "object" (A664/B692), but I do not think it undermines my reading, because in the passage the understanding is only said to be reason's "object" in what is in effect a metaphorical extension of the term.

faculty is "to understand", or to have knowledge. Further faculties are distinguished simply as ways of identifying fundamental aspects of theoretical cognition. As I will argue in Chapter 3, one of those aspects is the self-conscious activity that Kant means to attribute to reason, and the other aspect is the judging activity that Kant quite explicitly attributes to the understanding (in the narrow sense). The *Critique* aims to show, in effect, that there could be no judging understanding were there no reflecting reason. This is the philosophical lesson that we are liable to miss if we fail to appreciate the overarching framework of the project as *reason*'s investigation of its theoretical capacity.

So how are we to understand the idea that the *Critique* is a project of rational selfknowledge? Perhaps we can capture the basic idea, at least in a preliminary way, with the following slogan. *In the* Critique, *reason makes its own self-conscious activity explicit as the source of theoretical cognition in general.* I will say more about this in the following two chapters. But if this is the right way to conceive of Kant's idea of a "critique of pure reason", then does it not imply that the work is descriptive? When we make something explicit, we draw attention to something that is "already there", as it were. A project of "making explicit" would be descriptive, it seems, as opposed to constructive.

It is useful, I think, to see the *Critique* as a project of making something explicit that is generally only implicit in our cognitive lives: it makes explicit the reflective character of reason, revealing this as the source of our capacity to have knowledge about objects in the domain of nature. As I hope to make clear in Chapter 3, this overarching interpretation of the work keeps us from misunderstanding Kant's conception of our cognitive "spontaneity". Yet at the same time, we should be careful that this conception of the *Critique*'s project does not obscure its normative aspirations.

To prevent this misapprehension, we might call to mind the conception of the work that figures throughout the *Critique*'s Doctrine of Method, and particularly in the History of Pure Reason. Kant suggests there that we think of the *Critique* as reason's upbringing, the perfection of rational nature.<sup>52</sup> According to the story, the "indifference" of reason's dogmatic childhood gives way to the "resting place" of Humean skepticism, the adolescence in which reason scornfully looks back on its dogmatic childhood (A761/B789). The *Critique* is supposed to complete the development: reason reaches "adulthood" under the guidance of critical philosophy, and comes to know itself for the first time. If it is apt to think of this as reason's self-description, then it is equally well reason's self-transformation.

<sup>&</sup>lt;sup>52</sup> Velkley draws our attention to R 4468 (1770s): "That reason requires an upbringing [einer Zucht]. That, if it is not cultivated [gezogen ist] but extends its branches wildly, then it brings blossoms without fruits" (17:562-3). The idea that reason requires an upbringing, and further that the *Critique* is in effect bringing that off, leads to the unfortunate implication that critical philosophy is an "art" rather than a "science", given that (e.g.) Aristotle thought of upbringing as an art (*techne*). But in the Doctrine of Method, it becomes clear that critical philosophy is not an art. Speaking of philosophy in a way that clearly includes critical philosophy and excludes dogmatic metaphysics, Kant writes: "philosophy is the science of the relation of all cognition to the essential ends of human reason (*teleologia rationis humanae*), and the philosopher is not an artist of reason [Vernunftkünstler], but rather the legislator of human reason. It would be very boastful for one to call oneself a philosopher in this sense, and presume oneself to be equal to the archetype, which lies only in the idea" (A839/B867). This passage is crucial — but in different ways — to both Velkley's and Neiman's accounts. See also R4467 (17:562, early 1770s) and R4925 (18:30, c. 1776-8), which both make more or less the same point.







## Figure 2 Kant's Division of the Philosophy of Pure Reason (from A841/B869 ff.)

## **Comment on Figure 2:**

Kant's division of the philosophy of pure reason is straightforward up until the division of the metaphysics of nature. After that point confusion reigns, as it is not clear whether Kant's taxonomy is supposed to concern solely projects considered legitimate by the argument of the *Critique* itself. It seems it cannot, given the heading "transcendent rational physiology", which would include the dogmatic projects deemed fraudulent in the Transcendental Dialectic.

It is also unclear what to make of the heading "transcendental philosophy", included under the metaphysics of nature: what is its relation to critical philosophy, which is on another branch altogether? Transcendental philosophy, Kant says, "considers only the *understanding* and reason itself in a system of all concepts and principles that are related to objects in general, without assuming objects that *would be given (Ontologia)*". Ontology was traditionally the "first part" of what Kant deems "dogmatic metaphysics". Perhaps Kant has traditional ontology in mind as "transcendental philosophy" in this passage. The future metaphysics for which critical philosophy is the "propaedeutic" would probably be the "rational physics" listed here under "immanent rational physiology". (I am, however, unsure what to make of the "immanent rational psychology" listed here.)

In the *Metaphysical Foundations of Natural Science*, Kant remarks that transcendental philosophy is concerned with the "laws that make the concept of a nature in general possible, even without reference to any determinate object of experience" (*MAN*, 469). This sounds like a perfectly good description of critical theoretical philosophy — and so it doesn't help to explain why Kant lists them under different phyla in the diagram here. The most reasonable conclusion to draw is probably the following: for general Kantian purposes, the term "transcendental philosophy" simply picks out the first episode of critical philosophy (i.e., the *Critique of Pure Reason*).

## CHAPTER 2 REFLECTIVE METAPHYSICS: A PRELIMINARY ACCOUNT

Already in the nature of our reason lies this distinction of matter and form. Matter is the datum, what is given [...]. But form is how these data are positioned, the way the manifold stands in combination. [...] The ancients placed a great deal on the form; they said that it was the essence of things. And that is entirely right; for in no thing can we produce the matter, but rather only the form [...].

*— Metaphysik L*<sub>2</sub> (28:575)

In the form consists the essence of things (*forma dat esse rei*, as the scholastics say), in so far as this is to be known by reason. [...M]etaphysics as pure philosophy grounds its cognition above all on *forms of thought* under which every object (matter of cognition) may then be subsumed.

—"On a Newly Elevated Distinguished Tone in Philosophy" (8:404)

1.

The previous chapter concluded with a suggestion about how we might come to terms with the idea that the *Critique* is a project of self-knowledge: namely, that we should see it as simply making something explicit that is ordinarily only implicit in our cognitive lives. This makes the spirit of Kant's work seem remarkably democratic: there would seem to be no particular prerequisite to pursue critical philosophy. But we could put a finer point on this: the ideal preparation for critical philosophy would be the practice of judgment that animates the various sciences of nature. For the successful pursuit of these sciences, Kant tells us in the B-edition Preface, rests on some tacit appreciation of the distinction between what we put into our demonstrations and what we do not. The banner slogan of scientific cognition is that reason "has insight only into what it brings forth according to a plan of its own" (Bxiii). Scientific cognition, his account suggests, turns on some kind of self-concern of our cognitive capacity, which is concretely (but tacitly) made manifest through the particular investigative practices of the sciences themselves. Shortly thereafter, Kant suggests that *critical philosophy* requires a skilled capacity to distinguish the "addition" (*Zusatz*, B1-2) provided by our cognitive capacity and separate from the composite known as "experience" or empirical cognition. Conceivably, the aim of the *Critique* is to bring its reader to recognize the sources, and hence the possibility, of scientific cognition in general.

Yet this gloss on the *Critique* as a project of 'making explicit' obscures an equally important characteristic of it. Indeed, Kant says the *Critique* must accomplish something more than merely making explicit "what we should have already practiced in the merely empirical use of the understanding" in order to be worth its taxing effort (A237/B236). The *Critique* has metaphysical aspirations: it is a reflective project of self-knowledge that is supposed to yield positive conclusions about the viability of a metaphysics of material nature. Its ideal reader, then, would have metaphysical aspirations as well. He should not have been left entirely satisfied by his investigation into nature's phenomena, no matter how successful he may have been. Kant's ideal reader would conceivably have been compelled, even in the midst of his research, to wonder about the very domain into which he inquires — to wonder not just about some particular array of phenomena, but also to wonder about nature *as such*.

At the close of the Transcendental Analytic, in Phenomena and Noumena, Kant paints a fresh picture of his ideal reader. "[T]he understanding that is occupied merely with its empirical employment [...] may get along very well," he admits. Yet there is one thing that this "understanding" cannot achieve if it "does not reflect [nachsinnt] upon the sources of its own cognition": it cannot "determine for itself the boundaries of its employment and know what may lie inside or outside its entire sphere" (A238/B297). Its "sphere" is nature itself; and its reflection upon the "sources" of its cognition apparently enables it to determine — in an exhaustive or comprehensive way — what belongs to nature *as such*. The Transcendental Analytic is supposed to have brought about a reformation of the reader's metaphysical aspirations: the reader now recognizes that metaphysics is, at least in its first part, a *reflective* project.

The aim of this chapter is to offer a preliminary account for the unity of the two fundamental conceptions of the *Critique*: it is both a project of self-knowledge and it lays the foundation for 'any future metaphysics' of material nature.<sup>1</sup> The unity of these two projects could be expressed simply by saying that the *Critique* is *reflective metaphysics*. In broad outline, the idea of reflective metaphysics is this: an account of our cognitive capacity yields conclusions about the domain of nature as such. My aim in the present chapter is to give some preliminary idea of what this would involve, focusing on what Kant takes "reflection" to be. I will have more to say about reflection in Chapter 3, where I examine the status of the *Critique* as a project of rational self-knowledge more closely.

<sup>&</sup>lt;sup>1</sup> The first conception is stated most clearly at Axi-xii, Bxxxv, and A849/B877, and the second conception at A11/B25 and A841/B869.

2.

If the *Critique* is to be a project of reflective metaphysics, then it goes without saying that its subject matter would be, at least in part, our cognitive capacity. But if it is to allow for genuinely *metaphysical* conclusions at all, then it cannot abstract from the relation of this cognitive capacity to its objects. The objects at issue belong to the realm of material nature. Our concern with them is "metaphysical" rather than "physical" — and this, I take it, means that we are not concerned with *particular* objects of nature, but with what it is to *be* an object in that realm.

In empirical natural science, we are directly concerned with objects in the domain of material nature. We are concerned with actual objects as well as with "possible" ones — that is, possible objects of actual experience or perception, which we might call "theoretical entities". (The fact that we can legitimately postulate theoretical entities, things known "comparatively a priori", stems from the fact that there is an a priori basis to natural science, a "form of possible experience"; in arriving at theoretical entities we trace from actual perceptions to possible perceptions under the guidance of the "analogies of experience" (A225/B273). Theoretical entities are entities that *could* be observed, given the right equipment.<sup>2</sup>) But in critical philosophy, we are not directly concerned with these objects; we are, nevertheless, still concerned with nature. We are concerned with the *possibility* of objects of experience. The critical-metaphysical investigation aims to determine the 'being' of nature — i.e., what it is to figure in its realm.

<sup>&</sup>lt;sup>2</sup> I suppose that we are no longer in a position to say the same things about the theoretical entities of (e.g.) quantum physics, at least not in any straightforward sense. But for more of Kant's views about theoretical entities, see *Entdeckung*, 8:205, where he argues that there is no ground for supposing that theoretical entities are "non-sensible [...] objects of reason": "Newton's *lamellae*, of which the colored particles of bodies consist, have not yet been seen through a microscope. Nevertheless, the understanding not only recognizes (or supposes) their existence, but also that they really are represented in our empirical intuition, albeit without being consciously apprehended".
In putting things this way, I reveal my hand: the metaphysics that the *Critique* is engaged in is ontological in character.<sup>3</sup> It is a contentious claim, but only if the notion of ontology is attributed wholesale to dogmatic metaphysics. In Phenomena and Noumena, Kant himself disclaims the "proud name of an ontology", which "must give way to the modest one of a mere analytic of the pure understanding" (A247/B303). Critical philosophy is not "ontology" in the dogmatic sense that Kant has in mind here, simply because it is not making claims about the being of objects independently of what it is for them to be *known* — at least by us.

Among Anglo-American commentators, the idea that Kant's critical philosophy is not "ontological" but rather "epistemic" in character is commonplace. Henry Allison's influential work speaks of the categories and the pure forms of intuition as "epistemic conditions".<sup>4</sup> These are the conditions in virtue of which representations relate to objects or have "objective reality" (leading Allison to the alternate terminology "objectivating conditions", 10). Epistemic conditions are not ontological conditions, Allison goes on to emphasize; the former are conditions for the possibility of *experience* or knowledge of things, and the latter conditions for the possibility simply of the being of things. And thus, the former are conditions of the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are for us, and the latter conditions for the possibility of things as they are in themselves (11).

<sup>&</sup>lt;sup>3</sup> Commentators writing in English tend to deny the "ontological" and emphasize the "epistemological" aspect of the project, while commentators writing in German often seem to do the reverse. As far as I can make out, 'ontological' readings of the *Critique* among German commentators vary widely. Heidegger's account, put forward in *Kant and the Problem of Metaphysics* and *Phenomenological Interpretation*, is — like much of Heidegger's exegetical work — keyed to his own philosophical project. Somewhere closer to standard exegesis, Manfred Baum urges us to recognize that the *Critique* is a work of metaphysics — a "metaphysic of metaphysics", as Kant noted — and claims that one must appreciate its ontological concerns in order to grasp the Transcendental Deduction properly (see "Transcendental Proofs", *passim*; and *Deduktion und Beweis in Kants Transzendentalphilosophie*, 12). Hans Graubner argues that Kant reconceives of ontology through his concept of form: there is an "ontological component" to Kant's critical philosophy, which nevertheless avoids the errors of previous metaphysics (*Form und Wesen*, 12 ff.).

However, Kant's exclusion of "things in themselves" from any science of nature ought not to merit the inference that his critical metaphysics is non-ontological in character. My objection — which stems from a concern to direct our attention to Kant's idea of "form" — is not directed so much against Allison's fundamental point but against his injudicious ascription of "ontological". Allison's terminology of "epistemic conditions" papers over Kant's calling these representations "forms"; as a result, the ancient connection between the notion of form and that of being also falls out of view.

Consider the very passage where Kant most emphatically suggests that the *Critique*, or the Transcendental Analytic at any rate, is not an "ontology":<sup>5</sup>

The Transcendental Analytic accordingly has this important result: that the understanding can never accomplish a priori anything more than to anticipate the form of a possible experience in general, and since that which is not appearance cannot be an object of experience, it can never overstep the limits of sensibility, within which alone objects are given to us. Its principles are mere principles of the exposition of appearances, and the proud name of an ontology, which presumes to give synthetic a priori cognitions of things in general in a systematic doctrine [...] must give way to the modest one of a mere analytic of the pure understanding. (A246-7/B303)

Kant appeals to the notion of form to distinguish what the *Critique* is doing from what it is not: its claims address the "form of a possible experience in general". Elsewhere, Kant

<sup>&</sup>lt;sup>4</sup> Kant's Transcendental Idealism, 10ff.

<sup>&</sup>lt;sup>5</sup> N.B.: In several other places, Kant indicates that he thinks of the Transcendental Analytic as a reconceived, i.e., non-dogmatic, ontology. This is implied by his remarks about the need to separate the concepts of reflection from the pure concepts of the understanding in the *Prolegomena*; the concepts of reflection "cannot intrude into ontology", he says, suggesting that the account of the pure concepts of the understanding in the Transcendental Analytic proper *is* an ontology (*P* §39, 326). And in the *Critique*'s second edition Preface, Kant claims that his 'Copernican' experiment for philosophy "promises the sure path of a science to *metaphysics in its first part*" (Bxviii, my emphasis): the "first part" of metaphysics, traditionally, was ontology. The project that he seems to conceive as a non-dogmatic ontology is "concerned with the a priori concepts of which the corresponding objects appropriate to them can be given in experience".

speaks of "pure intuition" as containing "solely the form under which something is intuited", and "pure concept" as containing "only the form of thinking [Form des Denkens] of an object in general" (A50/B74). At the same time, Kant generally speaks of "forms of thought" when he wants to emphasize the emptiness of the pure concepts of the understanding in the absence of intuition.<sup>6</sup> As we saw in chapter 1, the rational selfconcern proper to critical philosophy would be distinguished from the rational selfconcern proper to pure general logic because it is self-*knowledge*. To distinguish critical philosophy from logic is to understand it as a kind of metaphysics. But the *Critique* is not "contentful" in virtue of making claims about objects. It is contentful because it demonstrates the a priori basis on which the pure forms of intuition and the forms of thought cooperate in the production of knowledge. It "anticipates the form of a possible experience in general". In this respect, it yields an account of the "being" of objects in the realm of nature.

3.

In order to take the idea of a "reflective metaphysics" any further, it will help to consider what Kant tells us about *reflection*. In both the Amphiboly and the *Jäsche Logic*, Kant claims that reflection plays a role in the development or formation of conceptual capacities. Kant's remarks about reflection in the Amphiboly are complicated by the fact that his immediate goal there is to advance a polemic against Leibniz. For this reason, I will mostly draw on the *Logic* to examine Kant's conception of reflection. Still,

<sup>&</sup>lt;sup>6</sup> Indeed, the phrase is usually prefixed with a "mere" ("bloße Gedankenformen", B148; B150; B288; B305; and *P* §57, 355). See also the first paragraph of an unsent draft of Kant's 13 October 1797 letter to Tieftrunk, addressing J.S. Beck's suggestion that the Analytic precede the Aesthetic, where he speaks of the

we should handle those remarks with care, since they belong to the work of "pure general logic" rather than the "transcendental logic" with which we are principally concerned in the *Critique*.<sup>7</sup> Notwithstanding this limitation, Kant's account of reflection in the *Logic* will allow us to appreciate what "reflection" and "form" have to do with one another. And this will put us in a position to return, at the end of this chapter, to the idea that the *Critique*'s reflective metaphysics is a reconceived ontology.

Concepts are universal representations: particulars fall under them. In the *Logic*, Kant claims that concepts owe their universality to a three-fold "logical act of the understanding [logischen Verstandes-Actus]" (§6, 94), which is comprised of *comparison, reflection,* and *abstraction.* Our *given* representations are particular or singular;<sup>8</sup> they are presentations of actual or putatively actual states of affairs. Concepts may indeed be involved in our passive enjoyment of given representations — involved, that is, in some way that is distinct from judging or predication — but that is an issue that I want to set aside for the time being, to focus on the idea that given representations are particular or singular, while concepts are universal.

The account in the *Logic* is illustrated by the formation of an empirical concept (*tree*). In the example at hand, the three aspects of this "logical act of the understanding" are presented as following one another: I compare several given representations and notice that they are different from one another in various respects, but also have

<sup>&</sup>quot;mere form of thinking" as "concepts without objects, concepts that as yet are without any meaning" (13:463; Zweig, 238).

<sup>&</sup>lt;sup>7</sup> The *Jäsche Logic* presents an account of the origin of concepts, which is advertised, not surprisingly, as a logical account: it 'abstracts from all content'. It is concerned only with "the origin of concepts with respect to mere form" (*Logic* §5, 93), and not with how concepts may differ according to their content, or objective purport.

<sup>&</sup>lt;sup>8</sup> See also A320/B376-7, Kant's taxonomy of "representation in general".

something in common ("the trunk, the branches, the leaves themselves"); I abstract from, or leave out of consideration, what makes them different from one another, and arrive at the concept *tree*. Representations are *compared* "under one another in relation to the unity of consciousness"; we *reflect* as to "how various representations could be conceived [begriffen] in a single consciousness", and "finally" comes the *separation* (*Absonderung*) "of everything else in which the given representations distinguish themselves" (§6, 94).<sup>9</sup> Reflection, in this account, is some kind of appreciation of a possible unification of consciousness.

The fundamental difference between particular and universal representations, or between intuitions and concepts, can be expressed modally: intuitions pertain to the actuality of an object, while concepts have to do with the possibility of an object. What if we imagine a non-present object? We have a given representation; it may not be immediately caused by the presence of an actual object (because we hallucinate or deliberately conjure the image), but it is a representation which purports to put something actual, in all of its particularity, into view. Perhaps it is obvious that particular, given representations announce the actuality of an object; less obvious is how concepts pertain to the possibility of an object. In the *Critique of Judgment*, Kant suggests that the fundamental capacity of the understanding is to distinguish between the possibility and the actuality of things:

It is unavoidably necessary for the human understanding to distinguish between the possibility and actuality of things. The reason for this lies in the subject and the nature of its cognitive faculties. For if there were not two entirely heterogeneous elements required for this exercise of these

<sup>&</sup>lt;sup>9</sup> See also *Anthropology* (134n.), where the "inner activity (spontaneity) whereby a *concept* (a thought) is possible" is glossed "*reflection* (*Reflexion*)"; it is distinguished from "the receptivity (Rezeptivität) whereby a *perception* is possible" (134 n.).

faculties, understanding for concepts and sensible intuition for objects corresponding to them, then there would be no such distinction (between the possible and the actual). For if our understanding were intuiting, there would be no objects except the actual. Concepts (which merely pertain to the possibility of an object) and sensible intuitions (which give us something, without thereby allowing us to cognize it as an object) would both fall by the wayside. Now all of our distinction of the merely possible from the actual rests on the fact that the former signifies only the position of the representation of a thing to our concept and in general to our faculty for thinking, while the latter signifies the positing of the thing in itself (apart from this concept). (KU §76, 401-2)

Certainly a heady passage — but the thought it expresses is contained in the more sober lines from the *Logic*. In the *Logic*, Kant makes a point of correcting what he takes to be a misuse of the term "abstraction": "We must not say that we 'abstract something' [...] but rather we 'abstract from something' [...]" (§6, 95). It looks, at first blush, to be a comment of miniscule significance, but it is actually quite fundamental. Abstraction, he goes on to say, is a "negative condition under which objectively valid representations can be developed", while comparison and reflection are presumably the positive ones. Abstraction is a kind of not-considering. Now, let us admit as a hypothesis that concept formation involves a kind of *separation* (i.e., abstraction). Kant is warning against our supposing that this separation is a *positive* activity, a thing we do; abstraction is expressly something that we *don't do*. Abstraction is leaving the particularity of our given representations out of consideration. One may then wonder: what then is left? If we are dealing with particular, given representations, and we leave out of consideration their particularity, then what exactly do we have in view? The answer is that we have in direct view not so much those given representations themselves, but rather a relation by means of which those given representations can be considered as belonging together. Concepts are ways in which given representations *can* stand in some cognitively significant relation to one another. Particular (i.e., given) representations can be thought of as *matter*, and the concept expresses the possibility of a determinate relation of those given representations. Abstraction is a disregard of matter (everything "in which given representations distinguish themselves" (95)). This amounts to a kind of disregard of the given representation itself — or rather, an *indirect* regard of it. Reflection is the positive counterpart of abstraction. It is attentiveness to form, or the way in which matter can be arranged, composed, or related.

In the passage from the *Critique of Judgment*, Kant claims that the ability to distinguish between "the possibility and the actuality of things" is "unavoidably necessary for the human understanding". A creature that could not draw this distinction between the possibility and the actuality of things would either have no self-consciousness at all (e.g., a beast) or it would be an intuiting intellect, representing the actuality of things through sheer self-consciousness. In either case, the creature in question would not be able to *distinguish* itself (as a representing capacity) from any object. For Kant, reflection involves indirect regard of the "actuality of things" (given representations) that yields some appreciation of the "possibility of things" (a concept). A concept is thus an expression of an active unification of consciousness. Reflection is an awareness of this active contribution of our cognitive capacity, that is concretely determined by the particular representations that are held in merely indirect regard.

In the *Anthropology*, Kant claims that reflection is pure apperception, or "consciousness of the understanding" (134n.); I will discuss this idea further in Chapter 3. Yet this remark tells us something about the limitation of the *Logic*'s account: it presents reflection as a consciousness of particular conceptual capacities. It does not manifestly present reflection as a consciousness of the *faculty* for concepts — i.e., reflection as the consciousness of the understanding as such. Moreover, the account in the *Logic* is about our 'getting from' singular representations to universal representations. It thus seems to be suited to empirical concepts alone. Perhaps we should question its relevance. Is this the *same* "reflection" that figures in the idea that the *Critique* is "reflective metaphysics"?

Yes, at least generically. Kant says that the account in the *Logic* holds of all concepts, whether "*empirical*, [...] *arbitrary* [...], or *intellectual*" (94); intellectual concepts, he says, stem from the "nature of the understanding" alone. So the account in the *Logic* should indicate something about the intellectual concepts that are at issue in the *Critique* (e.g., the pure concepts of the understanding). Let us consider what that lesson could be.

The account in the *Logic* presents reflection as an awareness of conceptual capacities. In the case of empirical concepts, this capacity is specified by the singular representations that are held in indirect regard. These singular representations can be thought of as the "matter" of the concept; the "form" is presumably what is directly attended to in reflection. The form that is appreciated in reflection is that in virtue of which particular, given representations are seen as potentially standing in cognitively significant relations.

Empirical concepts do not arise from the nature of the understanding alone, but require a set of given representations which are appreciated as sharing some feature or set of features. The "matter" (the set of given representations) is there *prior* to our having the concept in question. Any member of this set of given representations which constitutes the "matter" of an empirical concept can be enjoyed without the concept in question: in this respect, the matter (as an intuition) does not depend upon the determining form merely in order to show up for us at all. Thus, while I may need the concept *rotund* to see or appreciate something *as rotund*, as standing in a certain unified relation to other rotund things, I do not need the concept *rotund* to have the item in question before my empirical consciousness in the first place. The objects are there anyway, whether or not we ever bring this or that empirical concept to bear upon them. They may be given merely in intuition, independently of my appreciation of the rotundity that they all share.

Kant's account in the *Logic* invites us to make a form/matter distinction when thinking about empirical concepts. The same sort of form/matter distinction is not naturally made with respect to pure concepts, i.e., the categories. The categories are the "forms of thought", and any particular empirical concept, I want to suggest, sets the given representations in question in relation according to one or more of these forms. Or rather, any empirical concept sets given representations in relation according to one or more of these *schematized* forms — i.e., according to one or more of the principles of the pure understanding. This is implied by Kant's claim that the principles of the pure understanding (mathematical and dynamical alike) are *constitutive* with regard to *experience*, which he glosses in the following way: the principles of the pure understanding "make possible a priori the *concepts* without which no experience would take place" (A664/B692). This stands in contrast to the principles of pure reason, which "cannot be constitutive even in regard to empirical *concepts*, since no corresponding schema of sensibility can be given for them, and therefore they can have no object *in*  *concreto*". The idea that the principles of the pure understanding are *constitutive* with regard to experience means that they make empirical concepts possible; it suggests that empirical concepts owe their "form" to these principles, and their "matter" to whatever given representations are at issue. Kant makes the point explicitly, in his remark that an empirical concept (*Erfahrungsbegriff*) is "nothing but a concept of the understanding<sup>10</sup> *in concreto*", and that "in appearances" the pure concepts of the understanding "originally have the matter [Stoff] for the empirical concept [Erfahrungsbegriff]" (A567/B595). In this sense, the categories — and indeed even the principles of pure understanding — are *just* forms, forms without matter.<sup>11</sup> Empirical concepts, by contrast, are composites.

For Kant, reflection is most generically thought of as a "consciousness of the understanding". In the *Logic*, reflection figures in the account of the origin of particular conceptual capacities. But in order to carry out a project of reflective metaphysics, we would need something of an entirely different register: we would need, in the first place, an exhaustive account of our cognitive capacity — and not this or that particular conceptual capacity. If there is a role for reflection to play in such a project, it would need to be a "consciousness of the understanding" *as such*. The *Critique*'s Analytic of Concepts promises exactly this: "*an articulation of the faculty of understanding itself*, in order to examine the possibility of a priori concepts" (A65/B90). It is the first main part of the *Critique*'s Transcendental Logic, and it prepares us for the conspicuously "metaphysical" work of its second part, the Analytic of Principles.

In Chapters 3 and 4, I will focus on the role of reflection in the Analytic of Concepts. In preparation for that, it will help to have a better sense of where Kant is

<sup>&</sup>lt;sup>10</sup> "Verstandesbegriff" — the context indicates clearly that he means a *pure* concept of the understanding.

headed. What sort of claim about nature can we expect to make on the basis of an account of our cognitive capacity? To answer this question, I will first consider what conception of nature figures in the background of Kant's project, focusing on Kant's distinction between *formal* and *material* senses of "nature". Finally, I will conclude with some programmatic remarks about *how* Kant expects to move from an account of our cognitive capacity to a metaphysics of material nature.

## 4.

It is noteworthy that the concept of nature does not figure in the main argument of the *Critique* until the concluding paragraphs of the Transcendental Deduction. As we will see in Chapter 4, this has to do with the *Critique*'s synthetic method: the argument is supposed to take nothing as given except reason itself and argue *to* an account of the fundamental principles of material nature. When we get to the end of the Deduction, Kant distinguishes between two senses of nature, material and formal. We will need to understand this distinction in order to understand the metaphysical project of the *Critique*. But since our immediate goal is to acquire some preliminary grasp of what reflective metaphysics involves, it may help to turn to the *Prolegomena*, where the concept of nature figures prominently from the outset.

In the *Prolegomena*, Kant distinguishes between two senses of nature, material and formal. Nature in the material sense is conceived as the "totality" of everything that is there to be experienced. Kant also distinguishes *two* formal senses of nature. One is recognizably Aristotelian, and is rejected; the other is not recognizably Aristotelian, and

<sup>&</sup>lt;sup>11</sup> For an opposing view, see Richard Aquila, *Matter in Mind*.

is embraced. According to the formal sense of nature that Kant wants to accept, nature is the "totality of the rules under which all appearances must come in order to be thought as connected in an experience" (§39, 318). This can only be "possible", he says, "by means of the constitution of our understanding". Presumably, this "formal sense" of nature could be revealed through *reflection*, if reflection is a "consciousness of the understanding".

Kant presupposes at least the minimal parameters of the dominant early-modern conception of nature: there are no little bits, gaps, or dark corners, of nature that are *as such* cognitively inaccessible. Nature is commensurate to, or in some relation of "affinity" with, our cognitive capacity. The Kantian version of the modern conception of nature stresses the idea that the "constitution of our understanding" is a condition of the possibility of nature, and is therefore a kind of idealism: nature is not what it is independently of its relation to our cognitive capacity. This issue is obviously related to the basic problem about how the two conceptions of the project are united — that is, how a "consciousness of the understanding" could lay the foundation for any future metaphysics of material nature.

