THE UNIVERSITY OF CHICAGO

# KANT AND THE ORIGIN OF THE CONCEPT OF MIND

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE HUMANITIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

BY

PIRACHULA CHULANON

CHICAGO, ILLINOIS

AUGUST 2022

# Table of Contents

# Note on Citations

Citations from Kant's *Critique of Pure Reason* are given by the pagination of the first (A) edition of 1781 and the second (B) edition of 1787. Citations from Kant's other works are given by the volume and page number of the Academy Edition (Ak.), i.e., *Kants Gesammelte Schriften*, edited by the Königliche Preussischen Akademie der Wissenschaften (Berlin: Georg Reimer, later Walter de Gruyter, 1902-). Unless otherwise indicated, translations are taken from *The Cambridge Edition of the Works of Immanuel Kant*, edited by Paul Guyer and Allen Wood (Cambridge: Cambridge University Press, 1992-). In addition, the following abbreviations are used along with the volume and page number of the Academy Edition for works other than the first *Critique*:

| | |
|---|---|
| *Anth.* | *Anthropologie in pragmatischer Hinsicht abgefaßt* (1798), Ak. 7:117-334 = *Anthropology from a Pragmatic Point of View* |
| *BW* | *Briefwechsel*, Ak. 10-13 = *Correspondence* |
| *Diss.* | *De mundi sensibilis atque intelligibilis forma et principiis* (1770), Ak. 2:385-420 = *On the Form and Principles of the Sensible and the Intelligible World* ("*Inaugural Dissertation*") |
| *EE* | *Erste Fassung der Einleitung in die Kritik der Urteilskraft* (1794), Ak. 20:193-251 = *First Introduction to the Critique of the Power of Judgment* |
| *GMS* | *Grundlegung der Metaphysik der Sitten* (1785), Ak. 5:385-464 = *Groundwork of the Metaphysics of Morals* |
| *JL* | *Immanuel Kants Logik, ein Handbuch zu Vorlesungen* ("*Jäsche Logik*") (1800), Ak. 9:1-150 |

| | |
|---|---|
| *KpV* | *Kritik der praktischen Vernunft* (1788), Ak. 5:1-164 = *Critique of Practical Reason* |
| *KU* | *Kritik der Urteilskraft* (1790), Ak. 5:165-485 = *Critique of the Power of Judgment* |
| *LB* | *Logik Blomberg* (1771?), Ak. 24:7-301 |
| *LDW* | *Logik Dohna-Wundlacken* (1792), Ak. 24:697-784 |
| *MAN* | *Metaphysische Anfangsgründe der Naturwissenschaft* (1786), Ak. 4:465-566 = *Metaphysical Foundations of Natural Science* |
| *MD* | *Metaphysik Dohna* (1792-93), Ak. 28:610-704 |
| *MH* | *Metaphysik Herder* (1762-64), Ak. 28:1-181 |
| $MK_2$ | *Metaphysik* $K_2$ (early 1790s), Ak. 28:705-816 |
| $ML_1$ | *Metaphysik* $L_1$ (mid-1770s?), Ak. 28:185-350 |
| $ML_2$ | *Metaphysik* $L_2$ (1790-91?), Ak. 28:531-594 |
| *MM* | *Metaphysik Mrongovius* (1782-83), Ak. 29:747-940 |
| *MV* | *Metaphysik Vigilantius* (1794-95), Ak. 29:943-1040 |
| *MVo* | *Metaphysik Volckmann* (1784-85), Ak. 28:351-459 |
| *MvS* | *Metaphysik von Schön* (1780s?), Ak. 28:460-524 |
| *Prol.* | *Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können* (1783), Ak. 4:255-383 = *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science* |
| *SF* | *Der Streit der Facultäten, in drey Abschnitten* (1798), Ak. 7:1-116 = *The Conflict of the Faculties* |
| *UE* | *Über eine Entdeckung, nach der alle neue Kritik der reinen Vernunft durch eine ältere entbehrlich gemacht werden soll* (1790), Ak. 8:185-252 = *On a Discovery Whereby Any New Critique of Pure Reason Is to Be Made Superfluous by an Older One* |

| | |
|---|---|
| *WF* | *Welches sind die wirklichen Fortschritte, die die Metaphysik seit Leibnitzens und Wolffs Zeiten in Deutschland gemacht hat?* (1793/1804), Ak. 20:253-332 = *What Real Progress Has Metaphysics Made in Germany Since the Time of Leibniz and Wolff?* |
| *WL* | *Wiener Logik* (1780-82), Ak. 24:787-940 = *The Vienna Logic* |