The two conceptions of the project seem to be united through the concept of *form*: the "formal" sense of nature is spelled out through the exhaustive account of our cognitive capacity, which presents the principles that constitute the "form" of an object of experience in general. These principles purport to tell us 'what it is' to figure in the domain of nature at all. In this respect, Kant's appeal to form borrows *something* from the Aristotelian conception of form, according to which *form* is that in virtue of which something is 'what it is', and that without which it could no longer be 'what it is'. A

finger is 'what it is' in virtue of being able to perform a finger's characteristic functions; these performances are the expression of the finger's form. Here Aristotle might distinguish between the "second actuality" of the finger (its essential or characteristic performances as a finger), and the "first actuality" of the finger, which would be its *capacity* to be a finger. When the finger is "dead" — severed or paralyzed, say — it loses this capacity. It is no longer a finger, according to Aristotle, or is a finger only by courtesy and only in name (*Metaphysics* Z.10, 1035b23 ff.).

As I already noted, there are two "formal" conceptions of nature floating around in Kant's discussions. One is obviously Aristotelian in origin; it draws on the Aristotelian idea that the "nature" of a thing is its "internal principle of change and rest".<sup>12</sup> But for Kant, the relevant sense of "form" pertains to that which determines 'what it is' to figure in the domain of nature at all. In the *Metaphysical Foundations of Natural Science*, Kant first gives a definition of nature "merely in *formal* meaning" as "the first inner principle of everything that belongs to the existence [Dasein] of a thing" (467). Kant says that this conception of nature is "adjectival": nature or form is *essential to a kind of thing*, and by extension, to particulars considered as belonging to natural kinds. Kant distinguishes this conception of nature from a "substantival" conception of nature. This is the conception of nature that Kant is after: it is "substantival" because it pertains to nature *as such*, and not the nature of this or that kind of thing.

In the *Metaphysical Foundations*, the substantival conception of nature is associated with the "*material* meaning" of nature: "the totality of all things in so far as they can be *objects of our senses*, and thus also of experience, under which is therefore

<sup>&</sup>lt;sup>12</sup> See (e.g.) Aristotle, *Physics* B.1.

understood the whole of all appearances, i.e., the world of sense to the exclusion of all non-sensible objects" (467). Very similarly, in a footnote in the Antinomies, he writes:

'Nature', taken adjectivally (*formaliter*), signifies the connection of determinations of a thing, according to an inner principle of causality. However, one understands under 'nature' taken substantivally (*materialiter*) the totality of appearances in so far as these stand in thoroughgoing connection by means of an inner principle of causality. In the first sense, one speaks of the nature of fluid matter, of fire, etc., and employs this word adjectivally; however, if one speaks of the 'things of nature', then one has in mind a subsisting whole. (A418-9/B446 n.)

What does this suggest about the role the concept of form plays in Kant's conception of nature, and ultimately, his conception of metaphysics? The two passages quoted here are merely lexical: they alone cannot answer the question. However, the "substantival" sense of nature invokes an idea of nature in its totality, as a *whole*, suggesting that Kant may prefer that conception. Could there be a "formal" aspect of this "substantival" sense of nature? Yes — but this emerges more clearly in the *Prolegomena*.

In the *Prolegomena*, Kant argues for a formal conception of nature which is distinct from the "adjectival", blatantly Aristotelian one invoked in the two passages above. He begins his analysis of pure natural science in §14 with the following definition of 'nature': "*Nature* is the *existence* of things in so far as it is determined according to universal laws" (§14, 294). After ruling out of court a conception of nature which admits "things in themselves", he again invokes the Aristotelian conception of nature, and contrasts it with some other conception of nature that is clearly his principal concern.

The word *nature* assumes another meaning, namely [that] which determines the *object*, whereas in the former meaning it signifies only the *conformity to law* [*Gesetzmäßigkeit*] of the determination of the existence

of things in general. Thus nature, regarded *materialiter*, is the *totality of all objects of experience*. We are here concerned only with this [...]. (§16, 295; see also *KrV* §26, B163)

Here he contrasts what is presumably the adjectival, Aristotelian conception of nature (now presented as "what determines the object") with the conception of nature that he introduced at the outset of §14: the "*existence* of things in so far as it is determined according to universal laws". Again the adjectival conception is set aside, and a material conception of nature is introduced.

This remark gives the false impression that Kant only cares about the "material" conception of nature. In fact, Kant goes on to express his interest in "the *formal* [conception] of nature" in §17, which is "the conformity to law of all objects of experience, and in so far as they are cognized a priori, the *necessary* conformity to law of them" (296). Thus there is clearly some formal conception of nature that Kant wants, and it seems to follow from the material conception of nature.

Let us consider the formal sense of nature that Kant wishes to reject. In this sense, nature is a determination of the *object* full stop, i.e., irrespective of any *relation* in which that object stands to a cognizing subject. Moreover, this rejected formal conception of nature admits of the possibility of a kind of fragmented nature: a collection of singletons or unit sets, each containing a single entity governed by its own law. The determination of nature's entities might still be "according to law", but there could be a law for each individual thing. Each law of nature could hold universally of a singleton! The rejected formal conception of nature, after all, says only that nature is what determines the object; it does not suppose that there is any necessary connection between the determination of one object and the determination of any other. A nature of this sort

would not admit of formal expression as a totality of laws, because any one law would not necessarily reach through the whole of nature, but possibly only through a singleton.

One way to understand all of this is to see that Kant is rejecting a Leibnizian conception of nature (one which accommodates self-contained monads), which admittedly owes something to the Aristotelian notion of form.<sup>13</sup> But that does not mean that Kant advances a complete rejection of an Aristotelian conception of form. He in fact embraces the notion of form, as long as it is recognized that "form" is something the subject provides for the cognition of objects (which are *given* from elsewhere). This point emerges in Kant's defense of his "formal idealism", which he distinguishes from the "material idealism" of Berkeley and Descartes.<sup>14</sup> Material idealism makes a claim about the ideal or mental nature of the *actual* existence of objects. There are two kinds: the *dogmatic* brand, found in Berkeley, claims that all objects are merely 'ideal' or 'in the mind'; and the *skeptical*, found in Descartes, which claims that the existence of objects outside of us is merely uncertain (B274). Formal idealism, on the other hand, is not concerned to make any claims about the *actual* existence of objects at all; its point is that the *possibility* of objects is 'ideal' or mental.

In the formal conception of nature that Kant rejects, "form" is the "inner principle" which determines the thing; and the rejection lies in the fact that the form is *proper to the thing alone*. The idea that nature is what "determines the object" does not involve any implicit thought to the effect that the determination of one object would be

<sup>&</sup>lt;sup>13</sup> Of course, Leibniz had ideas about the orchestrated harmony of these self-contained entities. I am not claiming to present Leibniz's conception of nature, but merely one that we could reasonably call "Leibnizian".

<sup>&</sup>lt;sup>14</sup> Kant agonized in print over having named his brand of idealism "transcendental", thinking that the moniker was partially responsible for the widespread misunderstanding of the *Critique*. He said that he ought to have called it "critical idealism", or "formal idealism" (P, 375).

related to the determination of another, let alone all others. Nature on this conception is not a relation: it is neither a relation among objects, nor a relation of what is present in nature to the possibility of its being known. That these two ideas of nature as a relation are intimately related becomes clear when we consider the "formal" conception of nature that Kant wants to embrace: "the conformity to law of all objects of experience [...]". The formal conception of nature that Kant wants to accept arises through consideration of the material conception of nature. It distinguishes the formal basis of that idea of nature, namely that any object in nature conforms, *as* an object of experience, to certain laws.

This "formal" conception of nature is an expression of nature as a totality. No *actual* experience is adequate to represent nature as a whole; but the "form of a possible experience in general" is an expression of nature as a totality, which has its basis in the principles of the pure understanding, and thus ultimately the categories. These categories are "forms of thought". But as "forms", they are proper to the cognizing subject, and are not, as in a certain reading of the Aristotelian conception of form, proper to the object. At bottom, Kant is rejecting a Leibnizian appropriation of the notion of form, in which form may have to do with being, but not necessarily with being known.<sup>15</sup>

Leibniz's invocation of the Aristotelian idea of "form" as the "inner principle of a thing" leads to a world of monads, as Kant (in effect) argues in the Amphiboly discussion of the pair "inner" and "outer" (A265-6/B321-2). The "internal" properties of "an object of the pure understanding" have no relation to anything outside of, or different from, it. But if we prescind from the rationalist view that sense representations are confused

<sup>&</sup>lt;sup>15</sup> At least by us. An intellectual intuition would indeed have *no use* for the forms of representation which for a finite understanding are the very conditions for the possibility of synthetic a priori cognition. This point is related to Kant's claim at KU §76 (quoted above) that the cognition of an intuiting understanding (i.e., intellectual intuition) could bear only upon the *actuality* of objects.

versions of intellectual representations, then the "inner determinations of a *substantia phenomenon* in space are nothing but relations [Verhältnisse], and it itself [is] entirely and utterly a totality of sheer relations [Relationen]" (A265/B321). Thus, only under the rationalist assimilation of sensibility to the understanding does the notion of an "inner principle" or an "inner determination" lead to the supposition that the world must be constituted by self-contained, perceiving monads. Without this rationalist presupposition, the talk of an "inner principle" does not bespeak an *absolute* entity; instead we get an idea of an "inner determination" of a thing that does not seal it off from all other things, but rather locates it in a common space *with* all other things. Kant is not banishing the notion of form from his conception of nature, and thus he is not banishing the notion of form from his conception of theoretical metaphysics. Kant may be rejecting a Leibnizian appropriation of the Aristotelian concept of form; but he is himself reinterpreting the Aristotelian idea of form, with the very idea of "formal idealism".

5.

How can we connect what we know about reflection from the *Logic* with Kant's interest in a "formal" conception of nature? Reflection, the *Jäsche Logic* taught us, can be thought of as an awareness of form; it presents reflection as an awareness of conceptual capacities. In the *Critique*, we are interested in the *faculty* of concepts; the reflective work of the *Critique* is a "consciousness of the understanding" *as such*. What preliminary conclusions can we now draw about the project of reflective metaphysics?

Kant claims that the fundamental capacity of the understanding is to distinguish between the possibility and the actuality of things: it is "unavoidably necessary for the human understanding" to make this distinction, he says (KU §76). What is it to draw this distinction? The account from the *Logic* suggests an answer: it is to *reflect*. Perhaps this is what Kant has in mind when he claims that "the understanding does not intuit anything, but rather only reflects [reflektiert nur]" (P, 288).

The phrase "reflective metaphysics" is my own; Kant, as we have seen, prefers the term "formal idealism". But the upshot is the same either way: the "possibility" of things is ideal, or "lies in us". Kant's formal idealism, or reflective metaphysics, is concerned with the contribution that the cognizing subject makes to experience. To understand what this contribution is, Kant claims, is to have a handle on the very "possibility of things". The way to end up with an exhaustive account of the principles expressing what it is to figure in the domain of nature at all is to determine them as they are united in the concept of a cognizing subject.

In this chapter, I have focused on the idea that reflection is a consciousness of the *understanding*. But this may be a slightly misleading way of putting things. At the beginning of the *Critique*, as we have seen, Kant hints that its project of reflection will be concerned with the *contribution* of our cognitive capacity to experience (B2). There Kant tacitly invokes a distinction between form and matter: this contribution is distinguished from the "fundamental material" (*Grundstoffe*) of cognition. If this is meant to anticipate the reflective work ahead, then once again we have the idea that reflection is an awareness of *form*. But as we soon learn, right at the outset of the Transcendental Aesthetic, the formal contribution of our cognitive capacity is not to be attributed to the understanding alone: there seems to be some formal contribution that is attributed to *sensibility*. And this suggests that what we have said about reflection is still somewhat

confused. Is reflection a consciousness of the understanding, per Kant's gloss in the *Anthropology*? Or is it something broader: a consciousness of the contribution of our cognitive capacity?

Kant's initial remarks about being "attentive" to the contribution of our cognitive capacity are certainly underdetermined and preliminary: Kant has not yet laid out the parameters of his investigation. Until we become clearer about the parameters of critical philosophy, we cannot hope to answer the question that I just raised in a satisfactory way. Nevertheless, we do need to say something about it if we want to end up with a preliminary grasp of what reflective metaphysics involves.

The doctrine most closely associated with transcendental idealism is Kant's innovation that space and time are "pure forms" of intuition. Do these forms count as an "addition" of our cognitive capacity — something that we could be "attentive" to through reflection, as a "consciousness of the understanding"? To answer this, we might consider a passage from Phenomena and Noumena where Kant invites us to perform an abstraction on empirical cognition. At first, we lift away "all thinking (through categories)": doing this, Kant asserts, leaves us with "no cognition of any object at all" (A253/B309). Kant then invites us to consider a different abstraction, in which we lift "all intuition" from empirical cognition. In doing so, we take away everything particular and singular from the representation. This, Kant claims, would leave us with just "the form of thinking, i.e., the way of determining an object for the manifold of a possible intuition" (A254/B309). The first abstraction leaves us genuinely empty-handed; the second abstraction, however, leaves us with a mode of access to objects, even if it does not leave us with a representation of any actual object. The understanding is the source

78

of any cognitive relation we may have to objects. The passage suggests that when we take away the contribution of the understanding, we take away everything that pertains to the "possibility" of objects — and this, it seems, would include the forms of sensibility.

The metaphysical aspiration of the *Critique* is to give an account of the "possibility" of objects. The work rests on a reflective appreciation of what it is that "our own cognitive faculty (merely prompted by sensible impressions) provides out of itself" (B1). If my interpretation of the passage from Phenomena and Noumena is correct, this reflective work is principally a "consciousness of the understanding": it is a kind of *logic*, after all — *transcendental logic*. But this "consciousness of the understanding" would have to involve some account of the contribution of sensibility, in so far as that contribution rests on the capacity for thought. As I will show in Chapter 4, the primary concern of the Transcendental Deduction is to account for the respective contributions of sensibility and understanding on the basis of a rational principle. As we will see, the principle employed is a *reflective* principle: it has to do with the self-consciousness of our cognitive capacity. The concept of nature emerges out of the account of the relation between sensibility and understanding that is established on the basis of this principle. For this reason, the *Critique* is "reflective metaphysics".

79

## CHAPTER 3 SPONTANEITY AND SELF-KNOWLEDGE IN THE CRITIQUE

Thus when nature has unwrapped, from under this hard shell, the seed for which she cares most tenderly, namely the propensity and calling to *think* freely, it gradually affects the mentality of the people (who thereby become, little by little, more capable of *freedom* in acting), and eventually even affects the principles of the *government*, which finds it salutary to treat the human being, who is *now more than a machine*, in keeping with his dignity. —"What is Enlightenment?" (8: 41-42)

1.

Kant advertises the *Critique of Pure Reason* as a project of rational selfknowledge, *Selbsterkenntnis* (Axi-xii).<sup>1</sup> And looking back over the "the whole course of our critique" at the end, he proclaims it reason's "scientific and fully illuminating selfknowledge" (A849/B877). But do we know how to understand this? After all, the *Critique* also teaches us that knowledge is limited to objects of possible experience. And reason is not an object of possible experience according to Kant: it "does not at all belong in the series of sensible conditions which make appearances necessary according to natural laws" (A556/B584). However, Kant does not seem to think that this conception of reason rules out the possibility of rational self-knowledge. Notoriously, he writes: "Only the human being who is otherwise acquainted [kennt] with the whole of nature through the senses, knows [erkennt] himself also through mere apperception, and indeed in actions and inner determinations which he cannot at all count among impressions of sense" (A546/B574). Has Kant's pen slipped? Or perhaps these remarks about self-knowledge are merely rhetorical flourishes? For on the face of it, this characterization of the project is incoherent by the *Critique*'s own lights: its operative conception of reason precludes our assuming that reason investigates itself as a phenomenal object, while its arguments limiting knowledge to objects of possible experience preclude our assuming that reason investigates itself as a noumenal object. <sup>2</sup> Kant's division of objects into phenomenal and noumenal is exhaustive (A235/B294). So it seems there is one move left to make if we care to preserve the coherence of Kant's project: reason acquires knowledge of itself not as "object" at all, but as "subject".<sup>3</sup>

By itself, this is a thin response — something to say in order to avoid a clearly unsatisfying pair of alternatives: namely that the *Critique* is somehow engaged in the same sort of rational psychology incriminated by its own pages, or that it is essentially a project of empirical psychology and thereby in tension with its own aspiration to carry

<sup>&</sup>lt;sup>1</sup> See Bxxxv-xxxvi for a parallel passage that does not use the term *Selbsterkenntnis* explicitly.

<sup>&</sup>lt;sup>2</sup> Patricia Kitcher bites the bullet and claims that the *Critique* is concerned with the phenomenal self (see "Kant's Real Self", *passim* and especially 138; and *Kant's Transcendental Psychology, passim* and especially 22). According to Kitcher, the arguments of the *Critique* turn on what she calls "task analysis", or "psychological analysis" (*Kant's Transcendental Psychology*, 18, 84). The starting point is "experience, that is, various tasks that make up cognitive experience"; the analysis reveals that these tasks "require certain elements that cannot be supplied by the senses" (18). "In analyzing cognitive tasks, he looks at the ability, then looks at the known resources, and then argues that we can only explain the ability by assuming some additional resources" (84). If such an analysis is meant to reveal the role of pure apperception as the source of theoretical cognition in general, then it is hard to see how it could really be understood as knowledge of the "phenomenal self". Moreover, Kitcher draws on Kant's remarks at B1 — where he famously invokes the empiricist dictum that all knowledge begins with experience — as grounds for her interpretation of his method. But the idea that the reasoning of critical philosophy begins with experience stands at odds with the idea that its only starting point is "reason itself", as Kant claims in the *Prolegomena* (§4, 274). I discuss the *Critique*'s starting point in Chapter 4.

<sup>&</sup>lt;sup>3</sup> For versions of this response, see: Houston Smit, "The Role of Reflection in Kant's *Critique of Pure Reason*", 205; Henry Allison, *Kant's Transcendental Idealism*, 287, and "On Naturalizing Kant's Transcendental Psychology", 65ff.; and Lewis White Beck, "Towards a Meta-Critique of Pure Reason", 31. With the possible exception of Smit, none of these commentators tries to take the idea much further than this initial suggestion.

out an a priori, apodictically certain, investigation.<sup>4</sup> In order to make this response amount to something more than a move in a game of exegesis, we might consider what it is *trying* to say. Lewis White Beck avoids the dilemma in the following way: it is the "spontaneity of thought [that] cannot be made object; awareness of it is called 'transcendental consciousness', but Kant says regrettably little about it".<sup>5</sup> Knowledge of the "subject", it seems, must be concerned with the activity of knowing, or judging, rather than with what is known or judged, or with the thing that knows and judges. The project of self-knowledge is an investigation of spontaneity itself.

Unfortunately, and unnecessarily, Beck throws up his hands at the prospect of finding much guidance in the *Critique* itself about how to understand the idea that the spontaneity of theoretical reason could investigate itself, and in such a way that the results could count as knowledge. The aim of this chapter is to make it meaningful to say that the *Critique* is an investigation of the self as a "subject" — and to do so by taking Beck's cue, noting that this project of self-knowledge is at bottom an investigation of spontaneity.

Although spontaneity is clearly the central concept of Kant's *Critique of Pure Reason* — and arguably of his critical philosophy at large — Kant says relatively little about what exactly his appeal to spontaneity entails. I will begin by examining two of Kant's clues about it: namely that spontaneity is proper to the intellect, and that it is a kind of self-determination. I first consider what the intellect is according to Kant, and explain his characteristic distinction between reason and understanding. My account

<sup>&</sup>lt;sup>4</sup> On the "apodictic" status of the argument, see (e.g.) Axv-xvi. Empirical psychology, according to Kant, is an investigation of the self as "the *object* of perception, and thus of inner sense" (*Anthropology*, 7:134n.).

<sup>&</sup>lt;sup>5</sup> "Towards a Meta-Critique of Pure Reason", 31. On "transcendental consciousness", Beck cites Kant's *Reflexion*, "Ist es eine Erfahrung, daß wir denken?" (18:319).

gives special attention to the activity of reflection, which plays a special role in the critical project of reason's self-knowledge (\$2). I then turn to the other clue, working out precisely what sort of "self-determination" is thought in the concept of spontaneity (\$3). These considerations distinguish the Kantian conception of spontaneity from anything that could be explained in straightforwardly empiricist terms. For some commentators, this only betrays its illegitimacy. I respond with the suggestion that Kant's appeal to spontaneity is proper to the project of the *Critique* itself: in other words, we cannot adequately understand what Kantian spontaneity is if we do not also come to terms with the idea that the *Critique* is reason's self-knowledge (\$4).<sup>6</sup> I then bring this account to bear on the opening stretch of the *Critique*'s Transcendental Deduction, with the aim of attaining further insight into the critical project of self-knowledge (\$5). A fuller account of the Kantian conception of spontaneity comes in the conclusion (\$6).

## 2.

The concept of spontaneity is officially introduced at the beginning of the *Critique*'s Transcendental Logic. In these initial remarks, Kant attributes spontaneity to the capacity of the mind to "bring forth representations itself".<sup>7</sup> This spontaneous, productive capacity is distinguished from the capacity of the mind "to receive representations, in so far as it is affected in some way" (A51/B75). Spontaneity is attributed to the intellect, the capacity to "think" given representations, and to cognize

<sup>&</sup>lt;sup>6</sup> If this is right, then further debates about the legitimacy of Kant's appeal to spontaneity should then be part and parcel of debates about the viability of the method of the *Critique*. My aim in this chapter is limited to drawing the connection between Kant's appeal to spontaneity and the overarching conception of the project as reason's self-knowledge. In Chapter 4, I offer a closer account of the *Critique*'s synthetic method, which concludes with a brief critical assessment of it.

objects by means of them (A50/B74); this is the "higher cognitive faculty", which is sharply and unequivocally distinguished from sensibility or the "lower cognitive faculty".

Although Kant never offers any explicit justification for drawing this distinction between receptivity and spontaneity, it also appears not to require one. For any account of our theoretical cognitive capacity is likely to draw on the idea that we must be both receptive to the way things appear, and at the same time active in response to those appearances. Although this distinction is quite unsurprising, and apparently uncontroversial, Kant suggests that he alone understands it properly. At any rate, he charges his predecessors with failing to appreciate fully and consistently that there are two *heterogeneous* sources of our cognition — one receptive, the other spontaneous which can never "exchange their functions" (A51/B75).<sup>8</sup> At the same time, this spontaneous capacity — which is the discursive understanding — *essentially cooperates* with the receptive capacity in order to do what it does, which is to judge about given objects. Sensibility and understanding are distinct but necessarily unified — in so far as either can be a cognitive capacity at all.

Let us now focus simply on Kant's conception of spontaneity, which he presents to us through his account of the understanding. We must first take note of the fact that Kant uses the term "understanding" in both a broad and a narrow sense. The understanding in the *broad* sense comprises the three "higher cognitive faculties": the understanding in the *narrow* sense, the power to judge (*Urteilskraft*), and reason

<sup>&</sup>lt;sup>7</sup> Apart from a somewhat tangential remark about spontaneity in the Transcendental Aesthetic (§8, added in the B-edition), this opening passage of the Transcendental Logic (A50-1/B74-5) contains the first mention of "spontaneity" (*Spontaneität, Selbsttätigkeit*) in the *Critique*.

<sup>&</sup>lt;sup>8</sup> Sensibility and understanding are two distinct sources of cognition, with two distinct modes of representation, intuition and concept, which "constitute the elements of all of our cognition" (A50/B4). For

(A130/B169).<sup>9</sup> How do these capacities differ from one another, and how are they unified as a single spontaneous cognitive capacity?

In order to answer this question, I would like to review a certain case of cognitive activity that Kant discusses in the course of laying out the problem of the *Critique* in its Preface. The case is our reasoning with diagrams in Euclidean geometry.<sup>10</sup> This reasoning requires that one distinguish the non-contingent features of the diagram from the contingent ones. The non-contingent features have their source in the discursive part of the demonstration; the contingent features do not. In dealing with a particular drawn figure, Kant remarks that we attend "always only to the action of the construction of the concept" (A714/B742). There are "many determinations" of the particular drawn figure to which we are indifferent, in so far as we attend only to the action of the construction. These are the contingent features. Our ability to distinguish the non-contingent features of the diagram is, at bottom, an ability to distinguish what we *put into* the figure from what we did not. It is an awareness of ourselves: the cognitive practice of reasoning with diagrams in geometry is made possible by some kind of tacit self-consciousness.

Kant holds that rational cognition is marked by this self-conscious activity. Discussing physics in the Preface, Kant notes the tacit insight of its practitioners that "reason has insight only into what it brings forth according to a plan of its own" (Bxiii). The point, however, is general: the great light that dawns upon experimental practices of physics is the same great light that dawns upon the practices of the geometers. It is the

the charge against his predecessors, see the Amphiboly: "In a word: *Leibniz intellectualized* appearances, just as *Locke* [...] *sensualized* the concepts of the understanding" (A271/B327).

<sup>&</sup>lt;sup>9</sup> In this passage, Kant gives this whole a generic title, the "understanding in general" (A131/B169); elsewhere, his generic title for it is "reason" (A835/B863). To avoid confusion, my generic title for this whole is simply the "intellect".

<sup>&</sup>lt;sup>10</sup> I am rehearsing a point from Chapter 1, §4.

light of reason, which is to be understood as some kind of self-consciousness. As we shall see shortly, Kant calls this self-conscious activity "reflection".

To the extent that modern philosophy in general is preoccupied with selfknowledge, it is not surprising that many of its central texts draw on some conception of reflection or consciousness of the mind. For Descartes, the intellect is best aware of itself — best able "to distinguish without difficulty what belongs to itself, to an intellectual nature" — when it withdraws from the senses, and finds itself in its "own freedom".<sup>11</sup> The *Meditations* is inaugurated by a call for such a withdrawal, so that the work of reflection may begin. For Locke, reflection means turning our attention to the "operations" of the mind.<sup>12</sup> There are no such "operations", and so no reflection, without the stimulus of the senses: "And the use of Reason becomes daily more visible, as these Materials, that give it Employment, increase".<sup>13</sup> Reflection can clearly be understood in various ways, and Kant's idea of it is neither Cartesian nor Lockean.

In Kant's taxonomy of terms, there are two kinds of "apperception" or "selfconsciousness": pure and empirical. Pure apperception is *reflection*, and empirical apperception is *apprehension*.<sup>14</sup> Kant reserves the term "reflection" for the selfconsciousness that is internal to the intellect: it is "consciousness of the understanding", whereas apprehension is "consciousness of inner sense". Hence, the idea that the *Critique* is a project of rational self-knowledge, principally concerned with spontaneity,

<sup>&</sup>lt;sup>11</sup> Meditations, Synopsis, AT VII, 12.

<sup>&</sup>lt;sup>12</sup> Essay concerning Human Understanding, II.i.4. Locke does use the word "mind" here; he also uses "understanding" and "reason" without clear distinction.

<sup>&</sup>lt;sup>13</sup> *Essay*, I.ii.15.

<sup>&</sup>lt;sup>14</sup> Anthropology §4 (7:134n.).

implies that it is a project of reflection in *Kant*'s specific sense.<sup>15</sup> Locke's "reflection" is Kant's "apprehension"; and Locke's project, by Kant's lights, is a kind of empirical psychology, which strictly speaking has no place in metaphysics (A848/B876; see also P §21, 304).

Now, Kant claims that the *Critique* "lays down nothing as given [...] except reason itself".<sup>16</sup> The idea that it takes nothing as given except "reason itself" strikes a rationalist note; and yet, the *Critique*'s Paralogisms chapter should prevent us from misconstruing this remark.<sup>17</sup> Reason cannot come into view for itself (as a cognitive capacity, at any rate) through a Cartesian "meditation", i.e., through contemplation in a state of withdrawal from, or disregard of, material nature. Kant heeds an empiricist insight about the possibility of the intellect's self-knowledge — or, at any rate, the anti-rationalist spirit of it: namely, that our engagement with material nature must be held firmly under consideration in order for reflection to be possible, or at any rate, *meaningful*.<sup>18</sup> That Kant is able to acknowledge this empiricist insight without also having to accept the empiricist program will emerge more clearly in the following section.

<sup>&</sup>lt;sup>15</sup> This passage from the *Anthropology* also supports the idea that we can avoid the dilemma about reason's self-knowledge by supposing that reason investigates itself as "subject", and not as "object". Reflection, or "consciousness of the understanding", is concerned with "the I as *subject*", whereas apprehension is concerned with "the I as the *object* of perception" (7:134n.).

<sup>&</sup>lt;sup>16</sup>Prolegomena §4 (4:274); see also Critique, Axiv.

<sup>&</sup>lt;sup>17</sup> The Paralogisms chapter of the *Critique* exposes the fraudulence of "dogmatic rational psychology" — a general heading for rationalist investigations into "spiritual substance".

<sup>&</sup>lt;sup>18</sup> In the *Critique of Practical Reason*, Kant remarks that in the first *Critique*, we "consider reason [...] in relation to objects" (5:16) — in contrast to the second *Critique*, which considers reason in relation to the will. We might take this as a gloss for the suggestion above, namely that Kant accepts an anti-rationalist insight about reflection from the empiricists: i.e., when theoretical reason reflects on its own capacity, it must do so by considering its "relation to objects". Hence (as Kant continues this line of thought), the first *Critique* must begin with sensibility, whereas the second *Critique* will "only", at the end, "where possible" proceed from concepts "to the senses" (16). In other words, the anti-rationalist insight is reflected in the very structure of the *Critique*.

As we have seen, Kant attributes spontaneity to the higher cognitive faculty at large. One aspect of our spontaneity, therefore, has something to do with the reflective activity of reason. But what about the rest of the higher cognitive faculty? Every action of the understanding, Kant remarks, can be referred to judgment: it is a "*faculty for judging* [*Vermögen zu urteilen*]" (A69/B94).<sup>19</sup> It is distinguished from the power to judge (*Urteilskraft*) as a potentiality is distinguished from its actuality; as Béatrice Longuenesse has noted, the power to judge is simply the understanding conceived of as it is "actualized" under sensory stimulation.<sup>20</sup> So, even though Kant marks out three higher cognitive faculties, there are but two fundamental activities at stake: reflection and judging.<sup>21</sup> Given that the three higher cognitive faculties are presented as aspects of a whole, we can suppose that the discursive intellect is a unified capacity. Reflection is its self-concerned activity, and judging its object-concerned activity. The intellect's

We now have in place a rudimentary account of Kantian spontaneity. It is characteristic of the intellect as such, which means that reflection and judging are the two

<sup>&</sup>lt;sup>19</sup> Kant's conception of a faculty is teleological: judging is the *end* of the understanding. Whatever other activity may be attributed to the understanding is fully intelligible only with reference to this end.

<sup>&</sup>lt;sup>20</sup> See Kant and the Capacity to Judge, 7 ff.

<sup>&</sup>lt;sup>21</sup> Kant's terminology is not entirely rigid: as noted above, the general title for the higher cognitive faculty is sometimes "reason" and sometimes "understanding". Note also Kant's remark in the *Prolegomena*, that "the understanding intuits nothing, but only reflects [reflektiert nur]" (288). The point that really matters, which I will argue for here in this chapter, is that reflection and judging are necessary and interdependent aspects of Kant's conception of spontaneity; the context of this remark in the *Prolegomena* suggests that it pertains to a very general distinction between receptivity and spontaneity, and hence does not stand in serious conflict with my suggestion that reflection is the characteristic activity of *reason*.

<sup>&</sup>lt;sup>22</sup> We could draw a similar conclusion about Kant's distinction between reason and understanding from the overarching lesson of the Transcendental Dialectic. The Dialectic demonstrates that the errors of reason stem from its presumption to judge about objects — all on its own, as it were. The proper business of reason, in other words, is reflection. Reason reflects on the judging subject; the understanding judges about objects. Reason is a governing capacity, that without which sensibility and understanding could not be cognitive capacities at all. Kant draws inspiration here from Plato's *Republic*: see (e.g.) A294/B350 and A832/B860.

necessary aspects of it.<sup>23</sup> In order to arrive at a satisfying account of what Kant means by spontaneity, it will be necessary to consider in greater detail the relation between these two activities of the intellect (reflection and judging). We shall take this issue up again in §4. At this point, however, there is still more to learn from Kant's preliminary remarks about spontaneity. I shall turn now to the *unofficial* introduction of the notion of spontaneity in the *Critique*, in which Kant distinguishes his conception of the understanding from anything that could be accounted for in strictly empiricist terms.

3.

Kant opens the *Critique* with an underdetermined conception of the activity of the understanding, which he associates with the empiricists. He begins with a sincere nod to the empiricist thesis that "all of our cognition begins with experience", and continues, with his sincerity intact, saying that our cognitive faculty is "awakened into exercise [Ausübung] [...] by means of objects which affect our senses" (B1). This brings "the activity of our understanding [Verstandestätigkeit] into motion to compare these [sc. sensible representations], to connect or separate them". No Locke, no Hume, should dig his heels in at this and refuse to go on. But is it, as it stands, an appeal to the *spontaneity* of the understanding? The passage goes on to distinguish two sorts of questions about cognition, one about its temporal beginning, and the other about its *source*. Kant claims that the activity of the understanding is "merely occasioned" by "sensible impressions"

<sup>&</sup>lt;sup>23</sup> Spontaneity should not be identified with reflection (or pure apperception) alone, as some commentators suggest. Although it is impossible to address the Kantian conception of spontaneity without talking about pure apperception, or reflection, many commentators' accounts end up dealing with pure apperception almost exclusively. (For an example, see Robert Pippin's "Kant on the Spontaneity of the Mind". He admits, though, that it may look like an attempt to explain the "obscure through the more obscure" (42).) This is a mistake: pure apperception is only part of the story.

("durch sinnliche Eindrücke bloß veranlaßt", B1). This announces that we can take a view of the understanding as independent of the way in which it gets nudged into activity by these occasioning causes. In other words, the understanding is not the creature that it is in virtue of its being impressed upon by material nature. That, I take it, is the force of this rhetoric about "occasion".

With this remark about the activity of the understanding being "merely occasioned" by sensible impressions, we approach Kant's concept of spontaneity. For Kant, the appeal to spontaneity is of a piece with the invocation of the "pure understanding". To suppose that it could make sense to give an account of the understanding independently of the "occasions" of its activity is to take the notion of a "pure understanding" to be intelligible. A pure understanding is one that can be active — can judge — a priori, independently of experience. But it judges about objects that can only be given *in* experience: i.e., it is a capacity for *synthetic* a priori judging.<sup>24</sup>

Thus, the notion of a "pure understanding" contains the thought of its autonomy or self-determination — at least with respect to the idea that the actual occasions of its activity are not essential to its being the capacity that it is. For Kant, spontaneity is a kind of autonomy or self-determination.<sup>25</sup> A passage from Kant's lectures on metaphysics makes this point quite clearly: "The intellectual cognitive faculty rests on spontaneity, or

<sup>&</sup>lt;sup>24</sup> We could gloss Kant's distinction between synthetic and analytic judgments in this way: synthetic judgments concern objects, analytic judgments concern concepts. Thus, "synthetic a priori judging" — at least in the context of the first *Critique* — is judging a priori about objects that can only be given in experience. In the *Critique of Judgment* (§76), Kant glosses "judging objectively" with judging "synthetically" (5:401).

<sup>&</sup>lt;sup>25</sup> I am taking it that *spontaneity*, *self-determination*, and *autonomy* are interchangeable terms for our purposes. We can then distinguish between the spontaneity of the understanding (which is Kant's concern in the *Critique of Pure Reason*) and the spontaneity or freedom of the will (which is Kant's concern in the *Critique of Practical Reason*). In the third *Critique*, Kant is not shy about employing the notion of autonomy outside of the context of moral philosophy, to distinguish the "higher" from the lower "faculties

the faculty of determining oneself, for it is independent of sensation".<sup>26</sup> If the cognitive power is receptive in some respect, then it may be affected by bombardments from the causal order of material nature. But if the cognitive power is also spontaneous in another respect, then it is free from *determination* by that causal order. *This* is Kant's conception of spontaneity, which involves more than the mere idea that the understanding is aroused into activity by objects affecting the senses.

Some commentators have challenged Kant's right to appeal to such a conception of spontaneity — supposing, in effect, that Kant simply stipulates the crucial idea of a "pure understanding".<sup>27</sup> But this is a misreading, which stems from a disregard of the way in which Kant attempts to bring the idea of the pure understanding clearly and legitimately under consideration. As we have already seen, Kant claims that the *Critique* takes nothing as given except reason itself. Now, in the *Critique of Practical Reason*, Kant holds that the moral law is "merely the self-consciousness of a pure practical reason" (5:29); and from this the "concept of a pure will arises" (30). Perhaps, then, the

of the soul": the higher ones "contain an autonomy" (5:196). The autonomy of the faculty of theoretical cognition, Kant goes on to claim, is "contained" in the principles of the pure understanding.

<sup>&</sup>lt;sup>26</sup> Metaphysik Mrongovius (29:881).

<sup>&</sup>lt;sup>27</sup> Relatively few commentators have tried to deal with the meaning and status of Kant's appeal to spontaneity in the *Critique*, despite its obvious importance to the project. Karl Ameriks is alert to the problem of its status, noting that Kant included spontaneity in his conception of rational psychology in his lectures on metaphysics prior to the *Critique*, but then excluded it from the conception of rational psychology examined in the Paralogisms (*Kant and the Mind*, Preface to the Second Edition, §3). Yet he only raises this as a puzzle; and at the end of the day, he does not seem to be especially worried that Kant's appeal to spontaneity could be illegitimate in some way. Susan Hurley certainly is worried — or rather, avowedly skeptical. She writes that the central argument of the *Critique* turns on the idea that the "consciousness of the spontaneity enjoys an unjustified and unjustifiable "protected status within transcendental idealism" ("Kant on Spontaneity and the Myth of the Giving", 155). In other words: without spontaneity, Kant's argument in the Deduction would go out the window; and for this reason Kant fails to address the grounds on which he invokes the concept of spontaneity in the first place. If he did, he would be forced to recognize that his appeal to spontaneity is untenable; and such an admission would undermine the Deduction and everything else in the *Critique* that turns on it (i.e., most everything).

appeal to the pure understanding is akin to that of the pure will: it arises from the selfconsciousness of pure *theoretical* reason.

In this passage from the second *Critique*, Kant goes on to distinguish between "positive" and "negative" conceptions of freedom, in a way that may shed light on the appeal to spontaneity in the first *Critique*. The moral law, Kant claims, is "identical with the positive concept of freedom" (5:29). This implies that the negative concept of freedom points merely to the will's possible *freedom from* being determined sensibly, by inclination. We can point to this freedom from sensible determination without necessarily having in view a positive account of this freedom — which, I suppose, would be expressed in the various formulations of the categorical imperative. Now, at the beginning of the Transcendental Analytic, Kant implies that a *positive* account of the understanding, and presumably its spontaneity, is in the offing. So far, he says, the understanding has been "explained merely negatively, as a non-sensible cognitive faculty" (A67/B92). The "negative" account of the understanding merely opposes it to sensibility, as we find in most of the references to the understanding prior to this point.<sup>28</sup> The remark suggests that the Transcendental Analytic aims to give a *positive* account of the spontaneity of the understanding.

The merely negative conception of spontaneity, which is glossed in terms of the understanding's *freedom from* being determined by sensibility, might seem to be a mere embellishment upon the empiricist conception of the activity of the understanding that Kant invokes at the outset of the *Critique*'s Introduction. If this were so, perhaps we could grasp some aspect of Kant's conception of spontaneity without having to worry

<sup>&</sup>lt;sup>28</sup> The remark is anaphoric, most immediately, for A65/B89: "The pure understanding separates itself completely not only from everything empirical, but indeed also from all sensibility."

about the legitimacy of his appeal to the pure understanding. The embellishing move would look something like this:

The empiricists are right to think that all knowledge begins with experience, and that the understanding must be aroused into activity by the objects which affect the senses. At the same time, however, we can also suppose that the understanding is the source of its own activity; the objects which affect the senses are merely an initial stimulus, to which it responds according to its own principles.

Now, this embellishment of the empiricist conception of the understanding falls short of the idea that these principles of its activity are themselves expressions of its own selfdetermination. It says only that the understanding is active according to its own principles; its "motions", in other words, are not to be conceived as mere extensions of the causal order of material nature. This is supposed to make it be the case that we are talking about *spontaneity*, so that we might distinguish the activity of the understanding from the causal order of material nature.

We might, following this line, liken the activity of the understanding to the activity of a non-rational animal. The beast is not a machine, but a creature with perception; its actions are intelligent, in the sense that they are responses to the salient features of its environment. It acts according to its instincts, which might be thought of as the principles that make it the sort of beast that it is. In responding as it does to what is salient for it, the beast is expressing those principles, and living a life of a certain kind. In *this* sense it is self-determined, and in *this* sense it "acts". But because its principles

are a matter of mere instinct, it is not self-determined, or autonomous, in a deeper sense: it cannot be held responsible for the constitution of its will.<sup>29</sup>

According to some accounts, the idea of a "deeper" sort of spontaneity — at least in the practical arena — has something to do with the idea of *choice*. Practical cognition is concerned with how things ought to be, and not (or not entirely) with how things are. Particular maxims are chosen as principles of conduct, and it is on this basis that we can hold an individual responsible for the constitution of his will. So the concept of choice has an obvious place in an account of practical cognition. But there is no clear role for the concept of choice in an account of theoretical cognition. In theoretical cognition, we are simply registering or reporting on how things are, and the concepts that we may employ in doing so are representations that, in some sense anyway, we simply find ourselves with.

So, if we lean too heavily on an analogy with practical knowledge, we may lose our bearings. For it hardly looks promising to suppose that some notion of choice could lead us to a more satisfying account of the spontaneity of theoretical cognition. At any rate, we should not expect to map the structural elements of one account neatly onto the

<sup>&</sup>lt;sup>29</sup> In her recent John Locke Lectures ("Self-Constitution: Action, Identity, and Integrity"), Christine Korsgaard distinguishes between the action of rational and non-rational creatures in this way. She argues that non-rational animals are autonomous or self-determined when their movements are governed by the principles of their own causality — i.e., their instincts. "An animal's instincts are its will, the laws of its own causality. They determine what it does in response to what, what it does for the sake of what. When it acts from its instincts then, the animal's movements are its own. It acts according to its own laws, and therefore autonomously" (III:21). Animals are autonomous even though they are not self-conscious, or reflective. The autonomous movements of an animal express an intrinsic purpose - e.g., living a mouse's life. The crucial difference between this sort of autonomy and our autonomy, Korsgaard notes, is that its instincts are not chosen: the animal does not choose the principles that are constitutive of its will (III:25). The mouse has a will: it determines for itself when and where and under what specific circumstances to scurry, and for how long, and for how far. But it does not choose the principles that guide those determinations. A rational creature is autonomous in a deeper sense, because it is capable of choosing the principles that are constitutive of its will: "Every agent, even an animal agent, is autonomous and selfdetermined in the first sense, or it would make no sense to attribute its movements to it. Only responsible agents, human agents, are autonomous in the second and deeper sense" (III:26).
other. But perhaps the analogy between the spontaneity of the understanding and the freedom of the will — one which Kant, I think, invites us to entertain — is best appreciated from a higher perspective, somewhere above the distinction between theoretical and practical cognition. To take this perspective is to suppose that spontaneity and freedom essentially belong to reason, and that we can inhabit this perspective as rational creatures. This is the perspective of critical philosophy as such, which is concerned with the "deeper" sort of spontaneity that is expressed in theoretical and practical cognition alike.<sup>30</sup> In critical philosophy, this "deeper" sort of spontaneity is conceived as *reflective* spontaneity. So, if all of this is correct, then Kant could conceivably legitimate his appeal to spontaneity through some kind of demonstration establishing the fundamental role of reflection in living the life of a rational animal. In the first *Critique*, Kant is specifically concerned with the *theoretical* cognitive activity of the (human) rational animal.<sup>31</sup> This activity is paradigmatically expressed as judging. It has already been noted above (§2) that it will be necessary to consider the relation of the reflective activity of reason to the judging activity of the understanding in order to advance our account of spontaneity. Conceivably, then, the full story about the legitimacy of Kant's appeal to spontaneity might turn on some demonstration about the

<sup>&</sup>lt;sup>30</sup> Tyler Burge suggests that the knowledge involved in "understand[ing] the notion of reason" — is above the theoretical/practical distinction ("Reason and the First Person", 250 and 258). Cf. Nicholas Rescher, who suggests that the reasoning involved in carrying out critical philosophy is itself *practical* (*Kant and the Reach of Reason*, 176).

<sup>&</sup>lt;sup>31</sup> For Kant, the "human" restriction only comes into play owing to the role of sensibility in theoretical cognition. Moreover, some aspect of the *Critique*'s account of sensibility is free from this restriction. For it is owing to our rational nature that our sensibility is "formed", or constituted a priori, *in some way or another*. But it is a contingent fact, from the perspective of critical philosophy, that our sensibility is constituted a priori by space and time (as pure forms of intuition) *in particular*. I discuss this further in Chapter 4.

relation of reflection to judging — for example, one that establishes that there could be no judging understanding were there no reflecting reason.<sup>32</sup>

My aim in what follows is simply to show that the proper grasp of spontaneity requires the framework of reason's self-knowledge as it is set out in the *Critique*. We can get the lesser sort of spontaneity into view, perhaps, by embellishing the empiricist idea that the understanding is aroused into activity by the objects which affect the senses. But the robust conception of spontaneity — one which is proper to a *rational* nature — requires a different approach. The positive account of spontaneity that we are promised in the Transcendental Analytic cannot arise from an embellishment of the empiricist conception of the understanding; it is not enough to say that the understanding gets nudged into action, and then responds according to its own principles, whatever they may be. In order to see why, let us consider Wilfrid Sellars's exploration of Kant's concept of spontaneity in his insightful paper, "'… this I or he or it (the thing) which thinks".<sup>33</sup>

4.

Sellars, too, wants to tie Kant's appeal to spontaneity to the critical project of rational self-knowledge. He begins by suggesting that the aim of the *Critique* is in some sense expressed in the Paralogisms itself: it is to uncover "the nature of our thinking being" (A345/B403; quoted by Sellars, §3). He begins, that is, by entertaining the possibility that there may well be a viable investigation into the "nature of our thinking

<sup>&</sup>lt;sup>32</sup> This is the upshot of Kant's response to Hume. For more of a story about this, see Chapter 4.

<sup>&</sup>lt;sup>33</sup> References will be given by paragraph number (§), not page number.

being" that does not have dangerous Cartesian implications.<sup>34</sup> In this way, he operates with the assumption that the *Critique* is a project of self-knowledge.<sup>35</sup>

In broad overview, the trajectory of his argument goes like this: we start off trying to give an account of spontaneity as it figures in the first *Critique*, but we come up with something that is unrecognizable as *our* spontaneity. Realizing that something has gone wrong, we reconceive the idea that the *Critique* is a project of self-knowledge, an attempt to uncover the "nature of our thinking being". We only gain insight into this "nature" of ours when we consider the first *Critique*'s account of spontaneity within a broader framework, which includes the account of pure practical reason in the second *Critique*.

The problem with this story is that it leads Sellars to suppose that the first *Critique* lacks the philosophical resources to provide an adequate account of the spontaneity of our theoretical cognitive capacity. The first *Critique* would not be a free-standing project of rational self-knowledge on Sellars's view, but merely a provisional first part, a story awaiting crucial supplementation.<sup>36</sup> Let us turn now to some of the details of his account.

Sellars's discussion of spontaneity is predicated on an idea about the understanding which has already been considered above — namely, that it must be

<sup>&</sup>lt;sup>34</sup> Sellars makes the point that Kant's critique of dogmatic rational psychology need not imply that Kant supposes that *no* project of rational psychology (or, as I would put it, rational self-knowledge) is possible (§§2-3) — though a viable rational psychology must of course be distinguished from a Cartesian project which purports to give an account of the "mind as a *sort of being*" (§18), or as a "spiritual substance". Ameriks suggests something similar in *Kant's Theory of the Mind*, passim.

<sup>&</sup>lt;sup>35</sup> Sellars's paper undeniably lends itself to various interpretations. Some commentators hail it for introducing "functionalism" into Kant studies. I do not think that Sellars's paper ought to be inducted into such a project, but that is not something for which I am prepared to argue directly in this chapter. At any rate, my reading comes at right angles to that alternative account of it.

<sup>&</sup>lt;sup>36</sup> I do not mean to deny that the second and third *Critiques* can contribute much to our understanding of the first *Critique*'s appeal to spontaneity. But we have to wonder what the first *Critique* could claim to achieve if its reader must first stumble blindly through its pages and look back from the vantage point of the second *Critique* in order for the spontaneity of the understanding to come into focus at all.

awakened into activity by sensory stimuli. Attempting to acknowledge this point, Sellars remarks: "the structure of the first *Critique* highlights what I have called the relative spontaneity of the conceptualizing mind" (§67). It is not entirely clear what he means by the "structure" of the *Critique*, but presumably he means the basic division of the work (or its Doctrine of Elements) into a Transcendental Aesthetic and a Transcendental Logic, the one concerned with sensibility and the other with the intellect. The heterogeneity of sensibility and understanding is one of the *Critique*'s central theses; it is coupled with the thesis that they must cooperate with one another. For Sellars, this means that our spontaneity must be thought of as "relative", because the understanding *essentially* acts in cooperation with sensibility.

Sellars remarks that there are two modes of self-awareness for Kant: a "*passive* awareness of states that are *passive*"; and "an awareness of the self as having a certain 'spontaneity', as not merely responding to the impingement of 'foreign causes'" (§54). As we have already seen, Kant calls this latter mode of self-consciousness "reflection". Sellars suggests that it is through reflection that we gain an appreciation of ourselves both as moral agents, and as capable of making original contributions to art and science: the self-consciousness proper to spontaneity puts us onto ourselves as "philosophizing, [...] cooking up plots for novels or reflecting on what to do" (§52).

Sellars recognizes that the *Critique* is principally concerned with our spontaneity, and hence that reflection would play a special role in its argument. The problem, Sellars suggests, is that the *Critique* only has the resources to account for our spontaneity as "relative": [F]rom the standpoint of the problems discussed in the *Critique*, even if we take into account an awareness of self that is not that of "inner sense", the spontaneity of which we are conscious is, though not *sheer* passivity, nevertheless *a* passivity in [the sense] that the inner development is set in motion by a foreign cause and follows a routine. (§59)

This model of relative spontaneity, with its talk of routine movements, is not one which resounds with self-recognition. For surely the activities of "cooking up plots for novels" and the like had better come to more than the following of routines. As a putative self-portrait, the account is disturbing; at best, it might accommodate the spontaneity involved in "constituting phenomenal objects" (§57), but it falls short of accounting for the spontaneity involved in the more robustly creative pursuits. As a response to this worry, Sellars suggests that we have to look outside of the framework of the first *Critique* to account for these more robust expressions of our spontaneity. This gets recast by Sellars as a new worry: alone, the *Critique* manages to present a picture of ourselves as mere *thinkers*, but not as *persons*. In listing "it (the thing)" among the "I or he" of the things which think in the famous passage from the Paralogisms, Kant leaves open the possibility of conceiving of a thinker that is not also a person: something which, in Kant's words, "is not capable of any imputation".<sup>37</sup>

What is haunting Kant [...] is the concept of an *automaton spirituale*, a mind which conceptualizes, but only in response to challenges from without, and in ways which, however varied, realize set dispositions. He thinks of the perception of external objects as the paradigm of conceptual activity being called into play.  $(\$65)^{38}$ 

<sup>&</sup>lt;sup>37</sup> Here Sellars quotes from the passage from the *Metaphysics of Morals* where Kant distinguishes "person" from "thing" (6:223). The Paralogisms passage (from which Sellars draws his title) is at A346/B404.

<sup>&</sup>lt;sup>38</sup> On Kant's worries about the *automaton spirituale*, Sellars surely has in mind a particular passage about a "*comparative* concept of freedom" from the *Critique of Practical Reason* (5: 96-7). Kant mentions Leibniz in this passage, though it is not clear what particular text (if any) he has in mind. Most likely he is simply thinking about Leibniz's conception of the apperceiving monad — and not any particular passage about it.

Sellars associates the spontaneity involved in the "perception of external objects" with the troublesome relative spontaneity. He distinguishes it from more a robust (presumably 'absolute') mode of spontaneity — e.g., that involved in the pursuits of metaphysical truth, beauty, and the good ("philosophizing, cooking up plots for novels, and reflecting on what to do"). But can we draw the distinction that Sellars wants to draw between robust spontaneity and its poor cousin — and call them both "ours"?

Sellars's argument turns on the idea of self-recognition: we do not recognize ourselves in the *automaton spirituale*, and therefore we must reject it as a model of our spontaneity. To clarify: we do not recognize ourselves as philosophers, artists, and moral agents. Do we then recognize ourselves as *mere* "synthesizers of phenomenal objects"? For if Sellars wants to say that the first *Critique* is even a partial project of selfknowledge, and if Sellars supposes that the *Critique* can only bring into view a "relatively" spontaneous cognitive capacity like that of the *automaton spirituale*, then he must also be supposing that the mere synthesizer of phenomenal objects could be the "self" that is the subject of that inquiry. However, this is clearly unsatisfactory — for it is next to impossible to know what it could mean to think of ourselves as *mere* synthesizers of phenomenal objects, like the automaton of Sellars's account. But let us suppose that the "self" who signs up for the first Critique's program of self-examination is abstracted from the sociable, cultural, and even scientific, selves that we know ourselves to be; let us suppose that only the mere synthesizer of phenomenal objects is eligible. Surely Sellars's intuition is correct — that a project of self-knowledge must be responsible to the evidence of self-recognition. So, then: how would a mere synthesizer of phenomenal objects, Sellars's automaton, recognize itself?

We know that the automaton must be aroused into activity by "challenges from without" in order to be active at all. Suppose that the principles it employs in its response to outside challenges require that it register those responses. Would this mean that the automaton was genuinely aware of its own activities? In order to say 'yes' to this question, we must suppose that it can distinguish its own activity from what is external to it. If its activity consists of responses to these challenges, then its capacity could be described in terms of a set of rules ("respond to input *x* with activity *y*"). Its activity would be manifest to it through the application of the rules in question. Now, if it were capable of some such self-awareness, would this self-aware capacity genuinely count as *part* of the automaton? In order to say 'yes' to this second question, we must at least suppose that having such a capacity for self-awareness contributes to its end of responding to external challenges.

We are putting Sellars's automaton to the test in this way for a reason: the aim is to see what is entailed about the spontaneity of a creature that could intelligibly carry out its own self-investigation.<sup>39</sup> So we have signed the automaton up for some project of self-investigation, and we have accepted Sellars's tacit suggestion that such a project must be responsible to the evidence of self-recognition. The Sellarsian automaton is limited to a certain repertoire of routines. These routines constitute its capacity to "respond to external challenges" — i.e., to "synthesize phenomenal objects". If the self-aware capacity contributes to the end of responding to external challenges, then it could conceivably interrupt, and change the course of, the activity of responding to external

<sup>&</sup>lt;sup>39</sup>Incidentally, the proposal that I have put to Sellars's automaton is far weaker than the proposal that reason puts to itself in setting down the "critical path". As Kant presents it, the *Critique* fulfills a self-incurring *requirement* of reason. For more on this theme, see the *Critique*'s Prefaces and Doctrine of Method, and also Kant's famous remark about the "thorny paths" of the *Critique* in the *Prolegomena* (4:367).

challenges. But how would the self-aware capacity know when to interrupt, and when to change the course of the operation? It might not have any principles which "tell it" when to do so. In that case, its interruption would be like an arbitrary wild-card mechanism built into the machine; and if so, it would be hard to see that it is genuinely aware of anything at all. Suppose, then, that it does have some principles which "tell it" to interrupt the operation. Then it interrupts on the basis of some internal "demand".

Let us call the two parts of this machine — a machine that manifests its selfawareness in its capacity to guide and correct its own activity — "alpha" and "beta". Alpha responds to external challenges and beta monitors the activity of alpha. The capacity of alpha can be described in terms of principles that have the following form: "respond to input *x* with activity *y*". Beta has its own principles, which allow it to differentiate and guide the various activities of alpha. Its principles might look like this: "if you did *y*, follow with *z*". In virtue of these principles, beta can presumably register when something has gone awry with alpha, and correct the course of its activity. In this scenario, the activities of alpha could be carried out without the direct involvement of beta — at least until they go off the rails.

The capacities of alpha and beta are, as it were, merely linked and not unified. Beta is aware of the activity of *alpha*, but is it for this reason *self*-aware? This sort of linkage could be extended to include further capacities, like "gamma", that would monitor the conjoined activities of alpha and beta. Each additional monitoring capacity, however, would not bring the automaton any closer to becoming self-aware.

If we want to understand the automaton as a unified, self-aware capacity to respond to external challenges, then we would need to say that its self-aware activity is *internal to* its activity of responding to external challenges. This is what is thought in the Kantian conception of reflection. This sort of self-awareness is not observational: it is not a response to activities that are there before the awareness of them can be possible. Rather, the "operations" of the unified capacity are reflective through and through. The principles of the reflective capacity would not be an additional body of instructions added to the original capacity to respond to external challenges. Rather, its self-awareness would always already be manifest in its responses to external challenges, regardless of whether the process has broken down in some way or not. A single set of principles would express this unified capacity, and its activities of self-awareness and responding to external challenges would best be thought of as aspects of *one* activity.

In other words, the self-aware 'automaton' would be capable of taking up the activity of responding to external challenges as its own purpose. It would tacitly acknowledge this purpose in the very exercise of its capacity to respond to external challenges. It is for this reason that Sellars tries to fix the original automaton by attributing an "intrinsic purpose" to it. He thinks that the first *Critique* proper just gives us an account of an *automaton spirituale* as a "conceptualizing machine". In order to see it as anything more, we must find the "purpose which is, so to speak, *intrinsic* to the machine" (§71). This purpose is "intrinsic" because it would make it intelligible to think of its activity as dealing in genuine concepts at all. Here Sellars reaches outside of the context of the first *Critique*, finding this intrinsic purpose in the moral law (§76). The attribution of it to the "relatively spontaneous" cognitive capacity that he takes to be at issue in the first *Critique* is supposed to complete its account. For Sellars, this "intrinsic

103

purpose" — the moral law — is supposed to be the lens through which one must look to find the first *Critique*'s account of the *understanding*.

Yet if what I have argued is correct, then some such intrinsic purpose — one that accounts for the reflective employment of principles — must already inform the supposedly minimally spontaneous activity of "constituting phenomenal objects". Hence we cannot join Sellars in distinguishing relative and absolute spontaneity, at least as long as Sellars wants to call them both "ours".

### 5.

This meditation on the Sellarsian automaton allows us to appreciate why it might matter that the *Critique* is reason's self-knowledge. Our cognitive capacity, which is not happily described as a capacity to "respond to external challenges", is reflective in its every activity. Its activity is the expression of a rational nature, and the reflective activity of reason is not an afterthought or a further embellishment of its expressly 'outward'-directed activities. If the *Critique* is pure theoretical reason's self-knowledge, then reason is carrying out the critical work, and hence reason is manifest throughout. The entire book is reason's self-examination, and not merely the thematic treatment of reason in the second half of the book, the Transcendental Dialectic. In fact, the Dialectic is mostly a negative account of reason's true capacity, or *Unvermögen*, that is presented in the first half of the book. According to Kant, reason cannot meaningfully investigate its capacity if it considers itself in isolation from the understanding; and since the understanding is a capacity to judge about objects that can only be given in

experience, reason cannot entirely isolate its capacity from the receptivity of sensibility either.<sup>40</sup>

Kant supposes that reason could only know itself in so far as its reflective activity is the basis of some 'outward'-directed activity — e.g., theoretical cognition. Theoretical cognition has two basic moments: experience and judgment. In Kant's account, both necessarily involve the representations characteristic of sensibility and of the understanding together. Moreover, both are underwritten by the reflective activity of reason. The Transcendental Deduction establishes this. We shall take just a few steps into this notoriously challenging part of the book — just enough to leave us with some idea of how Kant shows that our cognitive capacity is reflective through and through. In the conclusion (§6), I will connect this thesis of the Deduction with the Sellarsian idea that robustly spontaneous activity requires an "intrinsic purpose".