# Acknowledgements

It is uncommon to express gratitude toward the subject of one's dissertation. Nonetheless, I would like to express my sincere gratitude to the late Immanuel Kant (1724-1804) of Königsberg for having written philosophical works of marvelous depths. These works have been a source of illumination and delight for over a decade of my life.

I am grateful to my teachers who first introduced me to Kant. I discovered the *Groundwork* in Nuangnoi Boonyanate's introductory ethics class in my second year in Bangkok and was immediately captivated by its intellectual beauty. When I moved to Berlin the next year, I was fortunate to have the chance to attend Katalin Makkai's seminars, which sustained and deepened my interest in Kant. I am indebted to Katalin also for her continuous support of my work and my career over the years. It must have been while sitting in one of Tobias Rosefeldt's seminars, in which I learned how to find my way through Kant's thorny texts, that I decided to become a Kant scholar.

The members of my dissertation committee, each in their own way, are rare models of scholarly excellence and generosity of spirit. I am extremely humbled and honored to have worked with them for so many years. Matthew Boyle has been a patient, kind, and wise mentor since I first arrived in the graduate program. The most important lesson I have learned from Matt, and from his work, is that one can make much headway in philosophy just by putting in effort to clarify and motivate the problem one seeks to solve. I am afraid I have never done that to Matt's satisfaction, but I hope it will be clear to the reader how this lesson guided this project from the very beginning. I have benefited greatly from Michael Kremer's brilliant ear for sound philosophical problems and theses. Michael was also one among us who always kept his eye on the practical side of things. I would not have finished my degree in a timely manner and secured a job without his invaluable advice on practical

and professional matters. I am grateful to Tobias Rosefeldt for his constant support of my career, his generous engagement with my work, and his hospitality during my numerous visits to Berlin, which were as crucial for my research as for my personal life. Conversations with Eric Watkins and his meticulous feedback not only saved this work from numerous errors but also transformed my thinking about Kant. Eric's work, for me, exemplifies the highest standard of Kant scholarship that I can only hope to approach asymptotically in my own work.

During my time at Chicago, I have learned and benefited from conversation with faculty members of the Philosophy Department, especially James Conant, Kevin Davey, Matthias Haase, Daniel Moerner, Anubav Vasudevan, Candace Vogler, and Malte Willer. My intellectual forays outside of philosophy have been kindly encouraged by David Levin, Richard Neer, and David Wellbery.

I have presented portions of this dissertation, in earlier versions, to audiences at the University of Chicago German Philosophy Workshop (multiple times), the Forschungskolloquium Deutsche Klassische Philosophie at the Humboldt-Universität zu Berlin (multiple times), the University of Leipzig, the Ludwig-Maximilians-Universität München, the Leuven Kant Conference, the NYC Workshop in Early Modern Philosophy, and Princeton University. I am grateful to these audiences for stimulating discussion and helpful feedback. Special thanks are due to Julian Bacharach, Karin de Boer, Andrew Chignell, Katharina Kraus, Béatrice Longuenesse, and James Messina.

My visit to the Humboldt-Universität zu Berlin in 2020-21 was enabled by a research fellowship from Deutscher Akademischer Austauschdienst (DAAD). I was able to complete the work on this dissertation through a dissertation completion fellowship from the Mellon Foundation. I thank both organizations for their generous support.

I owe special thanks to two friends for their help with my thinking on the topic of this dissertation. From conversation with Amy Levine over the years, I have learned about philosophy perhaps more than from all my graduate seminars combined. Andrew Pitel has been an erudite and

thoughtful interlocutor on all aspects of Kant's philosophy and on the history of philosophy. For their help, conversation, and companionship, I am grateful to Manasicha Akepiyapornchai, Natalia Albizu, Maddalena Alvi, Sarale Ben Asher, Christopher Benzenberg, Sebastian Bürkle, Rosalind Chaplin, Chanikan Charoensri, Nicholas Currie, Sonja Dorau, Arnika Fuhrmann, Rory Hanlon, Lukas Ishar, Mathis Koschel, David Kretz, Maya Krishnan, Hyunseoung Lee, Sam Lee, Marta Medico, Peerapat Ouysook, Peera Panarut, Thomas Pendlebury, Eléonore Rimbault, Christopher Roser, Ron Sadan, Sam Segal, Shubham Shivang, Johannes Sudau, Olivia Sultanescu, Sanittawan Tan, Josh Trubowitz, Troy Vine, Martijn Wallage, and Michal Zechariah.

My greatest debts are due to my family for their love, their support, and, above all, their forgiveness for the fact that I have spent so much time that rightfully belonged to them studying, writing, and being away. My parents, Chulaphun Chulanon and Veeraporn Nitiprapha, are my first teachers of *Weltweisheit*. Since the day we met, Ronen Shifron has been *another* source of illumination and delight.[1]

[1] If you did not realize what it meant to say 'another' here, as Ronen himself did not, read the beginning again.

# Abstract

This dissertation articulates and defends a new interpretation of the concept of mind which underlies Kant's account of mental faculties in the three *Critiques*. On my reading, this concept is to be construed in terms of the transcendental idea of the soul (an immaterial thinking substance) in its regulative use. This offers an alternative to previous accounts, which construe Kant's claims about the mind either as metaphysically neutral claims, as claims to cognition of the mind as an appearance, or as claims to cognition of the mind as a noumenal entity. On my reading, although Kant's concept of mind is metaphysically substantive and even purports to give us a glimpse of noumenal reality, Kant's use of this concept neither is grounded in cognitive access to an independently existing object nor gives rise to a claim to cognition of the mind as it is in itself. Rather, Kant's employment of this concept in Critical philosophy is warranted by reason's demand for explanatory understanding (or *Begreifen*), which cannot be fully met by our capacity for cognition.

# Introduction

> Transcendental philosophy is the propaedeutic of metaphysics proper. (*MM* 29:751, cf. A841/B869)
>
> Metaphysics has as the proper end [zum eigentlichen Zwecke] of its investigation only three ideas: God, freedom, and immortality. (B395n)

## The Metaphysics of Mind and the Problem of Metacritique

One remarkable aspect of Kant's Critical philosophy is the extent to which a conception of a mind endowed with faculties underlies both its content and its form as a whole. Kant frames his central philosophical theses—on issues ranging from the nature of knowledge and reality to morality and aesthetic experience—in terms of the character of our mental faculties. Together, the three *Critiques* draw a map of the human mind's basic faculties, specifying their hierarchy, their division, and their relation to one another. This map not only presents a certain image of the human mind but also acts as a peculiar kind of blueprint which Kant fills in with substantive philosophical theses about various aspects of reality in the broadest sense. More curiously, perhaps, the order in which these theses and their arguments are presented also mirrors the order of the faculties (think of the division between the three *Critiques* themselves or the division of the first *Critique* into the Aesthetic, the Analytic, and the Dialectic). This reflects Kant's own conviction that the systematic character of his philosophy, its comprehensive scope as well as its internal unity, is grounded in the inherent systematicity of the mind, its primary object of inquiry.[1] Finally, the idea of a mind endowed with

[1] Cf. *EE* 20:246, *KU* 5:168.

faculties is built into Kant's doctrine of transcendental idealism, according to which objects of experience are *mind*-dependent. Kant fleshes out the relevant notion of mind-dependence in terms of the claim that objects owe their forms, such as spatiotemporality and categorial properties, to mental faculties that represent them. This explains why these formal features of objects can be cognized *a priori* and, hence, why synthetic *a priori* cognition is possible.