The anti-rationalist insight about the possibility of meaningful reflection is borne out by the fact that the first words of the Transcendental Deduction concern sensibility. "The manifold of representations can be given in an intuition which is merely sensible, i.e., is nothing but receptivity" (B129). Kant begins the Deduction with a remark about sensibility, and continues, stressing once again the heterogeneity of sensibility and understanding. Combination, he says, is an act of spontaneity, and "can never come into us through the senses, and thus also cannot already be contained in the pure form of sensible intuition" (B129-130). In saying that combination cannot be "contained in the pure form of sensible intuition", Kant warns us against supposing that an analysis of the

 $<sup>^{40}</sup>$  On the impossibility of isolating the account of reason from the account of the understanding, see A305/B362.

intuitive character of sensible representation could yield any insight into the spontaneity of the understanding.

The problem of the Deduction stems from this heterogeneity: we face the possibility that the unity of what is given in sensible intuition is incommensurable with the logical unity of judgment that is expressed in the categories. The solution involves a rational principle, which establishes the necessary cooperation of sensibility and understanding in the production of knowledge. It is the fundamental principle of reason's reflective nature. This principle is revealed in the Transcendental Deduction; for this reason, we can think of the Deduction as the central text of pure theoretical reason's critical self-knowledge. I will give an abbreviated account of how this principle is revealed in the first bit of the Deduction, and how it directly implicates the spontaneous activity of the pure understanding.<sup>41</sup>

The principle in question is stated in a rough-and-ready way in the famous opening lines of \$16 of the Deduction:

The *I think* must be *able* to accompany all of my representations. For otherwise something would be represented in me which could not be thought at all, which is as much as to say that the representation would either be impossible, or at least nothing for me. (B131-2)

Now, it is surely not the case that the capacity expressed by the 'I think' is a condition of the possibility of representing as such. Non-rational animals have a faculty of perception, and so have representations in Kant's sense, but they cannot accompany these representations with the 'I think'. The 'I' who utters this is pure theoretical reason, the agent and subject of the critical self-investigation. Kant invites us to consider the

<sup>&</sup>lt;sup>41</sup> I will cover some of the same ground again in Chapter 4.

alternative proposition — that the 'I think' were not able to accompany all of "my" representations.<sup>42</sup> On this proposal, the subject in question could have sensible representations that could not be thought by it. These representations would "not be anything" for this intellect. At the very least, this means that they could not arouse it into activity, or provide an occasion for its thinking.

The introduction of this principle presents us with the idea that our capacity to enjoy given representations is made possible by the reflective capacity that is expressed by this 'I think'. This reflective capacity is reason. The next move is to recognize that given, or sensible, representations could not "belong" to this reflective capacity without the activity of the understanding. Kant dubs the principle under consideration the "necessary unity of apperception"; he says that it is an "analytic proposition", but one that nevertheless "declares a synthesis of the manifold given in intuition as necessary, without which this thoroughgoing identity of self-consciousness cannot be thought" (B135).<sup>43</sup> So, the "belonging" of sensible representations to this reflective intellect rests on some "synthesis". Again, the alternative is ruled out on the evidence of selfrecognition. An intellect that could enjoy given representations through its reflective activity alone would be the scarcely intelligible "intuiting intellect". Since "ours can only think and seek intuition in the senses" (B135), this alternative proposal is rejected. In so far as there is *something* that I think, this reflective activity is manifest in the synthesis of given representations. Sensible representations may belong or matter to pure theoretical reason owing to this "necessary synthesis".

<sup>&</sup>lt;sup>42</sup> This is implicit in the "for otherwise" in the quoted passage.

<sup>&</sup>lt;sup>43</sup> It is an "analytic" proposition because it is revealed analytically, upon consideration of the very idea of the project as reason's self-investigation; I argue for this view in Chapter 4.

This brief journey into the Deduction shows reason in its role as the agent of the investigation; this is indicated, at least in part, through the use of the first person. We would understand fully how reason is the subject of this investigation if we could continue our journey further, and see how its reflective activity is established as the basis of theoretical cognition in general; to do so, we would need to understand better what this "synthesis" is, and to appreciate its relation to experience and judging alike. Even without the full story, this glimpse of the argument leaves us with the idea that there is no moment of our cognitive activity that escapes the reflective capacity of reason. This is a crucial piece of Kant's conception of spontaneity. But in the complete story about our robust spontaneity, this idea is married to what was insightful about Sellars's move to look for an "intrinsic purpose" of our cognitive activity — without making his mistake of looking for this purpose in the moral law. I turn to these issues now, by way of conclusion.

## 6.

It is not just any sort of "intrinsic purpose" that underwrites the sort of spontaneity that is at stake in the *Critique*. Even the activity of a non-rational animal expresses an "intrinsic purpose", which is simply living a life of a certain kind. The kind of intrinsic purpose that is at issue belongs to reason. Sellars looks to reason in its practical capacity to find this intrinsic purpose. His move leaves reason looking incapable of carrying out the examination of its own theoretical capacity, which is precisely what the first *Critique* is supposed to achieve.

What exactly is this "intrinsic purpose" supposed to do for Sellars's automaton? It is supposed to make the automaton into a viable model of *our* spontaneity. The automaton's dispositions are "fixed" and its responses are "routine". Yet it is not obviously the case that having an intrinsic purpose would make its dispositions any less fixed, or its responses any less routine. After all, animals have an "intrinsic purpose" living a life of a certain kind — and yet the principles of their activity (their instincts) are quite fixed indeed. So perhaps the real worry with the automaton lies somewhere else altogether.

The automaton responds to an onslaught. The intrinsic purpose is supposed to fix this: it is supposed to account for a *passive* expression of its spontaneity, so that it could determine what figures as a proper challenge, or a worthwhile matter for its concern. But again, Sellars supposes that the account of spontaneity in the first *Critique* "abstracts" from such "considerations of purpose" which would make this direction of attention, and appreciation of relevance, intelligible.<sup>44</sup>

Although Sellars is wrong to point to the moral law as this intrinsic purpose, there is nevertheless something like a practical dimension to Kant's account of the spontaneity of the understanding. In the *Critique*, this becomes clearer once we come down from the rarefied plane on which the Transcendental Deduction is carried out, and are newly prepared to consider the judging activity of the understanding somewhere closer to ground level. The understanding, Kant remarks, can be thought of as the "faculty of rules", while the power of judgment is the "faculty of *subsuming* under rules" (A132/B171). According to Kant, the power of judgment needs to be developed.

<sup>&</sup>lt;sup>44</sup> "[T]he first *Critique* simple [i.e., on its own] *abstracts* from the purposive aspects of the conceptualization involved in experiential knowledge" (§66).

"Sound" judgment, he says, is a "peculiar talent, which wants only to be practiced, and not taught":

[...T]his is also what is specific to the so-called mother wit, the lack of which no school can make good. Even if such a school were to fill a limited understanding up with rules that were borrowed from foreign insight, to present them and graft them onto it (as it were), nevertheless the faculty of making use of them correctly must belong to the student himself; and no rule which one may prescribe to him for this purpose is, in the absence of such a natural talent [Naturgabe], secure from misuse. (A133/B172)

This passage is often misread as a throwing up of hands, as if Kant were saying: there is no real hope for those who lack this natural talent, because the skill in question is essentially inarticulate, or blind.<sup>45</sup> That is not Kant's point. Rather, Kant intends to liken sound judgment to virtue: a disposition acquired through practice, under the guidance of reason. The virtue in question is the ability to recognize the relevance of a rule. Considerations of relevance go hand in hand with considerations of purpose. Presuming that the foregoing considerations were on track, we can say that the spontaneity of

<sup>&</sup>lt;sup>45</sup> David Bell associates spontaneity with the "blind" following of rules ("The Art of Judgment", 226). This is surely my only complaint about his wonderful paper. Comprehending Kant's idea of spontaneity requires that we make sense of the apparently paradoxical idea of rule-bound spontaneity, Bell argues. Bell then names the following "principle of spontaneity", to obviate a certain misunderstanding of this apparently paradoxical idea: namely, that it must be possible for me to apply a rule (a concept) for the first time, without having "already performed an act of that type". This is to avoid a regress "of an infinity of acts of judgment [... so that] at some point we must judge immediately, spontaneously — and this means without having already judged, identified, understood, or grasped a thought on the basis of any prior such act" (226). But then Bell goes on to say, wanting to draw on Wittgenstein and Kant alike, that the spontaneity of judgment is *blind*; he interprets this to mean that at the core of our cognitive capacities lies an "ability to enjoy a spontaneous, criterialess, disinterested, presumptively universal, non-cognitive, reflective feeling that certain diverse elements of experience as such belong together" (239) — a capacity revealed in (Bell's reading of) Kant's account of the judgment of taste. I suppose, though, that Bell would think that there is hope for those lacking 'mother wit', after all — namely through the cultivation of taste, as this is the truest practice for judgment (by the lights of the third *Critique*). Yet this only refreshes my doubt that spontaneity, in any respect, could indeed be "blind".

judgment rests on an "intrinsic purpose" — a purpose without which the activity in question could not count as judging at all.

The spontaneous activity of our cognitive power rests on an intrinsic purpose that is articulated in terms of the principles of the pure understanding. This intrinsic purpose is that in virtue of which we are able to appreciate the relevance of a rule to a particular cognitive context. The purpose is expressed as an idea of a *whole* — of which, as such, we can have no experience. This idea of the whole is articulated as the "totality of rules under which all appearances must stand if they are to be thought as connected in an experience" — nature in its "formal" meaning, according to the Prolegomena (§36, 318; see also *Critique* §26, B164). At this level, the story is forbiddingly abstract. But the point is that any theoretical cognitive project is ultimately directed towards the same end — the knowledge of nature — even if it is immediately directed towards a more "local" end, like determining the aetiology of glioblastoma multiforme. As such, it is informed by some rational idea of the whole of nature, which is expressed in the principles of the pure understanding. It is these principles which, fundamentally at least, guide our attention to what is relevant to that end: in other words, it is these principles which direct our attention to *objects*. Just as the moral law expresses the idea of the whole of a kingdom of ends, and determines what it is to be a person, or a member of this kingdom, so the "intrinsic purpose" of theoretical judging would express the idea of the whole of nature, and would determine what it is to figure as an object in it.

The idea that the spontaneous expression of our theoretical cognitive capacity has this intrinsic purpose allows us to appreciate how experience and judging are alike expressions of our spontaneity. An empirical concept is a rule for the determination of appearances. In judging, we employ such rules on the basis of what is relevant to our cognitive end — broadly speaking, the knowledge of nature. That end informs the practice of judging by means of the categories, which are the "forms" of empirical concepts.<sup>46</sup> The act of judging expresses one's appreciation of the relevance of certain rules (i.e., empirical concepts), where the judgment itself is supposed to contribute to the overall end of the cognitive project at hand. In its ideal state, an act of judging announces — by the force of its assertion, as it were — that "*these* considerations are relevant". The spontaneity involved in experience is simply expressed in the direction of one's attention. Judging makes this attention explicit — and maintains its relevance for the broader end that we share as "theoretically" rational creatures.<sup>47</sup>

This appreciation of the relevance of rules is the mark of freedom in our theoretical cognitive activity. Our empirical concepts are reflected upon in the very employment of them. When we employ a concept, we make a claim for its relevance to a given situation, and for some cognitive end. This recognition of relevance is the source of the development of our conceptual capacities — why we, in contrast to the *automaton spirituale*, do not merely exercise fixed dispositions and follow routines if our judging is genuinely an expression of our spontaneity, and (in turn) objectively valid. For new concepts do not materialize out of thin air, but when we appreciate the relevance of an

<sup>&</sup>lt;sup>46</sup> See Chapter 2, §3.

<sup>&</sup>lt;sup>47</sup> In practice, however, the relevant cognitive ends are the more "local" ones; but these local ends themselves point towards the ultimate end of attaining comprehensive and systematic knowledge of material nature (for more, see A835/B863). The "public" character of Kant's conception of reason has received some attention of late; see Katerina Deligiorgi, "Universalisability, Publicity, and Communication: Kant's Conception of Reason", and Onora O'Neill, "Kant's Conception of Public Reason".

existing rule in a new way — by *extending* the sphere of its application, often metaphorically.<sup>48</sup>

Sellars's gesturing towards the sort of spontaneity involved in "philosophizing and cooking up plots for novels" was meant, I take it, to introduce a conception of spontaneity that may have a firmer grip on ordinary folk than it does on philosophers. If asked, the proverbial man on the street would not invoke Kantian ideas about rational self-determination to gloss the concept of spontaneity. He might begin by saying that someone acts spontaneously when she does something surprising, and unprompted. But if such an action is to be seen as something more than random, or arbitrary, then we would have to see it as being prompted by shared considerations — or at least considerations that *ought* to be shared. To act spontaneously, and non-arbitrarily, is to express one's recognition of something that is *worthy* of our attention.

So where does this account of spontaneity leave us with respect to our initial worries about the viability of reason's self-knowledge in the *Critique*? Perhaps we are now in a position to appreciate a clue about the nature of critical philosophy that Kant conveys early on in the *Critique*'s Introduction. As we have already seen, Kant begins there with a sincere acknowledgement of the empiricist thesis that our understanding must be aroused into activity by objects affecting the senses. He goes on to propose that empirical cognition might be "something composed of what we receive through impressions and of what our own cognitive faculty provides from itself" (B1). As critical

<sup>&</sup>lt;sup>48</sup> This must of course be complemented by further specification and narrowing of scope of our concepts — both belong to the full account of the refinement and development of our conceptual capacities. Kant alludes to this in the Appendix to the Transcendental Dialectic at A650/B678 ff. The picture I have sketched in these concluding remarks is meant to anticipate Kant's further account of the proper role of reason in the empirical employment of the understanding, in the Introduction and the Appendix to the Transcendental Dialectic — and even onwards, I suppose, into the *Critique of Judgment*.

philosophers, he hints, we will be expressly concerned with what he refers to as the "addition" that our cognitive faculty provides from itself. But we cannot "distinguish" it, he warns, "until long practice has made us attentive to it, and skilled in the separation of it" (B2).

The "long practice" that prepares us for Kant's critical path is nothing other than the practice of judgment.<sup>49</sup> To the extent that this practice expresses the "spontaneity" (rather than the "mechanism") of reason, we may hope to acquire the virtue of sound judgment through it.<sup>50</sup> We should be able to say something along similar lines with regard to Kant's practical philosophy. On several occasions, Kant emphasizes that we have no need for a philosophy of practical reason in order to recognize the claims of morality; and yet we certainly do need to be able to recognize the claims of morality in order to carry out a critical philosophy of practical reason. So it seems that we are prepared for critical philosophy to the extent that we already know ourselves — *tacitly*, through the perfection of our cognitive practices. Critical philosophy is simply the scientific expression of this self-knowledge.

<sup>&</sup>lt;sup>49</sup> Kant explicitly refers to a certain "ripeness", or maturity, of our power of judgment as the proper preparation for critical self-knowledge (Axi).

<sup>&</sup>lt;sup>50</sup> On the distinction between the "spontaneity" and "mechanism" of reason, see *Logic* (9:76).

# **CHAPTER 4** THE ROLE OF THE TRANSCENDENTAL DEDUCTION IN THE *CRITIQUE*'S SYNTHETIC METHOD

### I.

1.

In the *Prolegomena*, Kant remarks that the method of the *Critique of Pure Reason* is "synthetic" (§4, 274-5). The *Prolegomena* is offered as a popular presentation of the results of the *Critique* that does not submit itself to the same methodological restrictions. Its method is "analytic", and this is supposed to "lighten the task considerably" (279). One conspicuous difference between the two works is the absence of any "transcendental deduction" in the analytic *Prolegomena*. Might it be the case, then, that the need for a transcendental deduction in the *Critique* is closely related to the peculiarities of its method? In this chapter, I will argue that this is so, demonstrating the pivotal role of the Transcendental Deduction in the light of an account of the *Critique*'s synthetic method.

We cannot appreciate the function of the Transcendental Deduction apart from the aim of the *Critique* as a whole. That aim is to address the possibility of a viable metaphysics of nature.<sup>1</sup> True to the spirit of much of the philosophy of the

<sup>&</sup>lt;sup>1</sup> Since the *Prolegomena* is a presentation of the results of the *Critique*, it shares this aim. In the *Critique*, this problem is generalized in terms of the characteristic question about the "possibility" of synthetic a priori *judging* (B19). On this, the *Prolegomena* differs only in that it asks how synthetic a priori *knowledge* is possible (§5, 278). I take it that this small difference in formulation can be chalked up once again to the difference in method: the *Critique* is a "reflective" project, explicitly concerned with the nature of our cognitive faculties, thus it phrases the guiding question terms of a particular cognitive activity. The

Enlightenment, the *Critique* aims to achieve this end through a reflective project of selfknowledge. Now, the concept of nature does not appear in the main argument of the *Critique* at all until the very end of the Transcendental Deduction (§26): this "deduction" apparently allows for a conclusion about the possibility of nature on the basis of its account of the pure understanding.

The Transcendental Deduction is part of the first book of the Transcendental Analytic, the overall aim of which is to provide an account of the pure understanding as a capacity to judge a priori about objects that can only be given in experience. The Transcendental Analytic begins by announcing that the "pure understanding separates itself not only from everything empirical, but also completely [separates itself] from all sensibility" (A65/B89; see also A62/B87). On this ground, appeal is made to pure general logic — a science that deals with the intellect, but abstracts entirely from the content of concepts, and thus from any relation of the understanding to sensibility. It deals only with "the mere form of thought" (A54/B78). From the table of the "functions of judgment" borrowed from pure general logic, Kant presents a table of "categories". But their status as genuine concepts, applicable to objects a priori, hangs in the balance. At this point, they figure as mere "forms of thought".<sup>2</sup>

The Transcendental Deduction is supposed to show that these posited concepts are valid of "whatever objects *may come before our senses*" (§26, B159). It is the "explanation of the way that concepts can relate to objects a priori" (A85/B117). It is an argument of legitimation, showing that the "forms of thought" that are first revealed

*Prolegomena* takes as given certain rational sciences, which are clearly supposed to involve synthetic a priori claims. Thus, it phrases the question in terms of how such *knowledge* is possible — since it begins with it as a kind of *fait accompli*.

<sup>&</sup>lt;sup>2</sup> They remain "mere forms of thought" until rather late in the Deduction; see §24 (B150).

when the understanding is considered in isolation from sensibility are applicable a priori to appearances in general. When this is shown, we entitle ourselves to the idea that these forms of thought are indeed pure *concepts*, or "categories". From there, we claim that the categories "prescribe laws a priori to nature, as the totality of all appearances" (B163).

It is striking that — in crude overview at least — the parallel text in the Prolegomena argues in the opposite direction: it argues from some concept of nature to the categories as the principles of its possibility. The trajectory of this argument has everything to do with the *Prolegomena*'s analytic method. In the *Prolegomena*, Kant accounts for the difference between the two methods in terms of what each admits as given starting points. The *Prolegomena* takes as given certain "actual" bodies of synthetic a priori knowledge; through the analysis of each, it "ascends to the sources" to the "principle of the possibility" — of what is given (275). Pure natural science is one such starting point. Thus, the second main section of the *Prolegomena* begins by identifying the conception of nature that is implicit in the idea of a pure natural science, (§§14-17): we begin with a "material" conception of nature as the "totality of all objects of experience" (§16, 295), from which we derive a "formal" conception of nature as the "necessary conformity to law of all objects of experience" (§17, 296). Kant then argues that a certain sort of empirical judgment is proper to the sort of project that a science of nature is supposed to be (§§18-20). From there, Kant concludes that the understanding must be constituted a priori by certain concepts that are to be thought of as conditions of the possibility of experience (§21); the understanding *must* be conceived in this way, if the judgments of natural science are to be what they purport to be.

Schematically, the argument looks like this: p is not possible without q; p is given; therefore q. Our given, p, stands for something like "pure natural science is a viable cognitive endeavor that is everything that it purports to be". The argument establishes the necessity of q through an iterated analysis of p.<sup>3</sup> Our starting point is the conception of nature that is analytically implied by the idea of pure natural science. This starting point ensures that the further analysis that leads us to the claim that the understanding must be constituted a priori by certain pure concepts also leads to the recognition that those pure concepts would necessarily be valid of appearances in general. The argument of the *Prolegomena* does not require a separate deduction of the categories. Their legitimacy is established by default.

That there is no separate question about the legitimacy of the categories in the analytic *Prolegomena* certainly seems to be part of what lightens the philosophical load. In the *Critique*, the categories come into view for us through pure general logic, and for this very reason their applicability to given objects remains at stake until the end of the Deduction. Has this given us a glimpse into the "synthetic" method of the *Critique*?

<sup>&</sup>lt;sup>3</sup> Interpreters who suppose that the Transcendental Deduction is an anti-skeptical "transcendental argument" take its form to be analytic in this way. Of course, the issue of whether the Deduction is a "transcendental argument" turns on more than mere analyticity; it also has to do with what is admitted as a starting point for the analysis — whether it is some bare conception of self-consciousness, or a full-blooded conception of experience or empirical knowledge. The "transcendental argument" view of the Deduction (e.g., as found in Strawson and his followers) supposes that the Deduction refutes Cartesian skepticism by admitting as given only the minimal premise that one is self-conscious. Stephen Engstrom clearly shows that the Deduction is not offered as a refutation of Cartesian skepticism, and hence is not an example of a "transcendental argument" in this sense. At stake in the Deduction is Humean skepticism; still, the Deduction is not exactly even a *refutation* of Humean skepticism. Noting that Humean skepticism is simply purified empiricism, Engstrom suggests that the Deduction "removes skepticism" of the Humean sort by "doing away with its cause, empiricism" ("The Transcendental Deduction and Skepticism", 375). A somewhat similar point is made by Karl Ameriks, who also wants to argue against the idea that the Deduction is a "transcendental argument" in Strawson's terms; the Deduction, he argues, takes as given some conception of empirical knowledge, moves to the "universal validity of the categories" without being "trivial, question begging, or tied to the scientific presuppositions of Kant's day" ("The Transcendental Deduction as a Regressive Argument", 287). For still another swipe at the "transcendental argument" reading of the Deduction, see Manfred Baum ("Transcendental Proofs in the Critique of Pure Reason", 6 ff.).

Yes: for the fact that we begin by separating the pure understanding from everything empirical and from all sensibility means that we will need to "unite" the understanding with sensibility once again, at least as long as we are seeking an account of it as a capacity to judge a priori about objects that can only be given in experience.

But why this arduous method? Would we be so impoverished if we had only the analytic *Prolegomena* on our shelves, and not the synthetic *Critique*? The simplest answer that Kant would give turns on the idea that reason must come to complete self-knowledge, and that it is only the synthetic *Critique* that can carry this out. In the *Prolegomena*, Kant expresses the hope that its gentler pastures will kindle a longing in the reader to head down the critical path. "[...] I can already imagine beforehand that everyone whom I have led down the thorny paths of the *Critique*, and whom I have made weary and indignant, will ask me on what indeed I ground this hope. I answer: *upon the unavoidable law of necessity*" (*P*, 367). Kant says that his hope is founded upon some sort of necessity, which presumably has to do with the nature of reason. Reason cannot remain as it is, in its unsatisfied state of internal conflict. It cannot stay this way because of its reflective character, and its drive for complete, harmonious determination of all things — including itself. The thorniness of the critical path is cause for resentment to the extent that the need to head down it goes unrecognized.<sup>4</sup>

Thus the critical project of self-knowledge is a self-incurring requirement of reason. And the *Prolegomena*, apparently, does not yield a satisfying form of reason's self-knowledge. In its Preface, Kant suggests that it is only the *Critique* which "presents

<sup>&</sup>lt;sup>4</sup> Cf. Kant's remark at A88/B121, about the need for the reader to be "convinced of the unavoidable necessity" of the transcendental deduction of the pure concepts of the understanding "before he has taken a single step in the field of pure reason; for he would otherwise proceed blindly, and after he had wandered about in various ways, he would still have to return again to the ignorance from which he had set out."

the faculty of reason in its whole extent and bounds", something that the mere "preparatory exercises" of the *Prolegomena* are in no position to do (261). *Why?* After all, the *Prolegomena* seems to be concerned with the same things that the *Critique* is: we learn about space and time as "pure forms of intuition", we learn about the "categories" and the "principles of the pure understanding", and we learn about the "ideas" of reason. The separate analyses of the *Prolegomena* correspond roughly to the *Critique*'s Transcendental Aesthetic, Transcendental Analytic, and Transcendental Dialectic. What does the *Critique* give us that the *Prolegomena* cannot — or, at any rate, *does* not?

Kant would reply that the *Critique* gives us a "science". This is required because the problem at hand is foundational: we need to establish the scientific viability of any future metaphysics, and of synthetic judging as such. Kant thinks that only a genuine science, the insights of which rest on the discovery of a rational principle, could provide the sort of completeness in its results that would be necessary for this foundational project to be achieved.<sup>5</sup> But still this sort of response must grate upon contemporary ears, because it is precisely Kant's various claims for the scientific character of his project that sound so quaintly ludicrous to us now.

Yet we could begin to take this idea seriously on exegetical grounds: the *Critique* possesses a certain unity in its exposition that the *Prolegomena* lacks. The careful reader of the *Prolegomena* soon recognizes that the individual analyses that make up that work — the analysis of pure mathematics, then pure natural science, and even the analysis of the hitherto failed enterprise of metaphysics — are not really independent of one another at all. And yet there is no explicit account of how the three analyses relate to one

<sup>&</sup>lt;sup>5</sup> For consideration of this point in relation to one aspect of the *Critique*, see A67/B92 and A80-1/B106-7.

another.<sup>6</sup> The *Prolegomena* lacks a unifying principle: and the role of just such a principle in the *Critique* is precisely what makes it a "science" in Kant's view.

Kant understands the very idea of what a science is in methodological terms: a proper science "must be established according to a method" (*Logic* §95).<sup>7</sup> It is a "whole of cognition as a system" that is "established according to a principle" (*MAN* 467-8). The *Prolegomena* is not a science, because it lacks a principle that could establish the unity of the results of its three analyses. In the *Critique*, this principle is known as the "principle of the synthetic unity of apperception". It is revealed in the Transcendental Deduction, where it is also put to work in establishing a unified account of our capacity to judge a priori about objects that can only be given in experience. This makes it the highest principle of "pure theoretical reason". Now, this principle is merely *mentioned* (not discovered, and not employed) in the analytic *Prolegomena* — and only by way of referring us over to the *Critique* (*P* §36, 318). Hence the *Prolegomena* is not a science, but at best a mere *report* (which is laid out in analytic form) of the *results* of the *Critique*'s "science of an a priori judging reason".<sup>8</sup>

<sup>&</sup>lt;sup>6</sup> For example, the first analysis in the *Prolegomena* reveals space and time as pure forms of intuition — "principles of the possibility" of pure mathematics. In the second analysis, which takes pure natural science as given and reveals the pure concepts of the understanding, we suddenly learn that these pure concepts are also required for a science of pure mathematics (at the end of P §20). In other words, what was elucidated apparently *independently* in the first analysis — without any mention of the pure concepts of the understanding — turns out to require the concepts that only come into view in the second analysis. Similarly, it is only when we get to the *third* analysis that we discover that the operative conception of nature in the *second* analysis — i.e., nature as the "*totality of all objects of experience*" (§16, 295) — has its source in reason. The "*absolute whole of all possible experience*," Kant tells us "is not itself an experience but is a necessary problem of reason" (§40, 328). The three analyses are clearly not independent of one another; at the same time, there is no clear account of their relation to one another. We are lacking a unifying principle.

<sup>&</sup>lt;sup>7</sup> Kant consistently distinguishes between "method" (*Methode*, *Lehrart*), and mere "procedure" (*Verfahren*); see Chapter 1, §4.

<sup>&</sup>lt;sup>8</sup> See Kant's 1783 letter to Christian Garve (10:340) for this gloss on the *Critique*.

The *Critique*'s synthetic method is the source of its status as "scientific and fully illuminating self-knowledge" (A849/B877). My aim in this chapter is to present a textually specific account of the *Critique*'s synthetic method, and to tell the full story about the role of the Transcendental Deduction in carrying it out. To do so, I must first revisit the topic of Kant's general conception of scientific cognition.

2.

Kant claims that scientific cognition begins with an "idea of the whole": apparently, we start out with an indeterminate grasp of the very subject matter of which the science itself will provide the determinate account. An emphatic, if not particularly clear, statement of this idea can be found in a passage from the Appendix to the Transcendental Dialectic. Although this passage specifically concerns the proper role of reason with respect to *empirical* cognition, its main point holds of scientific cognition in general.

If we survey the cognitions of our understanding in their entire range, we find that what reason quite uniquely prescribes and seeks to bring about concerning them is the *systematic* of cognition, i.e., its interconnection from a principle. This rational unity always presupposes an idea, namely that of the form of a whole of cognition, which precedes the determinate cognition of the parts, and contains the conditions for determining a priori the place of each part and its relation to the others. (A645/B673)

One can find a very similar passage in the Doctrine of Method, which tells us that a science begins with a "rational concept" that "contains the end and the form of the whole". According to this passage, a science is a "whole" of cognition that is "articulated and not heaped up; it can grow internally [...] but not externally, like an animal body

whose growth does not add limbs but rather makes each limb stronger and fitter for its end without altering the proportion" (A833/B861).

The curious proposal here is that scientific cognition begins with an idea of the whole and moves *to* a determinate grasp of the parts. What makes for a determinate grasp of the parts is apparently some kind of appreciation of their relation to one another, and their contribution to the end of the whole. This need not imply that there can be no role for an "indeterminate" account of the parts or elements of a science. An indeterminate account of these elements would simply not address their relation to one another, nor would it explain how they contribute to the end of the whole. As I will argue later on, the account of sensibility in the *Critique*'s Transcendental Aesthetic, as well as the account of the understanding in the initial stages of its Transcendental Analytic, are "indeterminate" in this sense.

Let us examine these remarks about scientific cognition more closely. Kant implies that a science cannot get going without a conception of its *end*, which itself bears some very close relation to the "idea of the form of the whole" of the cognition. Now, Kant says that this idea of the whole *contains* the conditions for the determinate cognition of the parts, suggesting that it may be through some kind of *analysis* of the idea of the whole that we arrive at the conditions through which the determinate cognition of the parts is possible. Finally, it seems that the conditions for the determinate cognition of the parts are called *principles*.

In what sense are the modes of inquiry that *we* today most comfortably think of as "sciences" accounted for by this general model? In answering this, it may help to remind ourselves of the concluding theme of the first two chapters: namely, that the *Critique* is

concerned to make something explicit that is generally only implicit in our cognitive practices. Now, Kant's remarks about scientific inquiry as an articulation of the "idea of the whole" seem to pertain to a method of *exposition*, rather than a method of discovery. The demonstration of the unity of a body of knowledge, or the exposition, *follows* a stage of discovery. Lamenting this a bit, Kant writes: "It is unfortunate that only after we have spent much time collecting cognitions, as building materials, in a rhapsodic way at the suggestion of an idea lying hidden in our minds, and indeed after we have, over a long period of time, assembled them in a technical manner, does it first become possible for us to discern the idea in a clear light and draw up a whole architectonically according to the ends of reason" (A834-5/B862-3). So Kant is not supposing that this general model of scientific exposition pertains to the order of discovery in empirical natural science.<sup>9</sup> Nevertheless, a "revolution" in a cognitive practice such as Kant describes in the *Critique*'s Preface is supposed to bring it about that the practice of investigation thereupon acknowledges — tacitly — universal principles of reason as the source of the knowledge in question. It allows that, at least with sufficient progress and reflection, we can articulate the whole according to a priori principles.

Yet there are certain modes of inquiry that could never, at least in Kant's book, achieve the kind of systematic expression that is the mark of a proper science. What are sometimes called the "human sciences" could arguably never figure as proper sciences, for the simple reason that they cannot contain *rational* principles. Take art history as an example: there may well be propositions which function as "principles" for art historians,

<sup>&</sup>lt;sup>9</sup> The method of exposition, I take it, is concerned to give a clear account of the domain of inquiry. Critical philosophy is the 'science of science', because its principles (the principles of the pure understanding) are supposed to determine what it is to figure in the domain of nature *at all* — i.e., as an "object in general". In this respect, critical philosophy is the ultimate "expository" science.

that is, propositions on the basis of which they are able to give a unified account of the practice (say) of painting. These could be propositions about style, or the constraints of the medium, or what have you. But in this sort of case, the character and development of the practice itself is prior to and largely independent of the theoretical apparatus that accounts for it.<sup>10</sup> We might consider a more specific example, this time from second-rate theorizing about *couture*: the proposition that "hemlines rise in tough economic times" conceivably allows one both to account for given sartorial phenomena, as well as to anticipate future sartorial phenomena. In this sense, it is a principle. But again, "principles" like this one shift with the development of the practice itself. (After all, there may well come a time when hemlines *fall* with the stock market.)

The upshot is that we could never expect to give an *exhaustive* account of the principles involved in cognitive projects like these. But Kant thinks that it must be possible to give an exhaustive account of the fundamental principles involved in any cognitive enterprise that merits the title of *science*. Proper scientific cognition rests on *rational* principles. A rational principle is *prior* to the elements of the system. In other words, what figures as an element in a system of a priori rational cognition (e.g., a particular rule or law), as well as any given representation that can be cognitively accommodated by that system, owes that status to the principle or set of principles that express the "form of the whole" of the science in question.

Thus the crucial aspect of scientific investigation seems to be that the knowledge in question arises from an a priori principle. But how do we "get" this sort of principle? It is not helpful to say that "it comes from reason". We are now facing the

<sup>&</sup>lt;sup>10</sup> Or at least ought to be — there have certainly been *critics* (Clement Greenberg comes to mind) who may have had an undue influence on the development of the practice.

methodological question: how does all of this work? Kant's conception of a science as a whole of cognition that is established according to a method, and according to rational principles, suggests that there may be *two* aspects of scientific method. For one thing, we will need to arrive at these principles in some methodologically sound way. And for another, it seems that we must "order" or "establish" the science *according* to these principles once we find them. These aspects are called analysis and synthesis by Kant in his logic lectures. Analysis, he says, "begins with the conditioned and grounded, and proceeds to principles" (*Logic* §117).

Analysis figures in the work of a science proper as the path to the first (or highest) principle of the science. But the *end* of scientific inquiry is systematic knowledge; and it is not evident that analysis alone can give us this. It reveals the principles, and hence the elements, of the subject matter; but it does not account for their unity. In the Preface to the *Critique of Practical Reason*, Kant distinguishes between these two stages of inquiry when what is at stake is the "determination of a particular faculty of the human soul". The analysis yields a provisional grasp of the "parts" of the faculty in question (KpV, 10). Following the analysis is a task that is both "more philosophical and *architectonic*: namely, to grasp the *idea of the whole* correctly and from this to see all those parts in their reciprocal relation to one another by means of their derivation from the concept of that whole in a pure rational faculty" (10). This is the "synthetic" stage inquiry, which is supposed to afford some kind of "examination and guarantee" of the preliminary results of the analysis (10).

Before we could appreciate how this would work, we would need to clarify our conception of what a principle is. For the *Critique*, at any rate, seems to deal in at least

two different kinds of "principle". The bulk of the *Critique* — everything except the Prefaces, the Introduction, and the Doctrine of Method — is called the "Doctrine of Elements". This divides into a Transcendental Aesthetic and a Transcendental Logic, suggesting that the "elements" in question are sensibility and the intellect (broadly construed). There are principles expressing the nature of each of these elements. Space and time, as pure forms of intuition, are "principles" in this sense. So are the "principles of the pure understanding", which are developed from the categories. At the same time, the general account of scientific method that Kant expounds in his logic lectures suggests that there must be another sort of principle in play: the principle on the basis of which the *relation* of these elements to one another can be established. This sort of principle would count as the "first" or *highest* principle of a science. The elements of the critical science would have that status in so far as they are united, or ordered, by means of such a principle.

In order to understand how this general conception of scientific inquiry might actually play out, we need to be particularly clear about our starting points. For an analysis of any kind is only intelligible with respect to its starting point; so if we want to track the analyses in the *Critique*, we would need to be clear about where the work begins. Kant is not as clear about his starting point in the *Critique* as he is, say, in the *Groundwork*. The first two parts of the *Groundwork* contain an analytic ascent to the categorical imperative from the starting point of "ordinary moral rational cognition". The categorical imperative is not a principle of abstruse metaphysical navel-gazing; it is in play, tacitly, in ordinary moral life. Whatever doubts we might have about the analysis that follows, its starting point (the idea of a "good will") is clearly announced. If the general account of scientific inquiry cobbled together here is on target, and is applicable to the *Critique*, then we should first ask about the starting point of the *Critique*'s analysis. Does the *Critique* begin with an analysis of ordinary knowledge — perhaps an analysis of *experience*? Not exactly; but at one point Kant does characterize some unspecified stretch of the *Critique* as an "analysis of experience in general [Erfahrung überhaupt]".<sup>11</sup> The crucial point is that an analysis is only intelligible in relation to its starting point. Hence we will need to consider the starting point of the *Critique* in order to have any chance of discerning the relevant analyses in the text.

Early on in the *Critique*, Kant remarks that any analysis carried out in its pages must be "purposeful" (*zweckmäßig*), which Kant says means that it would be carried out for the sake of making a certain synthesis possible — this synthesis, he remarks, is that "for the sake of which the entire *Critique* actually exists" (A14/B28). The importance of this remark will become evident later on. For now, we should simply hold onto the idea that analysis and synthesis are complements in scientific investigation: the analysis is carried out for the sake of the ensuing synthesis, and it is in virtue of this synthesis that the investigation becomes a science.

Thus, in order to understand the synthetic method of the *Critique*, we will need to address the following questions:

- (1) What is the starting point of the *Critique*? What does it take as given?
- (2) What are the "elements" of this science?
- (3) What is the first principle of this science, and where is the analysis that uncovers it?
- (4) What is the "synthesis"?

<sup>&</sup>lt;sup>11</sup> See Kant's letter to J.S. Beck, 20 January 1792 (11:313; 315).

The first two questions will be addressed in the remainder of Part I of this chapter. The latter two questions will be addressed in Part II, through a close reading of the Transcendental Deduction. The Transcendental Deduction, I shall demonstrate, consists of complementary strands of analysis and synthesis. The first half of the Deduction is an analytic ascent to the first principle of the critical science; the second half is the synthesis that establishes scientific unity of the project. In Part III, I summarize this account.

#### 3.

Let us begin with the surprisingly vexed issue of the *Critique*'s starting points. What does it take as given? It is tempting to point to the remark at the beginning of the Introduction — where Kant nods to the empiricists, proclaiming "there can be no doubt that all of our knowledge begins with experience" (B1) — and conclude that the starting point of *Critique* itself must be experience.<sup>12</sup> But as we have already seen, Kant describes some unspecified portion of the text as an "analysis of experience in general [*Erfahrung überhaupt*]". The *Critique* does not begin with an empirical premise: the "überhaupt"-formulation tells us that it is not any particular experience, or collection of them, that is our starting point, but rather some idea of "experience as such". It should go without saying that it is far from clear what this is supposed to mean.

There are three candidate "answers" to the question about the *Critique*'s starting points. I will argue here that they come down to three different ways of saying the same thing. The first candidate is "experience in general"; I will return to it after considering

<sup>&</sup>lt;sup>12</sup> This is a pervasive feature of Patricia Kitcher's work on Kant, from *Kant's Transcendental Psychology* to very recent work ("The Presupposition of Kant's Transcendental Deduction"). Vasilis Politis argues

the other two. The second candidate comes attached to the idea that the method is synthetic. While the analytic *Prolegomena* takes as given particular systematic expressions of rational activity (particular sciences, or putative sciences), the *Critique* claims to admit only reason itself. Its starting point is the *faculty* of reason, rather than an expression of its actuality. I quote the passage now in full:

In the *Critique of Pure Reason* I intentionally went to work on this question [about the possibility of metaphysics] *synthetically*, namely in such a way that I investigated into pure reason itself and in this very source sought to determine the elements as well as the laws of its pure employment according to principles. This work is difficult and requires a resolute reader to think his way gradually into a system which lays nothing as given for its basis except reason itself, and thus, without resting on any other fact [Faktum] seeks to develop the cognition from its original germs. (*P* §4, 274)<sup>13</sup>

So Kant says. But when we turn to the text of the *Critique* itself, we find no clear indication of how to understand this, apart from the idea that the *Critique* is *reason*'s self-investigation. And *this* characterization of the project is so general that it offers little concrete guidance for getting our bearings in the text.

Moreover, when it comes time to specify what our starting point is supposed to be, Kant says nothing specifically about reason. Instead, at the end of the *Critique*'s Introduction — just as we are about to embark on the work proper, the Doctrine of

against Kitcher on this, presenting an account of the *Critique*'s starting point in "experience in general" ("The Apriority of the Starting-point of Kant's Transcendental Epistemology").

<sup>&</sup>lt;sup>13</sup> In the Doctrine of Method, Kant contrasts the genuine *criticism* of reason with Hume's "censorship" of reason. Hume's project "subject[s] the *facta* of reason to examination [Prüfung] and, when necessary, to blame" (A760/B788). The "*facta*" in question are presumably particular *claims* of reason. The Humean project of censorship is contrasted with critical philosophy, "which subjects to evaluation [Schätzung] not the *facta* of reason but reason itself, as concerns its entire faculty and suitability for pure a priori cognitions" (A761/B789). In the passage from the *Prolegomena*, Kant says that the *Critique* takes as given
Elements — Kant says that the only "preliminary" is his characteristic thesis about the heterogeneity of sensibility and understanding.

All that seems necessary for an introduction or preliminary is that there are two stems of human cognition, which perhaps arise from a common, but to us unknown, root — namely, sensibility and understanding. Through the first objects are *given* to us, while through the second they are *thought*. To the extent that sensibility may contain a priori representations, which constitute the condition under which objects may be given to us, it would belong to transcendental philosophy. (A15/B29-30)

So we have a third candidate starting point: the thesis about the heterogeneity of sensibility and understanding, the former a capacity to represent objects in so far as they can be given, and the latter a capacity to represent them in so far as they can be thought. I refer to this as the *heterogeneity thesis*.

On the face of it, the heterogeneity thesis is not a controversial starting point. Nearly any philosopher, with the possible exception of a Platonist committed to disparaging the cognitive value of sensory representations altogether, would be prepared to give some lip service to the idea that sensibility and understanding contribute in distinct ways to knowledge. In this passage, however, Kant suggests a further ramification of the view that clearly is controversial. For it seems that if this thesis is to serve as the starting point of critical philosophy, then it entails not only that sensibility and understanding contribute in distinct ways to our cognition, but also that each capacity is constituted by certain a priori representations. At any rate, this is the condition under which *sensibility* would figure in "transcendental philosophy", as Kant says in the passage above. If the heterogeneity thesis entails this much, even as a mere

nothing except the "fact [Factum]" of reason itself, meaning that the only starting point is the *faculty* of reason, and not any particular actualization of its nature (whether legitimate or fraudulent).

"preliminary", it is much more radical than any garden-variety empiricist or rationalist would be prepared to accept. The rationalist digs in his heels with the idea that sensibility is constituted a priori, while the empiricist would dispute that either capacity could be "constituted a priori".

Furthermore, this preliminary thesis shapes everything that follows. We suppose the heterogeneity thesis at the outset, it seems — and with it, the more radical implication that sensibility and (presumably) the understanding "contain" a priori representations. So, when we turn the page, and begin the Transcendental Aesthetic, we set out to give an account of sensibility that presupposes the possibility of "isolating" it both from "what the understanding thinks through its concepts", as well as from "everything that belongs to sensation". This is to leave us with the a priori representations that we suppose sensibility must "contain" — i.e., nothing but "pure intuition and the mere form of appearances" (A22/B36). As we have already seen, Kant makes a similar move at the outset of the Transcendental Analytic, announcing that the pure understanding "separates itself" from "everything empirical" and from "all sensibility". Hence, the heterogeneity thesis seems to underwrite a methodological strategy, which I shall call the *strategy of isolation*.

The heterogeneity thesis says that the elements are irreducible to one another, and this irreducibility is cashed out in terms of the idea that each capacity is characterized by a distinct mode of representation (sensibility by intuitions, understanding by concepts). This idea of a "mode of representation" pertains to the distinct *expression* of each capacity. Is there not something ambiguous about the notion of "representation" in this story? For Kant takes a further step beyond the mere idea that each element is distinguished by its mode of *representing*: he supposes, as part of the heterogeneity thesis, that each element "contains" a priori representations that are constitutive of its capacity.

The hidden presupposition here, it seems, can be traced back to the B-edition Preface, where Kant puts forward a new model for conceiving of the possibility of a priori cognition. It is an "altered method of our way of thinking, namely that we cognize of things a priori only what we ourselves have put into them" (Bxviii). We cannot conceive of the possibility of a priori cognition, Kant argues, if we suppose that our cognition must conform to its objects. With a nod to Copernicus, the typical presupposition about cognition — that it must conform to its objects — is inverted. A new explanatory model is admitted: we are to conceive of the possibility of a priori cognition in terms of the idea that our cognitive capacity makes some necessary *contribution* to cognition.

This contribution can only be an anticipation — not a projection — of what can be cognized at all. The contribution would have to be independent of any particular occasion of the exercise of these capacities; in other words, this contribution would not be generated as a mere *response* to the world. And this seems to lead Kant to the idea that we can conceive of this contribution in terms of certain a priori representations that distinguish each capacity.

When each capacity is considered in isolation from the other, and from everything empirical, we discover the a priori representations that are constitutive of the capacity. This means that we can give an *exhaustive* account of each capacity in terms of these representations. The point is made emphatically at the outset of the Transcendental

133

Analytic: the table of the pure, elementary concepts of the understanding must "fully exhaust the entire field of the pure understanding" (A64/B89). The categories exhaust the capacity of the understanding in the sense that any particular empirical concept (*Erfahrungsbegriff*) is simply a pure concept of the understanding considered *in concreto*; they are "forms" of concepts in general (A567/B595). Likewise, the pure forms of intuition exhaust the capacity of sensibility, at least in the sense that anything that can figure as a sensible representation must "stand under" these forms — no singular representation can register in our consciousness if it is not given in space or time.<sup>14</sup> Thus, Kant's heterogeneity thesis says *not only* that sensibility and understanding make distinct contributions to knowledge, *but also* that each "contains" a priori representations that are constitutive of its capacity.

Yet this "preliminary" for Kant's entire investigation is something that neither an empiricist nor a rationalist could accept. On what basis can it be invoked as a *preliminary*, rather than as a conclusion? We can address this question by considering the relation of the heterogeneity thesis to the other main candidate for the *Critique*'s starting point: "reason itself". Presumably, Kant says that the *Critique* takes nothing as given except reason itself because it is supposed to be reason's project of self-knowledge. But this oversimplifies matters: it is reason's examination of its *theoretical* capacity. Its central question concerns the viability of a rational capacity to judge a priori about objects that can only be given in experience (i.e., judging synthetically a priori). The passages leading up to the presentation of the heterogeneity thesis as our starting point

<sup>&</sup>lt;sup>14</sup> For one clear statement of this, see Kant's reminder in the Deduction about the principal result of the Aesthetic: "The highest principle of the possibility of all intuition in relation to sensibility was, according to the Transcendental Aesthetic, that all the manifold of sensibility stand under the formal conditions of space and time" (§17, B136).

lay out the framework of the project as reason's examination of its theoretical capacity (A14-5/B28-9; see also Bx).

The presentation of the heterogeneity thesis as a "preliminary" follows upon Kant's careful consideration of what he calls the "general problem of pure reason" (\$VI, B19). And the "general problem of pure reason" arises as a generalization of the problem of metaphysics. With either formulation, the issue at stake is the viability of judging that claims both apodictic necessity and objective validity. In the first-edition Preface, Kant attributes this problem about the possibility of apodictic, objectively valid a priori judging, to the nature of reason (Avii). He also supposes that the only resolution to such a problem is for reason to examine its own capacity. For the claims of reason are at stake, and reason "recognizes no other judge than human reason itself" (A752/B780; see also *P*, 263).

No doubt, the argument leading to the presentation of the heterogeneity thesis is somewhat suppressed. But the line it takes must be something like the following. If the cognition of objects is possible a priori, then it must be independent of any particular occasion of objects appearing before the senses. Yet if the critical science is to account for the possibility of a priori cognition of *objects*, then it must rely on the existence of a priori conditions of the *possibility* of objects appearing before the senses. This is why Kant admits the heterogeneity thesis as a "preliminary", underscoring the idea that sensibility must "contain" a priori representations if it is to figure in this science at all.

The heterogeneity thesis figures as a "preliminary" under the proviso that we accept Kant's description of the general problem of pure reason, and his prescription for

its resolution.<sup>15</sup> There is no real tension between the idea that the *Critique* takes nothing as given except reason itself, and the idea that its only preliminary is the heterogeneity thesis. Reason is investigating itself as a capacity for a priori theoretical cognition. It does so by inquiring about the possibility of judging synthetically and a priori. What distinguishes *synthetic* a priori judging from *analytic* (a priori) judging is that it concerns empirical *objects*, rather than the content of concepts. *Supposing* that it is possible to judging. As such, it would have to be constituted by its "own" a priori representations if it is to be genuinely distinct from the capacity for thought, and yet part of an "a priori judging reason". So if this is how pure theoretical reason investigates its own capacity, then sensibility and understanding are, in this respect, "elements" of it.<sup>16</sup> They are not

<sup>&</sup>lt;sup>15</sup> In the Doctrine of Method, Kant suggests that the starting point of critical philosophy is the "nature of synthetic a priori propositions" (A762/B790). In doing so, he compares his project that of Hume, whom he calls a "geographer of human reason" (A760/B788). In these pages, Kant argues that no putatively critical project — a positive determination of "the boundaries of my possible cognition" — can be carried out in an empirical manner (A758/B788). Kant sets up an analogy between an ordinary geographer and a geographer of human reason. As an ordinary geographer, I investigate the earth from the perspective of standing on its surface. What I see is always "a space around me [...] in which I could proceed farther" (A759/B787); the earth appears as a "plate" that always extends to an ever-present horizon. So how do I get so far as knowing the "complete boundary" of the earth? My investigation presupposes a conception of the whole — that the earth is a sphere — and I infer the magnitude of the whole from the measure of a small part of its surface. Kant cashes the analogy out a few pages later: "Our reason [...] must [...] be compared with a sphere, the radius of which can be found out from the curvature of the arc on its surface" (A762/B790). The measure of this curvature, he says, is "the nature of synthetic a priori propositions". From this, the "content and [...] boundary [of reason] can be ascertained with certainty". The analogy indicates that the starting point of the critical examination of reason is the consideration of the nature of synthetic a priori judging. Where does Kant consider the nature of synthetic a priori judging? Early in the Introduction, which leads to the presentation of the heterogeneity thesis.

This passage from the Doctrine of Method tells against Kitcher's claim that the starting point of critical philosophy is experience. To suppose that it could be is to confuse Kant's project — the true criticism of reason — with something closer to Hume's.

<sup>&</sup>lt;sup>16</sup> Kant also distinguishes sensibility and understanding as the "lower" and "higher" cognitive faculties as another way of making the point that they are elements of a larger whole (see, e.g., *Anth.*, 140).

*derived* from reason — reason is not the "unknown root" — but rather, they are part of reason only in so far as they are unified by it, as we shall see.<sup>17</sup>

Thus, the heterogeneity thesis falls out of the idea of the project itself, as pure theoretical reason investigating its own capacity. This is also, I think, another way of saying that the starting point is "experience in general", or *Erfahrung überhaupt* — the first candidate for the *Critique*'s starting point. Kant's "überhaupt"-formulation tells us that we are dealing with an idea of reason: i.e., not any particular experience, or finite set of them, but something like the "sum total" of them.<sup>18</sup> Of course, we can only grasp this modally, in terms of the *capacity* for experience. Bearing in mind that Kant takes experience to be "empirical knowledge" (B234), the compatibility of this candidate starting point with the other two should be clear.

## 4.

What does the heterogeneity thesis, and with it the strategy of isolation, tell us about the *Critique*'s "synthetic method"? To answer this, consider Kant's reminder of the heterogeneity thesis at the outset of the Transcendental Logic. Each capacity, he says, is characterized by a distinct mode of representation — sensibility by intuitions, and understanding by concepts — which are the "elements of all our cognition" (A50/B74). *Knowledge requires both*: "Thoughts without content are empty, intuitions without concepts are blind" (A51/B75). Thinking is only a source of knowledge to the extent that it is applicable to what can present itself to us in sensibility, our capacity to enjoy given representations. And intuition is only a source of knowledge in so far as it can be brought

<sup>&</sup>lt;sup>17</sup> I take it that the heterogeneity thesis would be compromised if we could indicate what this "unknown root" is.

under concepts. As cognitive faculties, sensibility and understanding are fundamentally distinct: the form of representation that characterizes the one can never be reduced to the form of representation that characterizes the other. At the same time, they are necessarily unified: "These two faculties or capabilities can never exchange their functions. The understanding can intuit nothing, and the senses can think nothing. Only through their unification can cognition arise" (A51/B75-6). This thesis about the unity of sensibility and understanding complements the heterogeneity thesis; I call it the *cooperation thesis*.

When we begin the Transcendental Deduction, we have in place a preliminary account of sensibility as constituted by two "forms of intuition", and a preliminary account of the understanding as constituted by twelve "forms of thought". But merely pointing to the a priori "forms" of representation that are constitutive of sensibility and understanding respectively is not enough to entitle us to a conception of these as *elements* of a capacity for a priori knowledge of nature. Their status as cognitive capacities hangs in the balance: the cooperation thesis tells us this. Thus, the strategy of isolation must be complemented by an argument that, in effect, brings these elements together. This is why the method is "synthetic", and this is why the Deduction is the pivot on which everything else in transcendental philosophy turns.<sup>19</sup>

<sup>&</sup>lt;sup>18</sup> See *P* §40, 328.

<sup>&</sup>lt;sup>19</sup> Dieter Henrich remarks that the argument of the Deduction is "synthetic" for something like this reason: the Deduction "proceeds on the basis of the fact that the two doctrines of the *Critique* are initially developed independently of one another — the doctrine of the categories as functions of unity in self-consciousness and the doctrine of space and time as given representations" ("The Proof Structure of the Transcendental Deduction," 649). He continues: "Within the structure which Kant had already given his book, the advantages of a construction according to the synthetic method were in any case obvious. This construction allowed him to ground the two fundamental positions of critical philosophy, the sensible a priori and the active role of the understanding in knowledge, separately — and unite them by means of a single argument" (650).

Henrich's gloss on the idea of an *analytic* method is quite odd, however; and it seems not to draw at all on what Kant actually says about it in the *Prolegomena*. Henrich, attempting to address J.S. Beck's suggestion that the Aesthetic should *follow* the Deduction (see Kant's 1797 letter to J.H. Tieftrunk, 13:463;

5.

The Transcendental Deduction is the part of the *Critique* where we finally overcome the 'provisionality' of the initial proceedings. Before turning to the details of the Deduction, I would like to consider this provisionality further. For it is questionable whether it is even possible to present an account — provisional or not — of one element in strict abstraction from the other.<sup>20</sup> At any rate, the cooperation thesis indicates that as long as one element is held off from view, the fundamental nature of the other is as well.<sup>21</sup>

Even if the Transcendental Analytic begins under the auspices of the strategy of isolation, it is not long before this is to be overcome in the Transcendental Deduction. But the Transcendental Aesthetic stands all on its own, outside of the Transcendental Logic altogether. Is this entire episode of the *Critique* truly "provisional", as I have suggested? For the Aesthetic at least appears to contain arguments that would establish the status of sensibility as a cognitive capacity. Let us try to ascertain the scope and force of the strategy of isolation in the Transcendental Aesthetic.

in Zweig, 238), replies on Kant's behalf that such an arrangement would prohibit the Deduction's appeal to "the results of the Aesthetic as to fact"; and this, Henrich curiously concludes, would mean that the Deduction would have to follow an analytic method, and would have to demonstrate "the *necessity* of the forms of intuitions" in order to complete its work (649-50).

I find this curious, but I suppose that Henrich's point is this: the argument of the Deduction succeeds precisely *because* the doctrine of the Aesthetic is developed independently of the Analytic. The pure forms of intuition, as principles of the "possibility of all intuition", can be considered to be genuinely independent of the principle of the synthetic unity of apperception for this reason. Their independence is secured because the Transcendental Aesthetic lies outside of the Transcendental Logic altogether. So, if Kant took Beck's suggestion, he wouldn't have *this* way of completing the Deduction's argument. For more on this, see §11, below.

<sup>&</sup>lt;sup>20</sup> This is a characteristically Hegelian criticism of Kant's method.

<sup>&</sup>lt;sup>21</sup> To clarify: their fundamental nature as *cognitive capacities*.

The Aesthetic opens with brief introductory remarks about a hylomorphic

distinction that one could draw between sensation and the form of intuition, and how the separation of the latter from the former reveals "pure intuition" (§1). There is some kind of argument here, but not a particularly robust one. The "argument" consists only of the remark that what orders sensation cannot itself be sensation. If we suppose that there is any "ordering" to mere appearances at all, then we invoke some notion of form. While the matter of appearances is certainly given to us a posteriori, the form — *if* appearances are necessarily ordered in a certain way — must "lie ready for [the matter] in the mind a priori, and can thereby be considered in abstraction from all sensation" (A20/B34). This remark may seem to offer further justification for at least one "side" of the heterogeneity thesis, the one that concerns the a priori constitution of sensibility. But in truth, this remark would do nothing to sway the rationalist who supposes that any such "ordering" in appearances is due to the intellect. Perhaps the remark would present to the empiricist a new, and possibly more palatable, way of conceiving of the very idea of a priori representation; but again, it is not clear that it would speak in favor of the idea that these representations must belong to *sensibility*. On balance, the remark is best conceived as a further clarification of this one "side" of the provisional heterogeneity thesis: it tells us how we could conceive of sensibility's "containing" a priori representations, but it does not demonstrate that we *must* conceive of sensibility in this way.

The Aesthetic then divides into two main parts, one concerned with space and the other with time. No reason is given for why space and time are named as the candidate pure forms of sensibility. Each of these two main parts consists of a Metaphysical

Exposition and a Transcendental Exposition.<sup>22</sup> Kant specifies the metaphysical exposition in the following way: "I understand by *exposition* [...] the clear [deutlich] (even if not complete) representation of what belongs to a concept; the exposition is *metaphysical* if it contains that which exhibits the concept as *given a priori*" (B38). The Metaphysical Exposition spells out what is already thought in the idea that sensibility is a source of cognition that is constituted a priori by space and time: namely that space and time are neither pure nor empirical concepts but rather are the pure forms of intuition, and that space and time are *themselves* intuitions. The heterogeneity thesis, understood as falling out of the idea that the *Critique* is pure theoretical reason's self-examination, allows us to postulate that sensibility is constituted a priori. Hence, the Metaphysical Exposition is an *analytic clarification* of the idea that sensibility is constituted a priori by space and time.