Historically, commentators tend to regard Kant's faculty psychology with suspicion. The philosophical motivations behind Kant's appeal to mental faculties in making philosophical claims are not obvious, and there are good reasons to suspect that the framework itself is unnecessary and incoherent. In almost every case, it seems possible to strip a philosophically respectable claim of its dubious psychological dress. We can separate Kant's claim that objects of our perceptual experience are necessarily in space and time from his claim that space and time are the forms of our sensibility, the way in which our mind is affected by objects. Similarly, we can separate his claim that we necessarily think of objects of experience as exhibiting an ontological structure describable in terms of ontological concepts, such as substance and causation, from his claim that these ontological concepts (the categories) are the forms of the faculty of understanding, the rules according to which the understanding undertakes its synthetic activity. In his practical philosophy, Kant's claim that the categorical imperative is the normative principle that governs our actions seems detachable from the claims that it is the law of the will or the higher faculty of desire. In his aesthetic theory, Kant's analysis of aesthetic judgment as subjectively universal judgment grounded in formal purposiveness of an object seems separable from his account of aesthetic experience as "free play" of the imagination and the understanding. In each of these cases, then, it seems reasonable to ask whether Kant could have affirmed the non-psychological part of his claim without its psychological part and whether the psychological part adds anything of significance.

But, first of all, what are mental faculties? What is the mind, whose character Kant purports to describe by identifying these faculties? This question has been raised already by the first readers of the *Critique of Pure Reason* as part of the so-called problem of metacritique, the problem of the justification of the basic concepts and propositions of the Critical system.[2] Johann Georg Hamann, who coined the term 'metacritique', was perhaps the first to raise this question in writing—in 1783, two years after the first publication of the *Critique*.[3] Deliberately parodying Kant's own question, Hamann asks: "How is the *faculty* of thinking possible?" In other words, he is asking what the faculty of thinking or reason is, to which Kant appeals in answering the question how synthetic *a priori* judgments are possible. In raising this question, Hamann has a specific criticism in mind. In his view, reason or the rational mind is not a self-subsisting thing but entirely depends on its concrete embodiment in human experience, culture, and language. Hamann thinks Kant fails to recognize this and "hypostatizes" the rational mind into an independently existing thing. In doing so, Kant falls prey to "the talisman and rosary of a transcendental superstition which believes in *entia rationis* [thought entities]". Setting aside the specifics of his polemic, Hamann's challenge targets Kant's (lack of) justification for employing what I will call an *objective* concept of mind, the concept of mind as *hypostatized* or *reified*.

The present study focuses on the problem of Kant's justificatory basis for employing a concept of mind that is objective in the relevant sense. In Kantian terms, an objective concept of mind is a concept of mind as an object that can fall under the categories, the concept of an object in general. A paradigm example of an objective concept of mind is the Cartesian concept of mind as an immaterial

[2] For a useful overview of Kant's early critics, who in one way or another have recognized this problem, see Frederick Beiser, *The Fate of Reason: German Philosophy from Kant to Fichte* (Cambridge: Harvard University Press, 1993).

[3] Johann Georg Hamann, "Metakritik über den Purismum der reinen Vernunft," in *Hamann's Schriften*, ed. Friedrich von Roth, vol. 7 (G. Reimer, 1825), 1–16.

thinking substance. However, this is not the only possible objective concept of mind; other examples from the history of philosophy include the Leibnizian concept of mind as a monad, the Spinozist concept of mind as the thinking attribute, the functionalist concept of mind as a cognitive system (which has been attributed to Kant; see Chapter 3). My purpose in this chapter is to motivate the problem concerning the use of an objective concept of mind in the Kantian critique in the generic form—as a problem that arises no matter what specific concept of mind is taken to operate in the critique. On the one hand, Kant's Critical account of mental faculties seems to presuppose a concept of mind that is objective in the sense of being a concept of an object to which the categories can apply. On the other hand, Kant's various philosophical commitments, including the "mere formality" of transcendental apperception, the apriority of Critical philosophy, and noumenal ignorance, seem to prevent him from employing any such concept in the critique. The overarching goal of this study is to respond to this puzzle by articulating and defending a novel position on what specific concept of mind Kant employs in his philosophy and how his employment of this concept can be seen as compatible with his other philosophical commitments. I should emphasize from the start that I will be concerned exclusively with the question of Kant's justification for the *general* conception of the mind, commonly presupposed by his claims about mental faculties. Therefore, I will have nothing to say on the equally important question of how Kant justifies his *particular* claims about the nature of individual faculties.