Kant defines a transcendental exposition in general as "the explanation of a concept as a principle from which the possibility of other synthetic a priori cognition can be understood" (B40). Such insight requires that we see that there really is some knowledge which "flows from" the relevant concepts, and that such knowledge is only intelligible under the presupposition of those concepts. The "concepts" at issue are space and time as the pure intuitions constitutive of sensibility a priori. Kant points to geometry (in §3) and the "general theory of motion" (in §5) as the knowledge which is not intelligible apart from the conception of space and time which was laid out in the Metaphysical Exposition. The Transcendental Exposition apparently shows that space

<sup>&</sup>lt;sup>22</sup> These remarks pertain to the second-edition Transcendental Aesthetic. When I refer to "the Metaphysical Exposition" or "the Transcendental Exposition", let that stand as shorthand for the Metaphysical Exposition both of space and time *together* (i.e., §2 and §4), and likewise with the Transcendental Exposition of space and time *together* (§3 and §5).

and time are sources of synthetic a priori cognition. Moreover, it apparently gives the sort of justification that was lacking before: for if we suppose that these sciences are "actual", then since we have shown that they are not intelligible apart from the conception of sensibility laid out in the Metaphysical Exposition, we would have also shown that we *must* conceive of sensibility as the Metaphysical Exposition instructs.

Indeed, Kant suggests that the Metaphysical Exposition is incomplete without the Transcendental Exposition. In the concluding remarks of the Aesthetic, Kant refers to the account presented in the Metaphysical Exposition as "our opinion [Meinung] in regard to the fundamental constitution [Grundbeschaffenheit] of sensible cognition in general" (A41-2/B59). We should raise our eyebrows at this, since Kant claims in the Preface that the *Critique* should deal only in "absolutely necessary" truths, for it is itself supposed to set the example of "apodictic (philosophical) certainty" (Axv); opinion and mere hypothesis should have no place in its argument (Axvii). This concluding remark in the Aesthetic continues by pointing to the example of geometry — implying that this "opinion" becomes, upon the consideration of geometry, "certain and indubitable". In the second-edition version, Kant also refers us back to the Transcendental Exposition of space (§3), which draws on geometry.

Thus, it appears that the Transcendental Exposition is supposed to establish the apodictic certainty of what was first presented as some kind of opinion. This impression is reinforced by Kant's remark against the Leibnizian, who supposes that space and time are "relations of appearances that are abstracted from experience although represented confusedly in this abstraction" (A40/B56-7). The Leibnizian denies that sensibility is constituted a priori. The unfortunate consequence of that denial is that the Leibnizian

cannot accommodate the "validity or at least the apodictic certainty of a priori mathematical doctrines in regard to real things" (A40/B57). Thus Kant implies that what is supposed to recommend his own account of the a priori constitution of sensibility, articulated in the Metaphysical Exposition, is that it can account for the apodictic certainty of mathematics as a science of nature.

We have reason to be wary of whatever legitimation the Transcendental Exposition purports to offer for the claim that sensibility is constituted a priori by space and time. For Kant claimed that the *Critique* takes nothing as given except the faculty of reason itself; this was contrasted with the argument of the *Prolegomena*, which takes particular scientific expressions of reason as given. The argument of the Transcendental Exposition is reminiscent of the *Prolegomena*, at least in its appeal to the viability of existing sciences.<sup>23</sup> Now, it may be granted that the Aesthetic does not consist of an analysis of these sciences that *culminates* in the revelation of space and time as principles of synthetic a priori knowledge. Nevertheless, appeal is made to geometry and the general theory of motion as "givens" to legitimate the idea that sensibility is a source of a priori knowledge in virtue of its a priori constitution.<sup>24</sup>

<sup>&</sup>lt;sup>23</sup> Admittedly, this remark only clearly pertains to the Transcendental Exposition's appeal to geometry. It is hard to know exactly what Kant may have in mind as the "general theory of motion": is this an existing science? Is it supposed to be some aspect of Newtonian mechanics, or is it a *merely* possible science? (N.B.: *Arithmetic* is the actual science that analytically leads to the representation of time as a pure form of intuition in the *Prolegomena*.)

<sup>&</sup>lt;sup>24</sup> Stephen Engstrom remarked (in conversation) that Kant may mean to appeal to geometry and the "general theory of motion" as *merely* possible sciences. The thought, then, runs like this: without the conception of (e.g.) space that is laid out in the Metaphysical Exposition, we would have no prospect for understanding how the science of geometry is possible. The argument of the Aesthetic does not appeal to the actuality of these sciences, because in order to do so it would have to address the applicability of their determination of a priori intuition to *empirical* intuition — since this is the only way in which mathematics (at any rate) yields knowledge (see §22, B147). This thought defends Kant's introduction of geometry and the general theory of motion into the pages of the *Critique*. It also limits the justificatory scope of the Transcendental Exposition: whatever justification it may provide for the account of sensibility laid out in the Metaphysical Exposition, it does not touch the crucial issue — the status of sensibility as a source of *cognition* in virtue of its a priori constitution. Still, I wonder if this really helps, since Kant's retort to the

The Transcendental Expositions rely on resources that are not to be admitted into the program of the *Critique*, according to Kant's remarks in the *Prolegomena*. Therefore, whatever legitimation the Transcendental Expositions may provide for the idea that sensibility is constituted a priori as a source of cognition is provisional.<sup>25</sup> The Transcendental Expositions of space and time are mere *parerga*. If Kant draws on the results of the Aesthetic later on (as he will, in the Transcendental Deduction), he may only legitimately refer to the Metaphysical Exposition — an analytic unpacking of what is thought in the idea that sensibility is constituted a priori as one of the two heterogeneous sources of knowledge. This latter idea, the implications of which are worked out in the Metaphysical Exposition, is itself introduced only on the basis of the recognition of the "general problem of pure reason", and the very idea of a "critical" solution to it.

Leibnizian is that he cannot accommodate the "apodictic certainty of mathematical doctrines in regard to real things" (A40/B57), which suggests to me that Kant supposes that geometry is meant to be in view as an *actual* (and not a merely *possible*) science.

<sup>&</sup>lt;sup>25</sup> Daniel Warren argues that the Metaphysical Exposition is concerned with the *origin* of the representations of space and time; Kant is not, as Allison and others have suggested, trying to demonstrate their objective validity. He claims that it is only when their a priori origin is established "that Kant feels that the objective validity of the representation can be assessed" ("Kant and the Apriority of Space", 224). Warren does not discuss the role of the Transcendental Exposition in any detail. It seems that he thinks that the demonstration of their objective validity is pushed off until the Transcendental Deduction (see the final sentence of his paper, and 220 n.53). Nevertheless, I don't find him to be especially forthcoming on this matter.

Warren is probably right to be non-committal, since Kant is rather confusing on this issue. Just prior to the Transcendental Deduction, Kant suggests that we already have in place a conception of space and time as objectively valid a priori (§13, A85/B118; see also A87/B119-120). The remark brings with it the unfortunate implication that the Transcendental Exposition is supposed to be doing some real work — i.e., establishing the objective validity of space and time. The waters are muddled further by Kant's remark that we see the need to demonstrate the objective validity of *space* (nothing is said about time) *only* once we consider the pure concepts of the understanding, and recognize that a deduction is required for them (A88/B120). Despite this remark, however, he speaks as if the transcendental deduction of space has already been given, presumably in the Transcendental Exposition. *Nevertheless*, Kant does not seem to allow himself to draw on the Transcendental Exposition's argument in favor of the objective validity of space and time in the text of the Transcendental Deduction.

If the Transcendental Expositions of space and time are mere parerga, then we must be awaiting an argument that will entitle us to the idea that sensibility is a source of cognition in virtue of its constitutive representations. The anticipated argument of legitimation does not receive help from existing sciences; for the *Critique*, Kant tells us, is "robbed of all help from other sciences" (P §5, 279). This methodological stricture is imposed upon its argument because the *Critique* is supposed to be "a determination of all pure cognitions a priori" and "the standard [Richtmaß] and thus even the example of all apodictic (philosophical) certainty" (Axv). It cannot be that standard if it relies on the apodictic certainty of other sciences. This methodological stricture is part and parcel of the synthetic method, as described in the *Prolegomena*. The argument that we are waiting for — *the* argument of legitimation in the *Critique* — is the Transcendental Deduction.<sup>26</sup>

## II.

6.

It is widely recognized that the Deduction is supposed to address a question of entitlement. But this point is somewhat obscured by its often rehearsed epithet — the demonstration of the "objective validity" of the categories. Perhaps we can better appreciate the justificatory task at hand if we consider the project at large: at stake is pure theoretical reason's entitlement to conceive of itself, and to act, as a cognitive capacity.

<sup>&</sup>lt;sup>26</sup> Lewis White Beck's remarks on the synthetic method in Kant's moral philosophy are similar in spirit to my account of the synthetic method in the first *Critique*. The "synthesis", Beck suggests, is to be found in the deduction argument, without which "all our statements must be hypothetical and problematic" (*A Commentary on Kant's* Critique of Practical Reason, 12). Of the *Groundwork* in particular, Beck claims that its "synthetic" third section "is aimed at justifying the assertion of what was only entertained in the

To follow this idea further, we might consider the general methodological framework of the Critique, which is supposed to be reason's scientific self-knowledge. This tells us that Kant's general conception of scientific method ought to serve as some kind of guide to the text. For a moment, let us just consider what that general conception of scientific method indicates, without worrying about the details of the text. Since the *Critique* is reason's self-knowledge, it would make sense that some preliminary conception of reason is the "idea of the whole" from which we begin. As I argued in Chapter 1, the front matter of the *Critique* — the Prefaces and Introduction — put in place a preliminary conception of *reason*, by considering the nature of rational cognition (see Figure 3, and Chapter 1, §4). What immediately follows this as we embark upon the main text of the *Critique* (the Doctrine of Elements) are separate accounts of sensibility and understanding, each considered in "isolation" or "separation" from the other. In Part I of this chapter, I argued that the strategy of isolation on its own cannot yield an account of either element as a *cognitive* capacity. As yet, we have no basis to claim that the characteristic representations that we take to be constitutive of each element are "objectively valid". In other words, we still have no basis to claim that these representations put us on to *objects*.

This gives us another way of thinking about the justificatory work of the Deduction, which need not be seen as incompatible with the standard line: the Deduction establishes that the elements in question are indeed *cognitive capacities*. Kant's general conception of scientific method gives us some idea of how this should work. We are still awaiting the principle that would yield *determinate* knowledge of the elements,

first two sections" (which Kant said were "merely analytic"). I am, unfortunately, in no position to evaluate Beck's reading myself.

sensibility and understanding. And the general account of scientific method tells us that this should involve our grasping the relation of these elements to one another, and their contribution to the end of the whole, on the basis of a rational principle. As I aim to demonstrate in Part II of this chapter, this is the work of the Transcendental Deduction.

Kant's general model of scientific method, I will argue, applies quite directly to the Transcendental Deduction chapter. Kant divides this chapter into two parts with a signpost, an infamous remark that Kant makes about half-way through the Deduction chapter. He tells us that "a beginning [Anfang] of a *deduction* of the pure concepts of the understanding has been made". It has, moreover, *just* been made — "in the above proposition [obigen Satz]" (B144). Halfway through a text of punishing difficulty, we suddenly learn that we are just getting started.

Many commentators have remarked on the Deduction's two-part structure, but there is little agreement among those commentators about what is going on in each of the two parts, and how they relate to one another.<sup>27</sup> The general account of the *Critique*'s method that I have offered prepares us to understand why the Deduction has two parts. The strategy of isolation leaves us with two separate and (therefore) preliminary accounts of the elements, sensibility and understanding. We still need the rational principle that will allow us to unify them into the account of a single cognitive capacity. The principle that we are looking for is uncovered in the first half of the Transcendental Deduction. It is revealed by further consideration of the preliminary conception of reason that is put

<sup>&</sup>lt;sup>27</sup> I allude to the debates about the "proof-structure" of the Deduction. Most of the contributors to this debate, I believe, have failed to consider adequately the role of the Deduction in the *Critique* as a whole, and how the methodological framework of the *Critique* as *reason's self-knowledge* should bear on one's interpretation of the Deduction. For this reason, my account of the Deduction does not sit squarely within the existing terms of the debate — although I certainly have learned something from following it.

forward in the *Critique*'s front matter. For this reason, we might think of the Deduction as the central text of reason's scientific self-knowledge.

When Kant announces that the deduction has just *begun*, he is telling us that he has found the *principle* that will make this deduction go through. At one point in the Transcendental Dialectic, Kant glosses the term "beginnings [Anfänge]" with "principles [Prinzipien]" (A652/B680). Kant alludes to an ancient idea about scientific investigation, according to which the path *to* first principles is distinguished from the path *from* first principles. Principles are beginnings, just as the ancient Greek term *arche* would suggest. Thus, the signpost announces that the analysis has drawn to a close, because some first principle has been uncovered.<sup>28</sup> It heralds the beginning of the synthesis. The deduction — in the *strict* sense — is the "synthetic" portion of the argument; it is the argument that should make the *Critique* a "science". To see whether and how this succeeds, we must examine the details of the Transcendental Deduction.

7.

The Deduction, I am suggesting, picks up where the *Critique*'s front matter left off. The Prefaces make the call for reason to examine its own *theoretical* capacity, and the Introduction works out a rough account of the parameters of this investigation. The front matter concludes with the postulation of the heterogeneity thesis as our only "preliminary". Now, all on its own, the heterogeneity thesis leaves us with a problem. This is the problem that the Deduction must address, and it could be formulated as

<sup>&</sup>lt;sup>28</sup> Liddell and Scott report that the primary meaning of the Greek *arche* is "source" or "beginning"; in the context of scientific investigation, however, it means "first principle" (*Greek-English Lexicon*, 252).

follows: we are faced with the possibility that the unity of what is given in intuition is incommensurate with the logical unity of judgment that is expressed in the categories.<sup>29</sup>

The first section of the Transcendental Deduction chapter (§15) is largely a restatement of the heterogeneity thesis. Since the problem of the Deduction is clearly connected to the heterogeneity thesis, this should not be surprising. The heterogeneity thesis allows for the possibility that we could have a *merely* sensible representation. So the Deduction begins: "The manifold of representations can be given in an intuition which is merely sensible, i.e., is nothing but receptivity; and the form of this intuition can lie a priori in our faculty of representation, without being anything other than the mode in which the subject is affected" (B129). The restatement of the thesis cashes it out in terms of the distinction between receptivity and spontaneity. Wanting to ward us off thinking that the nature of sensible representation could point analytically to some claim about the spontaneity of the understanding, Kant stresses that the combination of the manifold is a "spontaneous" representation that cannot be "contained in" the characteristic representation of sensibility. The Deduction continues:

Only combination (*conjunctio*) of a manifold in general can never come into us through the senses, and thus cannot be already contained in the pure form of sensible intuition; for it is an act of spontaneity of the power of representation, and since one must call this understanding, in distinction from sensibility, so all combination, be we conscious of it or not [...] is an action of the understanding [...]. (B129)

<sup>&</sup>lt;sup>29</sup> Before laying out the table of categories, Kant asserts: "The same function which gives unity to the various representations *in a judgment*, also gives unity to the mere synthesis of various representations *in a intuition*. Expressed generally, it is called the pure concept of the understanding" (A79/B104-5). The problem is revealed by sketching the nature of the solution.

The need for the Deduction stems from the fact that the heterogeneity thesis was taken as a "preliminary" for pure theoretical reason's self-investigation. Within this framework, the heterogeneity thesis entails that sensibility and understanding are each constituted a priori, and that the representations characteristic of the one can never be reduced to the representations characteristic of the other. This figures as a "preliminary" because it is recognized as the only way in which the general problem of the *Critique* can be addressed. Sensibility and understanding must be fundamentally distinct cognitive capacities — each constituted a priori — if *synthetic* a priori judging is to be possible. At this point in the proceedings, we have some preliminary grasp of each element in its distinctness from the other. We now need to find the principle of their cooperation *without losing sight* of their heterogeneity.

This way of putting the problem of the Deduction is entirely internal to Kant's particular way of proceeding in the *Critique*. Another perspective on the problem is expressed as the Deduction gets under way. In Humean terms, the issue is our capacity to "represent [something] as combined in an object" (B130). Hume, the great purifier of empiricism, questioned our entitlement to suppose that we are able to represent an objectively necessary combination with the concept of cause. Our minds, Hume argues, are shaped by repetitions in circumstances as we trace a path through the world. This repetition forms tacit expectations, or anticipatory dispositions, which Hume calls "customs", and which are the "foundation of all of our judgments".<sup>30</sup> These anticipatory dispositions are formed passively, through some kind of mechanism of the associative imagination. With consistent empiricist methodology, he derived the concept of cause

<sup>&</sup>lt;sup>30</sup> *Treatise* I.iii.13, 147.

from the "subjective necessity" arising from custom — i.e., that one who is sufficiently conditioned *can't help* but to expect event of type B given event of type A. Missing from his account is the spontaneity of the understanding. According to Kant, the account of our capacity to judge about objects, rather than merely associate perceptions according to custom, must be a story about our spontaneity. While Kant credits Hume for his sensitivity to a fundamental philosophical problem to which everyone else was blind, and even praises his methodological consistency (B127), he charges Hume for failing to account for our capacity to "represent something as combined in an object".<sup>31</sup>

Yet it is worth going beyond this cartoon-like presentation of Hume's account, since the further details will allow us to appreciate better Kant's own strategy. Now, in claiming that custom is the foundation of all of our judgments, Hume says in effect that we share the *foundation* of our cognitive lives with the beasts. Of course, Hume recognizes that self-consciousness distinguishes us: we can "reflect" upon our customs. Our customs come into view for us when they are derailed by an apparently anomalous encounter. This derailment is an occasion for reflection: for becoming explicitly aware of the way in which we have been repetitively determined, and, in turn, for refining and correcting our expectations for the future. Thus, one who "has become a drunkard by the use of red wines" will, in virtue of the associative imagination alone, tacitly associate the color of his drink with its inebriating effect. It is only when he discovers that "he will be

<sup>&</sup>lt;sup>31</sup> In §19 of the Deduction, Kant distinguishes between objectively and subjectively valid relations of representations; in the former case, the "representations are combined in the object" (B142). In the *Prolegomena*, Kant indicates quite clearly that his work began with a generalization of Hume's skepticism about causality, to a complete account of the pure concepts of the understanding. Without such concepts, we cannot represent things as "combined in an object". This phrase as it appears first in §15 of the Deduction already alludes to his generalization of the Humean problem about the possibility of pure concepts, which (as Kant had said earlier on) "relate to objects a priori [...] as actions [Handlungen] of pure thought" (A57/B81).

carried almost with the same violence to white" that his anticipations are derailed. This provides the occasion to recognize that his anticipations have been (as it were) *unjustifiably* informed by the association of a certain red color of his drink with his drunken state. The red color is a "superfluous" — rather than an "essential" — circumstance of the repetition that has shaped his expectations. Superfluous circumstances excite the imagination as well as essential ones do; and it is only through reflection that we may "correct this propensity" to be confused by such "complication of circumstances".<sup>32</sup> As Hume uses the term, "reflection" is nearly always associated with correction.<sup>33</sup>

Reflection is an afterthought in Hume's account. What distinguishes us from the beasts is not the foundation of our cognitive lives, but rather something that is added on top of that foundation: namely, our capacity — and propensity — to reflect. Custom is associated with prejudice, i.e., believing and acting without being open to evidence that runs against the grains of one's habits and anticipations. Reflection is associated with believing and acting as a rational creature. An individual who is not open to the lessons that a frustration of his anticipations has to offer is a kind of human brute. Kant remarked on this very connection between prejudice and custom: custom is the "mechanism of reason", and not the "spontaneity" of it (*Logic* §IX, 9:76).<sup>34</sup>

<sup>&</sup>lt;sup>32</sup> *Treatise* I.iii.13, 147-8.

<sup>&</sup>lt;sup>33</sup> Some of the passages that connect reflection with correction in the first book of the *Treatise* are the following: I.ii.4, p. 47 and p. 48; I.ii.5, p. 57 (where "reflection" has a particularly ordinary sense); I.iii.13, p. 148; and I.iv.6, p. 254. For more on the role of reflection in the *Treatise*, see Annette Baier, *A Progress of Sentiments*. On the importance of Hume's account of reflection for moral theory, see Christine Korsgaard, *The Sources of Normativity*.

<sup>&</sup>lt;sup>34</sup> See also *Logik Phillipi* (24:425), and *Logik Pölitz* (24:547-8). For more on the same theme, especially regarding the connection between spontaneity and being human, see Kant's 1784 essay, "What is Enlightenment?" (8: 35-42).

Kant's first appeal to the notion of spontaneity in the Deduction connects it with the normativity of action: combination is an "action of the understanding", a *Handlung*. His initial point seems to be that "representing something as combined in an object" implies that we can be held responsible for the combination in question. As putative "knowers" we are always already subject to imputation — and not only when our tacit anticipations, our customs, fail us. So he says: "we cannot represent anything as combined in an object without having previously combined it ourselves" (B130). The remark is programmatic. The plan is to show that the exercise of any cognitive faculty be it receptive or spontaneous — must always already be *reflective*. The "critical" solution, as I have suggested in earlier chapters, is for reason to make its own reflective activity explicit as the source of theoretical cognition in general.

## 8.

The first section of the Deduction (§15) is a reminder of context: the problem that the Deduction must address can be traced back to the presentation of the heterogeneity thesis. In other words, the first words of the Deduction should put us in mind of what went on in the *Critique*'s front matter, which included the preliminary conception of reason as the "idea of the whole" from which the *Critique* sets out. This preliminary account suggests that our capacity for scientific cognition is rooted in some kind of selfconsciousness that is attributed to reason. As I argued in Chapter 3, Kant wants us to think of reason as the reflective aspect of our cognitive capacity, and he wants us to recognize that it cannot meaningfully be separated from the judging aspect of our cognitive capacity (the understanding). Now, the application of Kant's general model of scientific method suggests that the first part of the Deduction should involve an analytic "ascent" to the first principle of our science. Given that Kant thought of the *Critique* as the science of "an a priori judging reason", we should expect that this analysis will set out from some preliminary conception of reason.

This, I take it, accounts for the fact that the actual argument of the Deduction begins in the next section (\$16) with the theme of reflection or self-consciousness striking the first note: "The *I think* must be able to accompany all of my representations" (B131). The argument sets out with the preliminary conception of reason as a reflective or self-conscious capacity. The crucial moves of the analysis are carried out in the wake of this remark. Some commentators suppose that the analysis that follows is an unpacking of this *as* the "first principle" that we are looking for.<sup>35</sup> That cannot be quite right, since if this first statement *were* the "first" or "highest" principle of the science, then it should admit of no further analysis. This first remark might be thought of as a rough-and-ready statement of the principle that we are after; only we cannot yet recognize it *as* a principle. It is simply laid before us on the basis of the preliminary conception of reason that was put in place earlier on. In the analysis, we consider what its implications would be given the general framework of critical philosophy.

Generally speaking, analysis begins with a given proposition, which is affirmed or denied on the basis of considering its negation. If the alternative is unintelligible, the original proposition stands. After the original proposition is put forward at the start of §16, the alternative is tacitly introduced: that the *I think* were not able to accompany all

<sup>&</sup>lt;sup>35</sup> E.g., Allison ("Reflections on the B-Deduction").

of my representations. This alternative is found to be unintelligible: "For otherwise something would be represented in me which could not be thought at all, which is as much as to say that the representation would either be impossible, or at least nothing for me" (B131-2). But what sort of unintelligibility is this? For it is surely not the case that the ability to say "I think" — or the faculty of self-consciousness — is a condition of the possibility of representing as such. Non-rational animals have a faculty of perception, and so have representations, but they cannot say "I think" in conjunction with them.

The "I" who utters this is pure theoretical reason, the agent and subject of the critical self-investigation. The principle initially formulated at the top of §16 is a principle of the possibility of representations belonging or mattering to the subject of the critical investigation. The mere consideration of what it would mean for representations to belong to such an intellectual capacity entails the rough-and-ready principle stated at the outset of §16.

The next move in the analysis momentarily narrows the scope of this principle, to consider its bearing on *intuitive* or sensible representation.<sup>36</sup> In short: we take the idea of a reflective, spontaneous capacity and consider the possibility of its "having" sensible representations. The ability of this reflective intellect to think of sensible representations as belonging or mattering to it requires the supposition that its relation to them consists in its own activity. This notionally points to the idea that its capacity to enjoy sensible representations rests on some kind of *synthesis*. At the end of §16, our rough-and-ready principle is dubbed the "principle of the necessary unity of apperception". Kant says that

<sup>&</sup>lt;sup>36</sup> Immediately after the initial statement of the principle, the argument introduces intuition as "[t]hat representation which can be given before all thinking" (B132). In the next move, the relation of the *I think* to intuitive representation is considered: "Thus, everything manifold in intuition has a necessary relation to the *I think* in the same subject in which this manifold is encountered" (B132).

it is an "analytic proposition", which nevertheless "declares a synthesis of the manifold given in intuition as necessary, without which this thoroughgoing identity of selfconsciousness cannot be thought" (B135).<sup>37</sup> Here, too, an alternative proposition is ruled out: namely that this spontaneous, reflective capacity could somehow "have" sensible representations, but not in virtue of any "synthesis". This would leave us having to suppose that the agent of our investigation would be the scarcely intelligible "intuiting intellect", which is abruptly introduced into the argument: "An understanding in which through self-consciousness everything manifold would be given would be an *intuiting* understanding; ours can only *think* and must seek intuition in the senses" (B135). The analysis is responsible to the evidence of self-recognition; and the opposing proposal is rejected. In sum, the second move in the analysis looks like this: the reflective capacity of reason is now considered as the principle of the possibility of cognitively significant *sensible* representation. When we consider a reflective intellect in relation to given

in its own activity. In other words, some "necessary synthesis" must make this relation possible. Without supposing this "necessary synthesis", we would be left supposing that the reflective capacity in question is an intuiting intellect. The recognition of this "necessary synthesis" effectively points to the *pure understanding* (by the lights of \$15).<sup>38</sup>

<sup>&</sup>lt;sup>37</sup> The "identity of self-consciousness" refers to the idea that the capacity to say "I think" unites all representations: it is "one and the same in all consciousness" (B132). This "identity" cannot be "thought" apart from the activity of the understanding. As Kant remarked at the outset of the Introduction, the initial stimulus for this activity comes from the senses (B1-2).

<sup>&</sup>lt;sup>38</sup> The argument does not suppose from the outset that this capacity to say "I think" is an activity of pure apperception. We first consider this unspecified reflective capacity as a principle of cognitively significant sensible representation, or intuition. We then recognize, on the basis of the heterogeneity thesis, that intuition can be given "before all thinking". Yet the *I think* must be a spontaneous representation ("it cannot be considered as belonging to sensibility", B132). So, Kant clarifies this idea of the capacity to say

The Deduction thus begins by considering the unity of intuition — only *not* in terms of the pure forms of intuition, but rather in terms of the principle of apperception. The result is the following claim: the complexity of what presents itself to one in intuition rests on a certain unity, and this unity is conceivable in terms of the activity of a discursive or finite intellect. The conscious enjoyment of given representations is perception. On the empiricist conception, perception is attending to what presents itself as given: consciousness is viewed as a kind of accompaniment to the given.<sup>39</sup> But Kant is trying to put in place a different conception of perception, where consciousness is not an accompaniment but the very means by which representations are given. Consciousness of given representations "does not yet come about by my accompanying each representation with consciousness, but rather by my *adding [hinzusetze*] and being conscious of the synthesis of them" (B133). *Our* perceptual consciousness, in other words, involves spontaneity.

9.

The idea that the intuition of a manifold rests on some "necessary synthesis" needs clarification. In regard to *what*, exactly, is this synthesis necessary? According to the analysis in §16, it is necessary in order that sensible representations may "belong" to pure theoretical reason. Yet the thought that sensible representations rest on some kind of synthesis must not forsake the idea that sensible representations are *given* representations. By the end of §17, we should have a better grasp of this difficult idea that lies at heart of Kant's formal idealism.

<sup>&</sup>quot;I think": we are talking about "*pure* or *original apperception*", and not empirical apperception. Empirical apperception would have to follow intuition, and so could not be the basis of it.

The basic outline of the solution looks like this: we draw a distinction between the nature of intuition as a merely given mode of representation and the nature of intuition as the means by which objects may be thought. The distinction is supposed to allow us to acknowledge that sensible representation rests on a necessary synthesis without compromising the idea that it is given. Kant lays these two principles of the "possibility of all intuition" side by side at the outset of §17.40 The first principle was identified in the Aesthetic, and it says that "all the manifold of sensibility stands under the formal conditions of space and time" (B136). This principle is supposed to be strictly internal to sensibility: it is merely a principle of the possibility of given representation, and not a principle of the possibility of representing *objects* given in intuition. The second principle considers sensibility in so far as it is determined by the ends of the understanding: it is the principle of the possibility of all intuition in so far as objects can be "thought or cognized by means of [intuitions]" (B137). It is not internal to sensibility.<sup>41</sup> It is already familiar from §16: it says that "all the manifold of intuition" stands under the conditions of the original-synthetic unity of apperception".

The two principles laid side by side at the top of §17 allow us to separate the aspect of sensibility that is a requirement for objectively valid combination (that "all the manifold of intuition stands under the conditions of the original-synthetic unity of apperception") from the aspect of sensibility that is not (that "all the manifold of sensibility stands under the formal conditions of space and time"). We can understand

<sup>&</sup>lt;sup>39</sup> See, e.g., Locke's account of perception as "taking notice" of impressions (*Essay* II.ix.3).

<sup>&</sup>lt;sup>40</sup> "Under the first stands all manifold representations of intuition, in so far as they are *given* to us, under the second in so far as they must be able to be *combined* in one consciousness [...]" (B137).

<sup>&</sup>lt;sup>41</sup> The one is the "highest principle of the possibility of all intuition in relation to sensibility"; the other is the "highest principle of [the possibility of all intuition] in relation to the understanding" (B136).

these as formal and material conditions *of combination*. The material condition — that the combinable manifold is given in *space and time* — is contingent with respect to the requirements of an "a priori judging reason". The formal condition is not.<sup>42</sup> The principle of the synthetic unity of apperception indicates what is to be included about the possibility of given representations in the general idea of an "a priori judging reason".