It is natural to interpret Kant's concept of mind as a concept of an entity or an object characterizable in terms of the categories. One reason is that Kant applies ontological concepts (derived from the categories), such as faculty [Vermögen] and power [Kraft], to the mind (see Chapter 2). More importantly, Kant's intent in positing these faculties is clearly to identify various ways in which we can *causally* interact with objects distinct from us—the notions of individual mental faculties pick out ways in which we can be *active/spontaneous* or *passive/receptive* in relation to things in the world. Sensibility, for instance, is defined by Kant as the way in which we are passively affected by external

objects. The will, by contrast, is what enables us to actively cause changes in the world. It seems that we cannot make sense of such causal interaction unless the mind is also a thing that exists in the objective world alongside objects it causally interacts with. It is particularly telling in this respect that Kant employs the same framework of metaphysical concepts, centered around 'faculties' and 'powers', both in his Critical analysis of the mind and in his metaphysical account of matter. If this diagnosis is correct, what kind of thing is the mind? Is it a phenomenal or a noumenal entity? How is Kant's investigation in a position to make any claim about the nature of such an entity?

Without endorsing Hamann's radical position on the mind, one can see that the broadly Cartesian conception of the mind as a knowable entity sits uneasily with Kant's epistemology (his view on what we can and cannot know), his metaphysics (his view on what sorts of things exist), as well as his account of self-consciousness. As far as the received view goes, Kant thinks that things exist either as appearances or things in themselves. However, the critique is supposed to neither conduct an empirical investigation of the mind as an appearance nor make claims to cognition of the mind as it is in itself, since we have no cognitive access to things in themselves in general. Moreover, the way in which one is conscious of oneself as a thinking subject (i.e., pure apperception) does not present the thinking subject as an object for cognition at all.

Because the broadly Cartesian conception of the mind seems incompatible with these commitments, some of the most astute readers of Kant have deemed hopeless the attempt to fit this conception within Kant's system. And since Kant's faculty psychology framework presupposes this conception of mind, they think we should better find a way to preserve Kant's philosophical insights while jettisoning that framework. An influential case has been made by Peter Strawson.

> It is useless to puzzle over the status of these propositions [about mental faculties]. They belong neither to empirical (including physiological) psychology nor to an analytical philosophy of mind, though some of them many have near or remote

> analogues in both. They belong to the imaginary subject of transcendental psychology.[4]

The "imaginary subject of transcendental psychology", which echoes Hamann's invective of "transcendental superstition", has since become something of a catchphrase among Kant scholars. But I believe that the attitude it expresses also still holds sway, explicitly or implicitly, over many Kant scholars today.[5] While many are perhaps not as hostile as Strawson toward the use of faculty-psychological "terminology", they would agree with Strawson that this terminology is entirely dispensable. That is, all of Kant's philosophically worthwhile ideas ultimately can be translated into terms that make no reference to mental faculties. Kant's faculty talk may be a useful shorthand, but it has no philosophical depth.[6] For these Kant scholars, it is useless to puzzle over the status of Kant's faculty psychology not because it presupposes a pernicious kind of psychology but because it is innocuously empty.

---

[4] Peter Strawson, *The Bounds of Sense* (London: Methuen, 1966), 97.

[5] In the continental tradition, a parallel influence has been exerted by Hermann Cohen and Martin Heidegger. For a helpful overview, see Béatrice Longuenesse, *Kant and the Capacity to Judge* (Princeton: Princeton University Press, 1998), 3–5. Ironically, despite that her work highlights the significance of Kant's characterization of the understanding as a *faculty* (or capacity) to judge, Longuenesse downplays the significance of Kant's own use of the notion of faculty [Vermögen], holding that it is "incorrect to identify the meaning of *Vermögen* and *Kräfte* when applied to the mind with the meaning these terms have in the metaphysics of substance […]" (7-8). In a later work, she explains that she departs from the standard practice by translating *Vermögen* as "capacity" instead of "faculty" precisely in order to keep a distance from Kant's "dubious faculty psychology" (Béatrice Longuenesse, "Kant's Categories and the Capacity to Judge," in *Kant on the Human Standpoint* (Cambridge: Cambridge University Press, 2005), 18n.). In this sense, Longuenesse is to some extent sympathetic with the broadly Strawsonian attitude I am describing.