We have found a principle that promises to complete the prior provisional accounts of sensibility and understanding, by enabling us to consider them as necessarily cooperating with each other. At the start of \$17, this principle is indicated as that on the basis of which intuition could be an objectively valid mode of representation. Thus, this principle could conceivably complete the account of sensibility as a source of cognition in virtue of its constitutive representations. Moreover, at this point in the Deduction, the operative conception of the understanding suddenly shifts: while it was previously a "faculty to combine a priori" (\$15, *passim*; \$16, B135), it is now "the faculty of *cognitions*" (\$17, B137).<sup>43</sup> This shift in the operative conception of the understanding understanding comes hand in hand with the discovery of the principle of the synthetic unity of apperception.

The line of analysis found in §17 is a bit harder to fathom than that found in §16. It begins with the idea that the understanding is "the faculty of cognitions". Straightaway

<sup>&</sup>lt;sup>42</sup> N.B.: the notions of "formal" and "material" conditions are *relative*. The representation of space and time is a formal condition of *human sensible intuition*, but a material condition of *combination*. Formal conditions are "conditions of the possibility" of *x*; material conditions are "conditions of the actuality" of *x*. The representation of space and time is a formal condition of human sensible representation, because whatever may register as an appearance to us must have some extension in space or duration in time. The idea that the representation of space and time is a material condition of combination means that *our* activities of combination are "realized" as determinations of appearances in space and time. Nothing rules out the possibility that a material condition could obtain a priori. It is important to Kant's argument in the second half of the Deduction that the representation of time figures as an a priori material condition of the synthesis of the pure understanding.

it defines cognition as "the determinate relation of given representations to an object", with the proviso that a *concept* is a rule for this determination (B137). The "unity of consciousness", Kant continues, "is that which alone constitutes the relation of representations to an object, and thus their objective validity" (B137). These remarks leave open the possibility that both intuitions and concepts could be objective representations. For the notion of a "unity of consciousness" is generic: intuitions and concepts both involve a "unity of consciousness", each in a different way. In an intuition, "many representations are found to be contained in one representation, and in the consciousness of it"; and in a concept, "one and the same consciousness is found to be contained in many representations" (B136). The overall strategy is to show that both modes of representation owe their characteristic unity of consciousness to the same principle.

At this point, Kant presupposes an argument from a footnote in §16, which showed that the unity of consciousness belonging to a concept has its basis in the unity of consciousness expressed in a judgment. A concept expresses an "analytic unity of consciousness", by means of which a "feature" is represented "which (as a mark) can be encountered in anything or can be combined with other representations" (B133n.). The ability to consider separately some such feature rests on an ability — be it exercised or not — to combine that feature with other representations, "even if only possible representations", Kant adds (B134n.). The content of concepts rests on the conditions of their employment. Thus, Kant claims that the synthetic unity of consciousness in a judgment is prior to the analytic unity of consciousness in a concept, glossing his earlier

<sup>&</sup>lt;sup>43</sup> I.e.: prior to §17, no claim is made either for the objective validity of its combination, or for its status as a source of cognition.

remark that concepts are "predicates of possible judgments" (A69/B94). The grammar of "potentiality"-talk alone could yield this reading: if the content of a concept can only be understood in terms of its employment in a possible judgment, then this analytic unity of consciousness rests on the synthetic unity of consciousness expressed in a judgment.

In §17, Kant focuses on what the very idea of "consciousness of an intuition" entails. The answer is embedded in the second principle that was indicated at the top of §17: that "all the manifold of intuition stands under the conditions of the originalsynthetic unity of apperception". According to this principle, if our "consciousness" of an intuition is unified, then it must be able to be determined with respect to a rule of synthesis. In a judgment, we draw attention to a particular way in which our consciousness of an intuition is unified through the employment of a concept. Kant tries to make this vivid with a slightly misleading example:

[I]n order to cognize something in space, for example a line, I must *draw* it, and thus synthetically bring about a determinate combination of the given manifold, so that the unity of this action is at the same time the unity of consciousness (in the concept of a line) and thereby an object (a determinate space) is first cognized. (B137-8)

In the example, the concept of a line figures as a rule for the construction of an object (a line). But concepts are employed in empirical cognition as rules for the determination of *given* (not constructed) objects. We could gloss this conception of judgment, taking a cue from Kant's geometrical example. In an intuition, a multiplicity is given that is ready to be determined in judgment. The determination could be likened to tracing the boundaries of objects that are given in intuition.

This does not mean that intuition could "contain" a synthesis — i.e., a ready-made unification of representations marking off the boundaries of an object. This would leave us needing to tell a story about how this ready-made synthesis becomes the property, so to speak, of the understanding: in other words, we would still need to make knowledge out of it, or *establish* its objective validity. Kant does not think that it is possible to tell such a story: this is the upshot of his response to Hume.<sup>44</sup> In its broadest overview, the Deduction aims to show that the possibility of objective representation can be addressed only if the heterogeneity thesis is maintained. This sort of "borrowing" of ready-made syntheses would clearly be incompatible with the heterogeneity thesis. Kant supposes that a multiplicity or "manifold" of representations can be given in an intuition, which could be determined in a range of possible ways, through the employment of a concept in a judgment. The distinction drawn at the top of §17 — the two distinct principles of the "possibility of all intuition" — is supposed to allow for this idea without undermining the heterogeneity thesis.

10.

At this point, the basic moves of the analytic ascent are in place: we have uncovered the first principle of pure theoretical reason's critical self-investigation, and we have accounted for its status *as* that principle, since it pertains to the unity of consciousness that makes objective representation (whether as intuition or judgment)

<sup>&</sup>lt;sup>44</sup> With Hume clearly on his mind, Kant colorfully remarks: "[C]ombination does not lie in the objects, and cannot as it were be borrowed from them through perception and by that means first taken up into the understanding" (§16, B134; see also §27, B166).

possible.<sup>45</sup> The signpost in §21 marks the end of this analysis; with the principle in hand, the synthesis can begin. Before turning to the synthesis, let us verify that the signpost is telling us what we think that it should be telling us. The deduction proper has begun "in the above proposition", and that proposition is:

A manifold that is contained in an intuition that I call mine is represented as belonging to the *necessary* unity of self-consciousness through the synthesis of the understanding, and this takes place by means of the category. This indicates, therefore, that the empirical consciousness of a given manifold of one intuition [Einer Anschauung] stands under a pure a priori self-consciousness, as empirical intuitions stand under a pure sensible intuition, which likewise holds a priori. (B144)

The first sentence presents the principle, recapitulating the manner in which it was revealed. The second sentence considers the empirical employment of that principle: although any actual sensible representation can only be the object of an *empirical* consciousness, this empirical consciousness is made possible by a "pure a priori self-consciousness". The analytic principle, we have already seen, points toward a "necessary synthesis"; the principles of this synthesis — expressed as the principles of the pure understanding, which are developed from the categories — are the principles of the possibility of empirical consciousness as such.

The general point here could be expressed as follows. In Kant's story, perception — at least in so far as it can contribute to our cognitive endeavors — involves spontaneity as well. This is indicated already in the idea that intuition "rests on" a necessary synthesis. We could interpret this as meaning that our cognitive lives are spontaneous through and through, and not only in one aspect. If we deny this, then we

<sup>&</sup>lt;sup>45</sup> This takes us through §19. (To a large extent, §18 previews what is to come in the second half of the Deduction.) §20 is recapitulation of the argument leading up to the signpost, and the signpost is found at

would suppose that our cognition can be — at least at times, or in certain respects — merely *mechanical*. It is commonly thought that perception is "mechanical" in a way that judging is not. Kant would say that this is false. In fact, on Kant's conception, what we *take* to be judging is often only mechanical and not spontaneous: this is what Kant has in mind when he opposes the "mechanism of reason" to the "spontaneity of reason",<sup>46</sup> and when he ridicules the good student of Wolffian metaphysics who "has grasped and preserved well [...] and is a plaster cast of a living human being" (A836/B864). Cognition — whether in perception or judgment — involves spontaneity. This claim does not run afoul of the heterogeneity thesis: for nothing in the mere "having" of an intuition determines the shape of our consciousness of it. The determination of our consciousness of an intuition involves a moment of freedom, whether this determination is expressed in a judgment or is merely manifest in our attentive perception.

The general model of scientific investigation that was sketched above tells us that the synthesis should employ the principle that was uncovered in the analysis. That principle says that sensible representations are potentially cognitively significant in virtue of some "necessary synthesis" that can be attributed to the spontaneity of our cognitive capacity. What, exactly, does this principle *do* in the argument of the Deduction? In the broadest terms, it sets a shared end for sensibility and understanding. This is important, because the strategy of isolation led us to consider each element in so far as it is determined by its own ('unshared') end. The Aesthetic left us with a conception of sensibility as a "faculty to intuit a priori", constituted by space and time; and the passages in the Analytic leading up to the principle present us with a conception of the

the start of §21.

understanding as a spontaneous capacity to *combine* representations, one that is constituted by the categories as mere forms of thought. Yet what we want is an account of their *unity* as elements of a single cognitive capacity. This account should establish that the characteristic modes of combination that are provisionally attributed to the understanding in laying out the table of categories are not arbitrary, but necessarily valid a priori of appearances in general. This will complete our account of the pure understanding as a cognitive capacity. At the same time, this synthetic argument will provide the context in which we can finally appreciate the Metaphysical Exposition of space and time as an account of sensibility as a cognitive capacity. In this respect, it will complete our account of sensibility as well.

11.

The solution to the problem of the Transcendental Deduction rests on our seeing how space and time — as the particular principles of our sensibility, our capacity to enjoy given representations — in some respect presuppose the spontaneity of the understanding. This argument must not threaten the independence of sensibility. We are meant never to lose sight of the heterogeneity thesis — *the* thesis that distinguishes critical philosophy from its empiricist and rationalist predecessors alike. The solution, though, is reached only after a long argument that culminates in §26.

At the beginning of the synthetic argument (§§22-23), Kant indicates that the account of the pure understanding was not completed in the first half of the Deduction. As long as we aim to establish its capacity to judge a priori about objects that can only be

<sup>&</sup>lt;sup>46</sup> For the distinction between the "mechanism" and the "spontaneity" of reason see §7, above.

given in experience, then we need to bring the conditions of sensible intuition into full consideration. For reasons that are not made entirely clear, Kant holds that the categories may extend "beyond our sensible intuition" (§23, B148). The thought behind this seems to be Kant's idea that the understanding is *spontaneous*, or self-determined. This is to say that nothing outside of it determines its capacity; it is not 'what it is' owing to its relation to sensibility. But, Kant remarks, this "further extension of concepts beyond our sensible intuition does not help us at all". In other words, we need to register that the pure understanding is restricted in its employment by sensibility, in order to hold on to the crucial requirement that objects can only be given to the understanding from outside of it. According to Kant's usage, a limit restricts something in a negative way, meaning that it does not belong to the thing that it restricts.<sup>47</sup> So the particular conditions of our sensible intuition are allowed to figure as a *limiting* condition on the employment of the understanding. Registering this limitation is necessary in order to "determine the boundaries of the employment of the pure understanding in regard to objects" (§23, B148) — i.e., it is necessary in order to give a positive account of the pure understanding as a *cognitive* capacity.<sup>48</sup> Hence, the particular conditions of our sensibility are introduced into the argument of the Deduction for the first time. The categories are only sources of cognition to the extent that they are applicable to *empirical* intuition, and so to objects as they can actually be given to us in space and time. (For a discussion of these issues as they figure in the Amphiboly, see Appendix A.)

<sup>&</sup>lt;sup>47</sup> This can be inferred from Kant's discussion of the distinction between "boundary" and "limitation" in the *Prolegomena* (§59, 361).

<sup>&</sup>lt;sup>48</sup> This notion of determining "boundaries" is metaphorical in Kant's usage for giving a *complete* account of a cognitive capacity. The metaphor informs Kant's comparison of the true critical project with Hume's "censorship" of pure reason: only through genuine criticism can the *boundaries* of reason be determined (see A758/B788 ff.).
According to Kant's general conception of scientific method, the synthetic argument is supposed to account for the relation of the elements to one another. The first stage of the argument after the signpost concerns the role of sensibility as a "limiting condition" on the employment of the pure understanding. Presumably, then, this counts as the beginning of the synthetic argument; at any rate, we are beginning to consider the relation of the elements to one another (specifically, the "limiting" relation of sensibility vis-à-vis understanding). Admittedly, the principle of apperception does not seem to be playing any real role in establishing this point: we simply recognize that the first half of the Deduction had not been able to account for the pure understanding, we would need to register certain conditions on its employment that come from outside of it. Hence, sensibility is admitted as a *limiting* condition.

The second stage of the synthetic argument concerns the idea that the understanding *determines* sensibility a priori. The notion of "determination" indicates that the understanding would — at least in some respect — make sensibility 'what it is'. Obviously this idea needs to be handled carefully: for, at least at first blush, it is hard to see how Kant can say this without running afoul of the heterogeneity thesis.

Kant pursues the thesis that the understanding determines sensibility a priori under the banner of his doctrine of "figurative synthesis". (See Figure 4.) As he introduces this "figurative synthesis", Kant harks back to the account of synthesis as it was presented in the analytic first half of the Deduction. At that point, we were able to consider the constitutive concepts of the understanding only in relation to "objects of intuition in general, without it being determined whether this [intuition] be ours or any other, provided only that it is sensible" (§24, B150). For this reason, Kant says that the synthesis that was indicated in the analysis was "purely intellectual". Sensibility figured only notionally at that point, merely as the source of representations that are the necessary means by which the understanding may judge about objects.

It is important to see that these "intellectual" and "figurative" syntheses need not be regarded as different syntheses at all — i.e., different activities that each accomplish different results, and may be conducted together in some kind of grand orchestra of cognition. The distinction is better understood in terms of the idea that we are now in a position to take a different perspective on the synthesis that was already at issue in the first half of the Deduction. We have won this vantage point by acknowledging sensibility's limitation on the employment of the pure understanding. With that in place, we are able to consider the understanding as a capacity to judge a priori about *objects that can only be given as appearances*, and this allows us to consider its characteristic synthesis in a different light. At any rate, this is precisely how Kant begins his account of figurative synthesis, with a reminder of the limiting condition:

But since there lies in us a certain form of sensible intuition a priori, which rests on the receptivity of the capacity of representation<sup>49</sup> [...] (B150)

So the account of figurative synthesis seems to rely on our prior identification of an *independent* claim about the a priori constitution of sensibility. Why must this be in place in order to reconsider the characteristic synthesis of the understanding as a "figurative synthesis"? To answer this, let us see how this sentence continues:

<sup>&</sup>lt;sup>49</sup> The phrase is *Rezeptivität der Vorstellungsfähigkeit*. Kant generally avoids using the term "faculty" (*Vermögen*) for sensibility, preferring "capability" (*Fähigkeit*). For an account of Kant's usage, see Appendix B.

[... so] the understanding, as spontaneity, can determine inner sense through the manifold of given representations in accordance with the synthetic unity of apperception, and thus think a priori synthetic unity of the apperception of the manifold of *sensible* intuition, as the condition under which all objects of our (human) intuition must necessarily stand, through which then the categories, as mere forms of thought, acquire objective reality [...] (B150)

This can hardly be read as an argument establishing the doctrine of figurative synthesis; it is best read as a précis of the rest of the synthetic argument. In order to see how that argument goes, it is important to understand better what is at stake.

This "figurative synthesis" is supposed to be a *spontaneous* synthesis. In this respect, it is to be distinguished from a Humean synthesis in which perceptions are merely associated according to empirical rules that have their basis in custom. The Humean synthesis is a "reproductive" synthesis, that cannot be attributed to the spontaneity of our cognitive capacity; in other words, this synthesis is not thought to determine the form of what may actually appear to us in the first place. Through the doctrine of figurative synthesis, Kant means to establish the spontaneous basis of whatever may appear to us at all.

In the Deduction, Kant distinguishes between an empirical and a transcendental figurative synthesis. The empirical figurative synthesis is the synthesis involved in perception; Kant calls it "synthesis of apprehension". We might think of it as the synthesis involved in our concretely taking up the world in perception. In the Deduction, Kant looks for what he might call the "transcendental ground" of this synthesis: that is the transcendental figurative synthesis, which I will explain shortly. Once he establishes this transcendental figurative synthesis — in §26 — Kant takes himself to have legitimated the very idea of figurative synthesis: namely, that the pure understanding can

determine sensibility a priori. The establishment of this doctrine underwrites the work of the Analytic of Principles, which focuses on the specific ways in which the pure understanding determines *inner sense* (i.e., through determinations of *time*) by means of the categories. I will not cover the transition from the Deduction to the Analytic of Principles in this dissertation: it is, however, the natural next thing to do, since this would fill out the story of "reflective metaphysics" that I began in Chapter 2. But let us now consider how Kant establishes the doctrine of figurative synthesis in §26.

The crucial move of the synthetic argument — the one that is supposed to establish the doctrine of figurative synthesis — only comes at the very end of the Deduction. At this point, Kant relies on a previously neglected implication of the heterogeneity thesis that figured in the Aesthetic's account of space and time as the constitutive representations of sensibility. This implication of the heterogeneity thesis is that space and time are *themselves* intuitions — i.e., they are unified, singular representations in their own right. It is not enough to say that space and time are the "forms" by which sensation (or the "matter of appearance") is necessarily ordered; we have to recognize that these forms are proper to our receptive capacity, and this means that space and time are *themselves* intuitions (rather than concepts). If we fail to say this, we leave open the possibility that these "forms" could ultimately be reduced to representations that are proper to the intellect. So, if space and time are themselves intuitions, then they must stand under the principle of the synthetic unity of apperception — for this principle, as we saw, is supposed to hold of cognitively significant representations in general.<sup>50</sup>

<sup>&</sup>lt;sup>50</sup> As Kant points out, geometry (at least) is the science of space itself (as a "formal intuition", B160n.).

The upshot is that the representation of space and time themselves — as unified, singular representations — would have to rest on a "necessary synthesis". If this move works, then the result is that the *very means* by which representations are given is subject to the principle of the synthetic unity of apperception. Since Kant shows (in §§18-19), that the conditions of this synthetic unity are expressed in the categories, he is then able to conclude that the categories are valid a priori of "whatever objects *may come before our senses*" (§26, B159). This is the "transcendental" figurative synthesis: we are being told that the nature of our capacity to intuit a priori cannot be determined in isolation from the pure understanding. Rather, it owes something to its relation to the pure understanding, in so far as this relation is established on the basis of the principle of the synthetic unity of apperception.

In simply appealing to the distinction between receptivity and spontaneity as characterizing two necessary aspects of our cognitive capacity, Kant is committed to establishing the priority of spontaneity over receptivity. In other words, if our cognitive capacity is "spontaneous" in some respect, then it is self-determined in this respect. And that means that the spontaneous aspect of our cognitive capacity is prior to — undetermined by — the receptive aspect of our cognitive capacity. But this raises the following crucial question: *is this heterogeneity thesis internally stable?* What happens when Kant tries to follow the appeal to spontaneity through to its full consequences? Do we still have sensibility figuring as an independent element?

In order to understand how Kant tries to avoid the difficulty that I have just raised, we need to recognize that Kant's solution rests on our having attained different perspectives at different stages of the *Critique*. The perspective that we attain in the Transcendental Deduction is different from the perspective that we had back in the Transcendental Aesthetic. The Aesthetic lies entirely outside of the Transcendental Logic, and for this reason Kant takes its results to be quite independent of whatever goes on under the rubric of transcendental logic. More precisely, Kant's solution relies heavily on the fact that the two principles of the "possibility of all intuition" were identified independently of one another — one under the auspices of the strategy of isolation, and the other not. The center of gravity of the Transcendental Deduction and indeed of the whole project of reason's self-knowledge — is the principle of the synthetic unity of apperception. This principle sets a shared end for sensibility and understanding: i.e., judging. But it does not follow from the principle of apperception that sensibility must be formed by space and time *in particular*. The principle only presupposes that sensibility is formed *in one way or another*; and from the perspective that we have in the Transcendental Logic, the particular forms that are constitutive of sensibility can only be recognized as a contingent fact about the constitution of our (human) cognitive faculty (§21, B146).

Kant's account of transcendental figurative synthesis is supposed to entitle us to the idea that sensibility owes its a priori constitution to its relation to the understanding. That it is constituted by "these intuitions" — space and time in particular — does not fall within the scope of this determination (B161). The principle that is identified in the Aesthetic — that "all the manifold of intuition stands under the formal conditions of space and time" — is not reduced to the principle of apperception. Hence, Kant thinks that sensibility retains its status as an irreducible source of cognition, even though it is essentially unified with the pure understanding.

With this, the synthetic argument draws to a close. Working from the principle that is discovered in the analysis (the principle of apperception), Kant has provided an account of sensibility and understanding as two necessarily unified elements of a single cognitive faculty. Through this argument, Kant concludes that the categories are necessarily valid a priori of whatever objects may come before our senses. This is meant to be the crucial move of the *Critique*, the demonstration of the capacity of the understanding to determine the possibility of appearances a priori through the categories; this establishes our capacity to judge synthetically and a priori, which is Kant's reformulation of the problem of metaphysics. At this point, Kant's argument makes contact with a conception of nature as the "totality of all appearances", and Kant takes himself to be in a position to conceive of his result in terms of the understanding's "prescribing the law to nature and even making it possible" (B159). So the *shape* of Kant's solution to the problem of metaphysics comes down to the idea that nature can be conceived in terms of a relation between sensibility and understanding that is established on the basis of a rational principle. That principle is the one revealed in the first half of the Transcendental Deduction. This is why that principle is the "beginning".

#### III.

12.

The idea that the *Critique* follows a synthetic method means that we start with an idea of the whole. It is to be the investigation that an "a priori judging reason" carries out with respect to its own capacity. Specifically at stake is reason's capacity to judge a

priori about objects that can only be given in experience. This is presented as a generalization of the historically specific problem that Kant confronted: the failure of metaphysics. The *Critique* is offered as a response to this "general problem of pure reason". Since the problem is rooted in the nature of reason as such, the only solution to it is for reason to come to some kind of self-knowledge that would not simply describe its conflicted state, but bring about its internal harmony. Kant proceeds with the presupposition that a solution to the "general problem of pure reason" is in the offing, presumably on the basis of the existence of thriving examples of synthetic a priori cognition, like mathematics.

Given these parameters, Kant presupposes that sensibility and understanding would each have to figure as heterogeneous elements of pure theoretical reason. In turn, this would entail that each element is constituted a priori: the mode of representation characteristic to the one element can never be reduced to the mode of representation characteristic to the other. This underwrites the strategy of isolation, in which we consider each element in isolation from the other and also from everything empirical. This strategy allows for the particular representations that constitute each capacity to come into consideration. In the *Critique*, this happens in the Metaphysical Expositions of space and time, and the so-called Metaphysical Deduction of the categories.

What cannot be addressed in the moment of isolation is the status of sensibility and understanding as sources of *cognition*. This is why we need the Transcendental Deduction. The status of the claim that sensibility and understanding are, in virtue of the distinct a priori constitution of each, the "sources" of all of our cognition — that they are *cognitive capacities* — is provisional until quite late in the Deduction. The first half of the Deduction is an analytic ascent to the a priori principle of the necessary cooperation of sensibility and understanding. This principle — called the "principle of the synthetic unity of apperception" — is a rational principle in the thickest sense: it pertains to the reflective activity that is characteristic of reason. All forms of rational cognition turn on reason's self-consciousness — its tacit awareness of its contribution to the cognition in question. The *Critique* contains a very special form of rational cognition, since in it reason is expressly concerned with its own theoretical capacity. In the *Critique*, the reflective basis of theoretical cognition in general is made explicit in terms of the principle of the synthetic unity of apperception. Thus, this principle is at once the highest principle of the critical science, and the highest principle of theoretical cognition in general.

At the end of the day, the "preliminary" mentioned at the end of the *Critique*'s Introduction really is the claim on which the rest of the account rests. We are meant never to lose sight of the heterogeneity thesis. And Kant thinks that we need to interpret the heterogeneity thesis in his radical way if we want to resolve the "general problem of pure reason". The *Critique* is not offered to one who denies, or is thoroughly skeptical about, the possibility of synthetic a priori judging. He who does not accept Kant's account of the "general problem of pure reason", or who does not agree that there is any problem there at all, cannot take the first step on the critical path.<sup>51</sup> But he who does must accept the radical implications of the heterogeneity thesis, at least as a preliminary — if he also wants to suppose that the general problem admits of a solution. And if he

<sup>&</sup>lt;sup>51</sup> A remark in the Preface to the *Critique of Practical Reason* helps us to see the fundamental importance of recognizing that critical philosophy is reason's self-knowledge: "Nothing worse could happen to these labors than if someone should make the unexpected discovery that there is and can be no a priori cognition

accepts the analytic ascent to the principle of the synthetic unity of apperception, then I think he should then grant that this "preliminary" is indeed the basis of a "whole new science".

Yet a serious question remains, about whether Kant meaningfully maintains his cherished heterogeneity thesis through the end of the Deduction. Kant's defense of the heterogeneity thesis turns on the fact that the particular constitution of sensibility was uncovered separately, when we labored under the strategy of isolation. So we apparently have an "independent" account of sensibility, and this allows Kant to maintain the heterogeneity thesis even as he argues (in the thick of the Deduction) that our spontaneity does indeed, at least in some respect, account for the nature of sensibility. The principle of the synthetic unity of apperception is presented as the principle that can account for the possible cognitive significance of any representation. Hence it would account for the status of sensibility as a *cognitive* capacity — in so far as space and time themselves "stand under" this principle. The fact that we have an "independent" account of sensibility in the Aesthetic suggests that this principle, or our spontaneity, can only account for the bare idea that sensibility is "formed" *one way or another* — and not that sensibility is constituted by *space and time* in particular.

This solution threatens to be somewhat unsatisfying. At any rate, it is closely tied to certain particular facts about the method of the *Critique*. So is Kant's solution cosmetic, *merely* 'architectonic'? Perhaps Kant had too much faith in proper scientific methodology. It is tempting here to pull an early, and unpublished, remark of Kant's out of the drawer, and raise an eyebrow. For it is one thing to remark that method is "the

at all. But there is no danger of this. It would be tantamount to someone's wanting to prove through reason that there is no reason" (12).

most important concern of a science"; but something else to claim that "there is less at stake in whether the propositions of a pure philosophy are true or false about the object; it is more important that they are thought in the appropriate method".<sup>52</sup> But that would clearly be unfair to Kant, whose faith in methodology seems to come down to the more reasonable idea that false results follow upon inappropriate methodology, but not that appropriate methodology is a surefire guarantee of true results.

More philosophically promising, perhaps, is the angle on this problem that I entertained briefly above, in asking about the possible internal instability of Kant's heterogeneity thesis. What happens when we begin with the idea that one aspect of our cognitive capacity is "spontaneous", and the other "receptive", and try to explicate the full consequences of this view? Does the appeal to spontaneity itself ultimately undermine the heterogeneity thesis? It is unfortunate to conclude with more questions than answers. But these questions only clearly presented themselves to me as I tried to be a conscientious (yet surely not an *ideal*) reader of the *Critique*, 'thinking my way through a system that takes nothing as given except reason itself'. Perhaps the answers to these questions lie somewhere in Hegel's challenging response to Kant's critical philosophy. Uncovering those answers, however, is the work of another day — or another dissertation.

<sup>&</sup>lt;sup>52</sup> Unpublished written remark dated from the 1760s (Erdmann, *Reflexionen Kants zur Kritik der reinen Vernunft*, Nr. 183-4).

**CODA.** The Theme of Self-Knowledge in the Deduction

My suggestion that the Deduction be treated as the central text of reason's critical self-knowledge is corroborated by the fact that Kant discusses self-knowledge implicitly throughout the Deduction and explicitly in §25. His discussion there will allow us to briefly recapitulate our journey through the end of the Deduction. The very idea that the *Critique* is pure theoretical reason's self-knowledge means that *it* is the agent of the investigation — the one who says "I" and "me" and "my" throughout. Reason is "an intelligence" that is engaged in a project of self-determination; the investigation concerns its spontaneity (see B158n.). Someone who rejects the idea of cognitive spontaneity out of hand — e.g., a consistent empiricist like Hume — would be unable to take the first step down the critical path.

But Kant opens the *Critique* with a nod to the most cherished thesis of empiricism. "There is no doubt that all of our knowledge begins with experience" (B1). The remark is prominently placed at the beginning of the *Critique* to entice the broadminded empiricist to set off down its "thorny paths". Kant continues, acknowledging a crucial aspect of the empiricist conception of "reflection": our cognitive capacity can only be aroused by objects affecting the senses (B1), and so can only become aware of itself in regard to that stimulus. The anti-rationalist spirit of these remarks informs the entire conception of self-knowledge that the *Critique* aims to manifest. The *I think* of dogmatic rational psychology is not the mark of meaningful reflection, and it cannot yield self-knowledge. This point is stressed in the midst of the Deduction: Since *knowledge* of ourselves requires, in addition to the action of thinking that brings the manifold of every possible intuition to the unity of apperception, also a determinate sort of intuition through which this manifold is given, [I can say that] my own existence is not an appearance (still less a mere illusion). But the determination of my existence can take place only in accordance with the form of inner sense, according to the particular way in which the manifold that I combine is given in inner intuition. Therefore, I have accordingly no *knowledge* of myself *as I am*, but rather as I *appear* to myself. (§25, B157-8).

This remark may seem to tell against the very idea that pure theoretical reason could know itself at all. For the passage seems to indicate that the only mode of self-knowledge that is possible is that which is manifest in the empirical endeavors of anthropology and psychology. But this would be a misguided reading of the passage. The "I" speaking here is again the agent of the critical self-investigation, and it means to acknowledge the empiricist insight once again. The idea that reason is spontaneous, or self-determining, tells us something important about the idea that the *Critique* is reason's self-knowledge. This spontaneous "intelligence" can only determine its own faculty with regard to "the particular way in which the manifold that I combine is given in intuition".