[6] William Walsh offers an explicit statement of this view: "Kant speaks in the language of faculty psychology, and introduces into his pages a cast of actors most of whom were already familiar to contemporaries. […] But he need not have put his points in this language" (William Walsh, *Kant's Criticism of Metaphysics* (Edinburgh: Edinburgh University Press, 1975), 251–52). See also William Walsh, "Philosophy and Psychology in Kant's Critique," *Kant-Studien* 57, no. 1–4 (1966): 186–98.

If one asks how this rehabilitative task can be done, one can still look to Strawson's probing exposition for an unparalleled example. Strawson proposes that we can extricate two strands of investigation in the *Critique of Pure Reason* from each other and coherently embrace one while rejecting the other. The first, philosophically respectable strand is the investigation into "the general structure of ideas and principles which is presupposed in all our empirical knowledge".[7] This seeks to identify structural features that must be exhibited by any type of experience which we can coherently or intelligibly conceive of. Hence, the first strand of Kant's investigation aims at the "determination of the fundamental general structure of any conception of experience such as we can make intelligible to ourselves".[8] Because intelligibility is what defines the scope of this conception of experience such that no part of it can be intelligibly negated, it can serve as a "first principle" from which Kant's inquiry can then proceed by means of conceptual analysis alone. However, there is also a second, dubious strand in Kant, namely, the investigation into "the structure and workings of the cognitive capacities of beings such as ourselves". The connection between the second and the first investigation is this: "whatever necessities Kant found in our conception of experience he ascribed to the nature of our faculties".[9] In Strawson's view, Kant needlessly conflates the limits of intelligibility with the limits of our mental faculties. In addition, he also thinks that Kant's "transcendental psychology" on its own fails Kant's own standard of intelligibility. Therefore, Strawson urges that we jettison the second strand of investigation and reformulate the results of the first in a way that does not rely on the idiom of faculty psychology. The point of this exercise is not merely to weed out errors from Kant's otherwise respectable way of thinking but also to render it more faithful to Kant's own intent. Strawson sees and

[7] Strawson, *The Bounds of Sense*, 19.

[8] Strawson, 44.

[9] Strawson, 19.

commends Kant as one of the few historical champions of "descriptive metaphysics", which aims, singularly, to "lay bare the most general features of our conceptual structure".[10]

Consider, for instance, the so-called 'discursivity' thesis: the thesis that experience requires the cooperation of the sensibility and the understanding—that objective representation requires both intuitions and concepts. In Strawson's view, the valid point of the discursivity thesis is that any intelligible account of experience must necessarily assume "the duality of general concepts, on the one hand, and particular instances of general concepts, encountered in experience, on the other".[11] Thus, for Strawson, the discursivity thesis is a quasi-logical claim about "the general structure of ideas and principles which is presupposed in all our empirical knowledge". However, there is no reason to express such a claim in the idiom of mental faculties (i.e., sensibility and understanding). In the same vein, Strawson argues, we can and should reformulate all of Kant's propositions about the structure and workings of mental faculties into propositions that are strictly about quasi-logical structure internal to knowledge. The persistence of the Strawsonian attitude can be seen, for instance, in a common assumption (characteristic of the anti-metaphysical reading, which we will examine in Chapter 3) that we can construe Kant's central distinction between spontaneity and receptivity as a purely epistemological distinction.[12]

Note that Kant not only employs the faculty psychology idiom to formulate his philosophical theses but also underscores the methodological centrality of mental faculties. He identifies the rational

---

[10] Peter Strawson, *Individuals: An Essay in Descriptive Metaphysics* (London and New York: Routledge, 1959), 9.

[11] Strawson, *The Bounds of Sense*, 20.

[12] Karl Ameriks's dictum serves as a representative example: "The main point of his transcendental account is simply that, no matter what the ultimate metaphysical story about our self is, knowledge requires more than mere sensibly given data; it must also involve conceptual functions, and thus some kinds of spontaneity that are not exactly the same acts as the reception of data. This is a distinction internal to epistemology" (Karl Ameriks, "Kantian Apperception and the Non-Cartesian Subject," in *Kant and the Historical Turn: Philosophy as Critical Interpretation* (Oxford: Oxford University Press, 2006), 51–66).

faculties as the primary objects of Critical philosophy and holds that the distinctive mode of cognition instantiated therein turns on the special cognitive relation we have to our own mental faculties (see Chapter 1). According to Kant, the critique of pure reason seeks to investigate the nature of the mind's rational faculties rather than any domain of objects, with which other sciences are concerned.