The agent of the critical self-investigation continues its soliloquy, shifting somewhat awkwardly from the first to the third person:

[...] I exist as an intelligence that is conscious solely of its faculty of combination, but one that is subject to a limiting condition in regard to the manifold which it is to combine. This limiting condition, which it calls inner sense, can make that combination intuitable only according to relations of time, which lie entirely outside its own concepts of the understanding. (B159)

The anti-rationalist insight about the possibility of meaningful reflection need not entail the empiricist idea that self-consciousness must follow upon some actual sensory stimulation. Kant says here that we need only to acknowledge the general limiting condition of combination — its *material* condition — in order to speak legitimately of reason's self-knowledge. For this reason, the Transcendental Aesthetic must lie outside of the Transcendental Logic: if it didn't, we wouldn't get the "limiting condition" that we need in order for the project to manifest reason's self-*knowledge*. At the same time, we cannot appreciate how sensibility is necessarily part of "an a priori judging reason" until we are well along the thorny paths of the Deduction.

Nevertheless, we should be left with the following thought: critical selfknowledge is not to be understood on any introspective model, be it dogmatic-rationalist, skeptical-empiricist, or any mongrel thereof. The insight of critical philosophy — both substantive and methodological — is rooted in the idea that there is no privileged viewpoint outside of our own cognitive activity from which to assess it.

### Figure 3 Schematic Account of the Critique's Front Matter







# APPENDIX A THE CONNECTION BETWEEN THE NOTION OF A 'NOUMENON' AND THE APPEAL TO SPONTANEITY

At the beginning of the Amphiboly, Kant suggests that the term "object" admits of a double sense. The dominant use of the term, of course, refers to the objects that are "given" in sensible intuition (even if they can only be *known* by means of concepts): these are "phenomena". Still, we can entertain the thought of objects that are cognitively accessible by the pure understanding alone: these are "noumena" or "intelligibilia". Now, Kant holds that the representations of the pure understanding are those without which no objectively valid representation would be possible at all. Kant has just reminded us of this in Phenomena and Noumena (A253/B309), which immediately precedes the Amphiboly. And since the basis of objectively valid representation lies ultimately in the pure understanding, Kant himself cannot simply dismiss the possibility of objects accessible through the pure understanding alone.

It is only at the end of the Amphiboly that Kant clarifies the issue: the notion of a noumenon is admitted *problematically* (A286/B343). It is not a concept of an object, but "rather the problem, unavoidably connected with the limitation of our sensibility, of whether there may not be objects entirely exempt from the intuition of our sensibility" (A287/B344). The idea that sensibility figures as a mere limitation on the pure understanding, the concepts of which could conceivably extend "beyond *our* sensible intuition", cropped up already in the Deduction (§23, B148). In the Deduction, Kant says

Appendix A

nothing about why he even entertains this thought. In Chapter 4, I suggested that it simply falls out of the idea that the understanding is *spontaneous*, or self-determined: the appeal to spontaneity entails that the understanding is not 'what it is' owing to any influence from the senses. If we admit that the understanding is "spontaneous", then there is no intrinsic reason why its constitutive concepts — which, on this view, exhaustively account for the capacity for thought at all — could not extend beyond *our* sensible intuition.

If we admit that the pure concepts of the understanding conceivably extend beyond our sensible intuition, and also suppose that these concepts are the ultimate source of objectively valid representation, then we are poised to entertain the possibility of objects accessible by the pure understanding alone. The notion of the "noumenon" seems to come as part of the package with our appeal to the spontaneity of the understanding.

Critical philosophy is supposed to give us the requisite insight — and speculative constraint — to handle the appeal to spontaneity properly. The dogmatic metaphysician, Kant charges, lacks this insight. At the end of the Amphiboly, Kant traces Leibniz's error as a metaphysician to his appeal to spontaneity: to the idea that apperception and thought "precede all possible determinate arrangement of representations" (A289/B345). Leibniz fails to recognize the limiting role that sensibility must play if the determinations of thought are to yield cognitive access to objects at all. In his story, "the objects, i.e., possible intuitions, are made to conform themselves to concepts, but concepts are not made to conform themselves to possible intuitions" (A289/B345). The diagnosis: Leibniz engages in merely "logical" reflection, which is a comparison of representations

184

that completely abstracts from any consideration of the "cognitive power to which the given representations belong" (A262/B318), and this, Kant argues, cannot lead to any possible *objective* determination of objects. To do metaphysics, we need *transcendental* reflection — reflection "which goes to the objects themselves" (A263/B319).

## APPENDIX B SENSIBILITY AS A "FACULTY"

Kant avoids using the term "faculty" (Vermögen) for sensibility. In the Aesthetic, Kant says that sensibility is the "capability [Fähigkeit] (receptivity [Rezeptivität]) of receiving representations through the way in which we are affected by objects" (A19/B33). And at the outset of the Transcendental Logic, he names sensibility as one of the "two fundamental sources of the mind", without employing any "capacity"-talk at all: sensibility is simply the "receptivity of impressions", while the understanding is the "faculty to know an object by means of these representations (spontaneity of concepts)" (A50/B75). (This may be clearer in the German: "Unsere Erkenntnis entspringt aus zwei Grundquellen des Gemüts, deren die erste ist, die Vorstellungen zu empfangen (der Rezeptivität der Eindrücke), die zweite das Vermögen, durch diese Vorstellungen einen Gegenstand zu erkennen.") He refers to sensibility and understanding in the same way at the top of the next paragraph: sensibility is merely "the *receptivity* of our mind, to receive representations in so far as it is affected" (no "capacity"-talk in play), whereas the understanding is "the faculty to bring forth representations from itself, or the *spontaneity* of cognition" (A51/B75).

This usage suggests that there may be some connection between the notion of "faculty" (*Vermögen*) and "spontaneity". However, any possible connection between the two would have to be indirect. For it turns out that Kant does, on occasion, refer to sensibility — or rather, a capacity for a priori intuition — as a "faculty". And sensibility

cannot be spontaneous. If we were to suppose that it could be, we would fly in the face of the *Critique*'s steadfast thesis about the heterogeneity of sensibility and understanding. Thus, the occasional characterization of sensibility as a faculty does not indicate that it is spontaneous, or that there is any direct connection between the notions of "faculty" and "spontaneity" for Kant.

However, the occasional designation of sensibility as a faculty may have something to do with sensibility's role as a "stem" or "source" of cognition. On the page just following the remarks quoted above, Kant underscores the heterogeneity of sensibility and understanding in the following terms: "These faculties, or capabilities, cannot exchange their functions [Beide Vermögen, oder Fähigkeiten...]" (A51/B75). After all of the evident care Kant had taken both to avoid calling sensibility a "faculty" and to designate the understanding as a "faculty", he suddenly implies that the term could serve equally well for both.

This ambivalence can be explained if we consider Kant's idea of a "source" of cognition. At the very end of the *Critique*'s general Introduction, where the thesis about the heterogeneity of sensibility and understanding is first introduced, Kant claims that sensibility should only figure in the *Critique*'s Doctrine of Elements to the extent that it contains, or could be shown to contain, a priori representations (A15/B29-30). The context suggests that sensibility counts as a "source" of cognition, for Kant, if it contains a priori representations. These a priori representations are the pure forms of intuition.

Now, the very idea that sensibility could be a source of a receptive mode of representation may — and should — seem odd. For "receptivity" implies a capacity to receive something coming in from the outside, whereas a "source" flows from the inside

Appendix B

out. But this is just Kant's hylomorphism in play again. His hylomorphism can be understood modally. Sensibility is "actualized" when objects affect our senses. This actualization entails the presence of sensation, which is the "matter" for intuition. Nevertheless, any such actualization of sensibility presupposes the *forms* of intuition. These forms express the *capacity*, or the *faculty* of sensibility — i.e., sensibility as the source of our receptive mode of representation.<sup>1</sup>

Sensibility figures as a source of cognition only in so far as it contains these a priori representations. Late in the Aesthetic, after the hylomorphic distinction has been drawn between sensation as the "matter" and pure intuition as the "form" of sensibility, Kant finally speaks of "a faculty for intuiting a priori [ein Vermögen, a priori anzuschauen]" (A48/B65). He has just reminded us of geometry, and remarks that its characteristic claims could only be possible under the supposition that we possess such a faculty. This faculty, in other words, is a *source* of a priori cognition, and it is in this respect alone that sensibility counts as a "faculty".

<sup>&</sup>lt;sup>1</sup> Another way to draw the distinction, Kant remarks, is in terms of "higher" and "lower" cognitive faculties: sensibility is the lower, "because [it] gives the mere material [Stoff] for thought", while the understanding is the higher because it "disposes [disponiert] over this material and brings it under rules or concepts" (*Logic* §V, 36). This remark reminds us that "matter" and "form" are relative terms. Singular representations (intuitions) are the matter for thought, but sensation is the matter for intuition.

## BIBLIOGRAPHY

Bibliographical data on Kant's works are given in the Note on Sources and Abbreviations, above. Listed here are further writings that have been of use in writing this dissertation. It includes works from which I have quoted, works to which I have referred in the text and figures of the dissertation, and also works that I found particularly useful to have read in the course of working on the dissertation.

- Adelung, Johann Christoph. *Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart*. Wien: B.P. Bauer, 1808.
- Agich, George J. "L.W. Beck's Proposal of Meta-Critique and the 'Critique of Judgment'." *Kant-Studien* 74 (1983): 261-270.
- Allison, Henry. "Apperception and Analyticity in the B-Deduction." *Grazer Philosophische Studien* 44 (1993): 233-52.

------. "Kant's Concept of the Transcendental Object." Kant-Studien 59 (1968): 165-86.

- ——. "Kant's Refutation of Materialism." In *Idealism and Freedom*. Cambridge: Cambridge University Press, 1996.
- ——. Kant's Transcendental Idealism. New Haven: Yale University Press, 1983.
- ———. "On Naturalizing Kant's Transcendental Psychology." In *Idealism and Freedom*. Cambridge: Cambridge University Press, 1996.
- ———. "Reflections on the B-Deduction." *The Southern Journal of Philosophy* 25, Supplement (1986): 1-15.
- ——. "Things in Themselves, Noumena, and the Transcendental Object." *Dialectica* 32 (1978): 41-76.

- Ameriks, Karl. "The Critique of Metaphysics: Kant and Traditional Ontology." In *Cambridge Companion to Kant*, edited by Paul Guyer. Cambridge: Cambridge University Press, 1992.
  - ———. "Kant on Spontaneity: Some New Data." *Akten des Siebenten Internationalen Kant-Kongresses* (1990), vol. 2, edited by Gerhard Funke. Bouvier: Bonn and Berlin, 1991.
  - ———. *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*. Second edition. Oxford: Clarendon, 2000.

———. "Kant's Transcendental Deduction as a Regressive Argument." *Kant-Studien* 19 (1978): 273-87.

Aristotle. *The Complete Works*. Edited by Jonathan Barnes. In two volumes. Princeton: University Press, 1984.

———. *The Ethics of Aristotle*. (*Ethica Nicomachea*.) Edited by John Burnet. London: Methuen, 1900.

———. *Metaphysica*. Edited by W. Jaeger. Oxford Classical Texts. Oxford: Clarendon, 1957.

*——. Physica.* Edited by W.D. Ross. Oxford Classical Texts. Oxford: Clarendon, 1950. Translated by W. Charlton, in *A New Aristotle Reader*.

- Baum, Manfred. Deduktion und Beweis in Kants Transzendentalphilosophie: Untersuchungen zur "Kritik der reinen Vernunft". Königstein: Athenäum, 1986.

Baumgarten, Alexander Gottlieb. Metaphysik. Halle: Hemmerdeschen Buchhandlung, 1783.

- Baier, Annette. *A Progress of Sentiments: Reflections on Hume's "Treatise"*. Cambridge, Massachusetts and London: Harvard, 1991.
- Beck, Lewis White. "Five Concepts of Freedom in Kant." In *Philosophical Analysis and Reconstruction: Festschrift for Stephen Körner*, edited by T.J. Srzednicki. Dordrecht: Martinus Nijhoff, 1987, 35-51.

-. "Kant's Strategy." In *Essays on Kant and Hume*. New Haven and London: Yale University Press, 1978.

———. "Towards a Meta-Critique of Pure Reason." In *Essays on Kant and Hume*. New Haven and London: Yale University Press, 1978.

Behboud, Ali. "Greek Geometrical Analysis." Centaurus 37 (1994): 52-86.

Bell, David. "The Art of Judgment." Mind. New Series, 96 (1987): 221-44.

Buchdahl, Gerd. "Transcendental Reduction: A Concept for the Interpretation of Kant's Critical Method." *Akten des Vierten Internationalen Kant-Kongresses*, 1974, edited by Gerhard Funke and Joachim Kopper. Berlin and New York: Walter de Gruyter, 1974.

———. "Reduction-Realization: A Key to the Structure of Kant's Thought." *Philosophical Topics* 12 (1981): 39-98.

Burge, Tyler. "Individualism and Self-Knowledge." Journal of Philosophy (1988) 85: 649-663.

——. "Reason and the First Person." In *Knowing Our Own Minds*, edited by Crispin Wright, Barry C. Smith, and Cynthia Macdonald. Oxford: Clarendon, 1998. (243-270).

Carl, Wolfgang. "Apperception and Spontaneity." *International Journal of Philosophical Studies* 5 (1997): 147-63.

——. "Ich und Spontaneität." In *Philosophie in synthetischer Absicht*, edited by Marcelo Stamm. Stuttgart: Klett-Cotta, 1998, 105-22.

Cassirer, Ernst. *Kant's Life and Thought*. Translated by James Haden. New York and London: Yale, 1981.

------. *The Philosophy of the Enlightenment*. Translated by Fritz C. A. Koelln and James P. Pettegrove. Princeton: University Press, 1951.

- Dear, Peter. "Method and the Study of Nature". In *The Cambridge History of Seventeenth-Century Philosophy*, edited by Daniel Garber and Michael Ayers. Two volumes. Cambridge: Cambridge University Press, 1998.
- Deleuze, Gilles. *La Philosophie Critique de Kant*. Translated by Hugh Tomlinson and Barbara Habberjam, under the title *Kant's Critical Philosophy: the Doctrine of the Faculties*. Minneapolis: University of Minnesota Press, 1984.
- Deligiorgi, Katerina. "Universalisability, Publicity, and Communication: Kant's Conception of Reason." *European Journal of Philosophy* 10 (2002): 143-159.
- Descartes, René. *The Philosophical Writings of Descartes*. Translated by John Cottingham, Robert Stoothoff, Dugald Murdoch. Three volumes. Cambridge: Cambridge University Press, 1984.
- Dörflinger, Bernd. "The Underlying Teleology of the First Critique." In *Proceedings of the 8<sup>th</sup> International Kant Congress* (Memphis, 1995), edited by Hoke Robinson. Milwaukee: Marquette University Press, 1995.
- Engstrom, Stephen. "Introduction." In *Critique of Practical Reason*. Translated by Werner S. Pluhar. Indianapolis and Cambridge: Hackett, 2002.

———. "The Principle of Sufficient Reason in Kant's Practical and Theoretical Philosophy." Ph.D. dissertation, University of Chicago, 1986.

- Euclid. *The Thirteen Books of the Elements*. Translated by Sir Thomas Heath, with commentary. New York: Dover, 1956.
- Evans, J. Claude. "Two-Steps-in-One-Proof: the Structure of the Transcendental Deduction of the Categories." *Journal of the History of Philosophy* 28 (1990): 553-570.
- Förster, Eckart. "Is there 'A Gap' in Kant's Critical System?" *Journal of the History of Philosophy* 25 (1987).

——. "Kant's Notion of Philosophy." *Monist* 72 (1989): 285-304.

- Friedman, Michael. *Kant and the Exact Sciences*. Cambridge, Mass.: Harvard University Press, 1992.
- Garber, Daniel. "Descartes' Method and the Role of Experiment." In *Descartes*, ed. John Cottingham. Oxford: University Press, 1998.
- Glouberman, M. "The Methodological Development of Critical Philosophy." *Journal of the History of Philosophy* 17 (1979): 217-42.
- Graubner, Hans. Form und Wesen: ein Beitrag zur Deutung des Formbegriffs in Kants 'Kritik der reinen Vernunft'. Kant-Studien, supplementary volume 104. Bonn: Bouvier, 1972.

Gulley, Norman. "Greek Geometrical Analysis", Phronesis 3 (1958):1-14.

Guyer, Paul. "The Failure of the B-Deduction." *The Southern Journal of Philosophy* 25, Supplement (1986): 67-84.

———. "Kant on Apperception and A Priori Synthesis". *American Philosophical Quarterly* 17 (1980): 205-212.

———. "Kant's Strategies in the Transcendental Deduction." *Philosophical Topics* 12 (1981): 157-99.

—. "Psychology and the Transcendental Deduction." In *Kant's Transcendental Deductions*. Edited by Eckart Förster. Stanford: University Press, 1989.

- Hamann, Johann Georg. *Briefwechsel*. Edited by Walther Ziesmer and Arthur Henkel. Wiesbaden: Insel, 1955.
- Heidegger, M. Kant und das Problem der Metaphysik. Frankfurt: Vittorio Klostermann, 1951.
  Translated by Richard Taft, under the title Kant and the Problem of Metaphysics. Fifth edition. Bloomington and Indianapolis: Indiana University Press, 1997.
  - ———. Phänomenologische Interpretation von Kants "Kritik der Reinen Vernunft." Translated by Parvis Emad and Kenneth Maly, under the title Phenomenological Interpretation of Kant's "Critique of Pure Reason." Bloomington and Indianapolis: Indiana University Press, 1997.
- Heimsoeth, Heinz. "Metaphysical Motives in the Development of Critical Idealism." In *Kant: Disputed Questions*, edited by Moltke S. Gram. Chicago: Quadrangle Books, 1967.

- Henrich, Dieter. "Kant's Notion of a Deduction and the Methodological Background of the First *Critique*." In *Kant's Transcendental Deductions*, edited by Eckart Förster. Stanford: University Press, 1989. 29-46
  - . "The Proof-Structure of Kant's Transcendental Deduction", in *Review of Metaphysics* 22 (1969): 640-59. Reprinted in *Kant on Pure Reason*, edited by Ralph C.S. Walker. Oxford: University Press, 1982. German version, appearing under the title "Die Beweisstruktur von Kants transzendentaler Deduktion." In *Kant: Zur Deutung Seiner Theorie von Erkennen und Handeln*, edited by Gerold Prauss. Köln: Kiepenheuer and Witsch, 1973.
- ———. "Über die Einheit der Subjektivität." Philosophische Rundschau 3 (1955): 28-69.

Hintikka, Jaakko. "Kant on the Mathematical Method." Monist 51 (1967): 352-75.

- Hintikka, Jaakko and Unto Remes. *The Method of Analysis: Its Geometrical Origin and Its General Significance*. Boston Studies in the Philosophy of Science, vol. 25; Synthese Library, vol. 75. Dortrecht and Boston: D. Reidel, 1974.
- Horstmann, R.-P. "Space as Intuition and Geometry: Observations on Kant's Transcendental Aesthetic." *Ratio* 18 (1976): 17-30.
- Humphrey, T. "The Historical and Conceptual Relations between Kant's Metaphysics of Space and Philosophy of Geometry." *Journal of the History of Philosophy* 11 (1973): 483-517.
- Hume. An Enquiry concerning the Human Understanding. Edited by Eric Steinberg. Indianapolis: Hackett, 1977.

—. *A Treatise of Human Nature*. Edited by L.A. Selby-Bigge. Second edition, P.H. Nidditch. Oxford: Clarendon, 1978.

Hurley, S.L. "Kant on Spontaneity and the Myth of the Giving." *Proceedings of the Aristotelian Society* 94 (1994): 137-64.

——. "Myth upon Myth." *Proceedings of the Aristotelian Society* 96 (1996): 253-60.

- Kawamura, Katsutoshi. Spontaneität und Willkür: die Freiheitsbrgriff in Kants Antinomielehre und seine historischen Wurzeln. Stuttgart-Bad Cannstatt: Friedrich Frommann, 1996.
- Kirkland, Frank M. "Apperception and Combination: Some Kantian Problems." *Philosophy* and Phenomenological Research 49 (1989): 447-461.

Kitcher, Patricia. "Kant on Self-Consciousness." Philosophical Review 108 (1999): 345-86.

—. "Kant's Dedicated Cognitivist System." In *Historical Foundations of Cognitive Science*. Edited by J-C. Smith. Dordrecht: Kluwer Academic Publishers, 1990.

———. "Kant's Epistemological Problem and Its Coherent Solution." *Philosophical Perspectives* 13 (1999): 415-41.

———. "Kant's Real Self", in *Self and Nature in Kant's Philosophy*, edited by Allen W. Wood. Ithaca: Cornell, 1984. 113-147.

——. Kant's Transcendental Psychology. Oxford: University Press, 1990.

———. "The Presuppositions of Kant's Transcendental Deduction." Paper presented at the Fourth Athens-Pittsburgh Symposium, Delphi, 2003.

———. "Revisiting Kant's Epistemology: Skepticism, Apriority, and Psychologism." *Nous* 29 (1995): 285-315.

Körner, S. "The Impossibility of Transcendental Deductions." Monist 51 (1967): 317-31.

Korsgaard, Christine. "Kant's Analysis of Obligation: the Argument of *Foundations I*." *Monist* 72 (1989).

———. The Sources of Normativity. Cambridge: Cambridge University Press, 1996.

- —. "Self-Constitution: Action, Identity, and Integrity." John Locke Lectures, delivered at Oxford University, 2002; text version available at <a href="http://www.people.fas.harvard.edu/~korsgaar/>centersearcollectrol
- Laywine, Alison. "Problems and Postulates: Kant on Reason and Understanding." *Journal of the History of Philosophy* 36 (1998): 279-309.
- Leibniz, G.W. *New Essays on Human Understanding*. Translated and edited by Peter Remnant and Jonathan Bennett. Cambridge: Cambridge University Press, 1981.
- Liddell, Henry George and Robert Scott. *A Greek-English Lexicon*. Revised and Augmented by Henry Stuart Jones, with assistance of Roderick McKenzie. With revised supplement. Oxford: Clarendon Press, 1996.
- Locke, John. An Essay concerning Human Understanding. Edited by Peter H. Nidditch. Oxford: Clarendon Press, 1975.
- Longuenesse, Béatrice. Kant and the Capacity to Judge. Princeton: University Press, 1998.
- Lovejoy, Arthur O. "Kant's Antithesis of Dogmatism and Criticism." In *Kant: Disputed Questions*. Edited by Moltke S. Gram. Chicago: Quadrangle Books, 1967.
- Lyotard, Jean-François. *Leçons sur l'analytique du sublime*. Translated by Elizabeth Rottenberg, under the title *Lessons on the Analytic of the Sublime*. Stanford: University Press, 1994.
- MacFarlane, John Gordon. "What does it mean to say that logic is formal?" Chapter 4, "Kant and the Formality of Logic". Ph.D. Dissertation, University of Pittsburgh, 2000.
- Mahoney, Michael S. "Another Look at Greek Geometrical Analysis." *Archive for the History of the Exact Sciences* 5 (1968-9): 318-48.
- Malter, Rudolf. "L'analyse comme procédé de la métaphysique: L'opposition à la méthodologie wolffienne dans la *Preisschrift* de Kant en 1763." *Archives de Philosophie* 42 (1979): 575-591.
- McDowell, John. "Two Sorts of Naturalism." In *Mind, Value, and Reality*. Cambridge, Massachusetts and London: Harvard University Press, 1998: 167-197.

——. *Mind and World*. Cambridge, Massachusetts and London: Harvard, 1994.

- ——. "Having the World in View: Sellars, Kant, and Intentionality." Woodbridge Lectures, 1997. *Journal of Philosophy* 45 (1998): 431-491.
- Meerbote, Ralf. "Kant's Functionalism." *Historical Foundations of Cognitive Science*. Edied by J.-C. Smith. Dordrecht: Kluwer, 1990. (Philosophical Studies Series, vol. 46.)
- ———. "Deleuze on the Systematic Unity of the Critical Philosophy." *Kant-Studien* 77 (1986): 347-354.

—. "Function and Purpose in Kant's Theory of Knowledge." In *Proceedings of the 8<sup>th</sup> International Kant Congress* (Memphis, 1995), edited by Hoke Robinson. Milwaukee: Marquette University Press, 1995.

- Mohr, Georg and Marcus Willaschek, eds. *Immanuel Kant: Kritik der reinen Vernunft*. Berlin: Akademie Verlag, 1998.
- Nayak, Abhaya C. and Eric Sotnak. "Kant on the Impossibility of the 'Soft Sciences'." *Philosophy and Phenomenological Research* 55 (1995): 133-51.
- Neiman, Susan. *The Unity of Reason: Rereading Kant.* New York and Oxford: Oxford University Press, 1994.
- Palmquist, Stephen. "Knowledge and Experience: An Examination of the Four Reflective 'Perspectives' in Kant's Critical Philosophy." In *Immanuel Kant: Critical Assessments*. Edited by Ruth F. Chadwick and Clive Cazeaux. Volume II. London and New York: Routledge, 1992.
- Parsons, C. "Infinity and Kant's Conception of the 'Possibility of Experience'." *Philosophical Review* 73 (1964).
- Paton, H.J. *The Categorical Imperative: A Study in Kant's Moral Philosophy*. Chapter 1, "The Critical Method". Reprint. Philadelphia: University of Pennsylvania, 1971.
- Pippin, Robert. "Kant on the Spontaneity of the Mind." In *Idealism as Modernism.* Cambridge: Cambridge University Press, 1997.
  - ——. *Kant's Theory of Form: An Essay on the "Critique of Pure Reason"*. New Haven and London: Yale, 1982.

Plato. Opera. Vol. 4. Edited by J. Burnet. Oxford: Clarendon, 1902.

- ——. *Republic*. Translated by G.M.A. Grube, revised by C.D.C. Reeve. Indianapolis: Hackett, 1992.
- Politis, Vasilis. "The Apriority of the Starting-point of Kant's Transcendental Epistemology." International Journal of Philosophical Studies 5 (1997): 255-84.
- Rescher, Nicholas. Kant and the Reach of Reason: Studies in Kant's Theory of Rational Systematization. Cambridge: Cambridge University Press, 2000.
- Rosales, Alberto. "Zur teleologischen Grundlage der transzendentale Deduktion der Kategorien." *Kant-Studien* 80 (1989): 377-404.
- Schiller, Friedrich. Über die Ästhetische Erziehung des Menschen, in einer Reihe von Briefen. German text with English translation, by Elizabeth M. Wilkinson and L.A. Willoughby. Oxford: Clarendon, 1967.

Schönecker, Dieter. "Zur Analytizität der Grundlegung." Kant-Studien 87 (1996): 348-54.

- Sellars, Wilfrid. "The Role of Imagination in Kant's Theory of Experience." In *Categories: a Colloquium*, edited by Henry W. Johnstone. State College, Pennsylvania: Department of Philosophy, The Pennsylvania State University, 1978.
- - —. "....this I or he or it (the thing) which thinks..." *Proceedings and Addresses of the American Philosophical Association* 44 (1970-71): 5-31.
- Smit, Houston. "The Role of Reflection in Kant's *Critique of Pure Reason.*" *Pacific Philosophical Quarterly* 80 (1999): 203-223.
- Strawson, P.F. *The Bounds of Sense: An Essay on the "Critique of Pure Reason"*. London and New York: Routledge, 1966.

- —. "Imagination and Perception". In *Kant on Pure Reason*, edited by Ralph C.S. Walker. Oxford: University Press, 1982.
- ———. "Kant's New Foundations of Metaphysics." In *Entity and Identity and Other Essays*. Oxford: Clarendon Press, 1997.
  - —. "Sensibility, Understanding, and Synthesis: Comments on Henrich and Guyer." In *Kant's Transcendental Deductions*, edited by Eckart Förster. Stanford: University Press, 1989.
- Tetens, Johann Nicolaus. *Philosophische Versuche über die menschliche Natur*. Twelfth essay, "Ueber die Selbsttätigkeit und Freyheit." Hildesheim and New York: Georg Olms, 1979.
- Tonelli, Giorgio. "Der Streit über die Mathematische Methode in der Philosophie in der Ersten Hälfte des 18. Jahrhunderts und die Entstehung von Kant's Schrift über die Deutlichkeit." *Archiv für Philosophie* 9 (1959): 37-66.
- Uehling, Theodore E. "Wahrnehmungsurteile and Erfahrungsurteile Reconsidered." In Immanuel Kant's Prolegomena to Any Future Metaphysics in Focus, edited by Beryl Logan, 226-39. New York: Routledge, 1996.
- Velkley, Richard L. Freedom and the End of Reason: On the Moral Foundations of Kant's Critical Philosophy. Chicago and London: University of Chicago Press, 1989.
- Walsh, W.H. "The Idea of a Critique of Pure Reason." In *Hegel's Critique of Kant*, edited by Stephen Priest. Oxford: Clarendon Press, 1987 (pp. 119-133).
- Warren, Daniel. "Kant and the Apriority of Space." Philosophical Review 107 (1998): 179-224.
- Wartenberg, Thomas E. "Order through Reason: Kant's Transcendental Justification of Science." *Kant-Studien* (1979): 409-424.

- Waxman, Wayne. "Kant on the Possibility of Thought: Universals Without Language." *Review* of Metaphysics 48 (1995): 809-858.
- Willaschek, Marcus. "Phenomena/Noumena und die Amphibolie der Reflexionsbegriffe." In Immanuel Kant: Kritik der reinen Vernunft, edited by Georg Mohr and Marcus Willaschek. Klassiker Auslegen, vol. 17/18. Berlin: Akademie Verlag, 1998.
- Zöller, G. "Kant on the Generation of Metaphysical Knowledge." From *Kant: Analysen, Probleme, Kritik*, edited by Hariolf Oberer and Gerhard Seel. Würzburg: Königshausen und Neumann, 1988, 71-90.
- Zweig, Arnulf. Editor and translator, *Kant: Philosophical Correspondence, 1759-99.* Chicago and London: University of Chicago, 1967.