> The critique of the faculties of cognition with regard to what they can accomplish *a priori* has, strictly speaking, no domain with regard to objects, because it is not a doctrine, but only has to investigate whether and how a doctrine is possible through it given the way it is situated with respect to our faculties. (*KU* 5:176, cf. 20:195, 20:202, *MM* 29:751-2)

By "the critique of the faculties of cognition", Kant means the critique of *reason*, understood in the broad sense of "the entire higher faculty of cognition" (cf. A835/B863) and, hence, the Critical system as a whole. The higher faculty of cognition is divided into three sub-faculties: the understanding, the power of judgment, and reason (in the narrow sense). Each of the higher faculties of cognition legislates laws for each of the three basic faculties of the mind (the understanding for the faculty of cognition, reason for the faculty of desire, and judgment for the faculty of feeling of pleasure and displeasure). In this way, the critique of pure reason, "taken in the most general sense", has as its object the *entire* system of faculties of the human mind insofar as they fall under reason's legislation and, hence, are the sources of synthetic *a priori* judgments.

Some of Kant's remarks also anticipate and warn against the temptation to interpret his project as aiming at something like the Strawsonian account of "the general structure of ideas and principles".

> I have learned from the critique of pure reason that philosophy is not a science of representations, concepts and ideas, or a science of all the sciences, or anything else of this sort. It is rather a science of the human being, of his representations, thoughts and actions: it should present all the components of the human being both as he is and as he ought to be – that is, in terms both of his natural functions and of his relations of morality and freedom. (*SF* 7:69)

This makes clear that Critical philosophy is not concerned (at least not primarily) with representations, concepts, ideas, or scientific knowledge, but with the human being and her capacities for representations, thoughts and actions. This is an outright rejection of the Strawsonian attempt to dissociate his quasi-logical account of representations from his doctrine of the human mind and its faculties.

These considerations point in the same direction: we cannot properly understand the nature of Kant's philosophical project unless we take his doctrine of mental faculties seriously. This speaks against Strawson's suggestion that his deflationary approach offers not only a correction but a faithful-to-spirit reconstruction of Kant's thinking. But if we must take Kant's faculty psychology framework seriously, can we possibly defend him against the charge of incoherence? I will argue in this study that Kant has the resources to withstand the charge of incoherence with respect to his use of an objective concept of mind that underlies his account of mental faculties. These resources lie in the availability of a non-cognitive use of metaphysically substantive concepts. For Kant, there are ways of using such concepts that are not meant to give rise to claims to cognition of reality or the way things are. This opens the way for reading him as making use of a concept of mind-as-entity without at the same time committing himself to any claims of cognition of the metaphysical character of the mind. This "regulative" mode of conceptualization is what explains the propriety of talking about the mind in metaphysically substantive terms. In this way, my proposed reading steers the middle way between the anti-metaphysical and the realist reading of Kant's view on the mind.

Before I go on to sketch this reading in more detail, I would like to briefly consider another brand of the deflationary treatment of Kant's faculty psychology. The difficulty facing this approach will bring into view the direction that my own interpretation will take. This other deflationary approach denies that the mind, for Kant, is a thing in its own right but rather an attribute of a more fundamental thing, namely, the *human being*. This may seem to be suggested in Kant's remark we have just

considered, in which the critique is described as "a science of the human being", as well as in a few other places. William Walsh offers a representative statement of this reading:

> [For Kant] to speak of Understanding is simply a short way of referring to certain abilities and, perhaps, dispositions possessed not by 'the mind', whatever that is, but by persons. The operations of the Understanding are carried out by live human beings. […] Judgments are made by men, in their capacity as thinkers. The judging subject is neither real nor phenomenal but an abstraction, an abstraction from the concrete person who moves and has his being in the experienced world.[13]

Walsh's suggestion is that we can circumvent the metaphysical dispute about what the mind is by construing mental faculties as an abstraction or a generalization about abilities and dispositions of human beings. In this sense, whereas the Strawsonian deflationism seeks to assimilate Kant's Critical philosophy to logic, this branch of deflationism seeks to assimilate it to anthropology (in the Kantian sense). The former approach treats judgments (or, more broadly, representations) as abstract entities; the latter treats them as what human beings do.

It is plausible that Kant thinks of the mind and mental faculties *in the final analysis* as part of the human being, a being that is both embodied and rational, possessing both physical and psychological attributes. This explains why he occasionally reverts to the concept of the human being to characterize his object of inquiry. However, anthropological deflationism assumes that the conception of the human being as the fundamental bearer of mental faculties is unproblematically available to the Critical investigation from the outset. Put differently, it assumes that the subject who embarks upon a Critical investigation of her own rational faculties can readily help herself to the conception of herself as a human being. This seems to me to get things wrong in the order of explanation. Like Descartes's

[13] Walsh, *Kant's Criticism of Metaphysics*, 251–52. More recent proponents of this approach include Longuenesse, *Kant and the Capacity to Judge*; and Johannes Haag, "Faculties in Kant and German Idealism," in *The Faculties: A History*, ed. Dominik Perler (Oxford: Oxford University Press, 2015), 298–246.

*Mediations*, the critique of pure reason unfolds as an inquiry conducted from the point of view of the subject of thinking (the I think the transcendental subject of apperception)—what Bernard Williams famously terms the "point of view of consciousness" and what Kant calls the "human standpoint" (A26/B43).[14] While this certainly is the point of view *of* the human being, it is not the point of view within which the subject of thinking can unproblematically identify herself with a human being, which is an object in the world among other objects (I will revisit this point shortly). The anthropological approach is problematic insofar as it takes for granted that I (understood as the transcendental subject of thinking) can unproblematically identify myself with the human being that I am, or for that matter with anything at all in the objective world. While my relation to the human being that I am is the same as my relation to other objects and other human beings, my relation to myself *qua* thinker is radically different, being constituted by a distinctively first-person and non-cognitive mode of self-consciousness. If the critique proceeds from the point of view of the transcendental subject, it cannot *assume* the conception of this subject as a human being as its starting point but, rather, must *achieve* this conception through philosophical analysis.

In this study, I propose that reflecting on the status and the origin of the concept of mind in Kant's philosophy suggests a solution to the philosophical problem of how the thinking subject arrives at *any* objective conception of herself (including her self-conception as a human being). I argue that it is only against the background of this problem that we can grasp what is philosophically at stake in

[14] Other commentators who have emphasized the human standpoint, as opposed to the standpoint of divine omniscience or (in Williams's famous terms) the absolute conception, as the standpoint of Kant's investigation include Graciela de Pierris, "The Constitutive A Priori," *Canadian Journal of Philosophy* 22 (Supplementary Volume 18) (1992): 179–214; Henry Allison, *Kant's Transcendental Idealism: An Interpretation and Defense*, 2nd ed. (New Haven: Yale University Press, 2004); Béatrice Longuenesse, *Kant on the Human Standpoint* (Cambridge: Cambridge University Press, 2005); Adrian Moore, "The Transcendental Doctrine of Method," in *The Cambridge Companion to Kant's Critique of Pure Reason*, ed. Paul Guyer (Cambridge: Cambridge University Press, 2010), 310–26.

Kant's employment of the concept of mind in the critique. The cognitive role of this concept is precisely to make available a way for the thinking subject to conceive of herself as a part of the objective order. In this way, the concept of mind plays an intermediary role in the explanation of how it is possible for the thinking subject to conceive of herself as a human being, which is situated in the objective world—the concept of mind is what fills in the gap between the thinking subject and the human being.

In the remainder of this chapter, I sketch and contextualize my interpretation by introducing three themes around which my interpretation will be developed. The first theme is the problem of the objective concept of mind. As I have mentioned already, my interpretation proposes to consider Kant's account of the concept of mind as a response to the problem of the gap between an objective conception of mind and the non-objective conception of the self as the subject of thinking. Kant's solution, as I understand it, draws upon his account of *reason* as a source of substantive concepts. This will be our second theme. In particular, I will argue that Kant's concept of mind is to be construed in terms of the idea of the soul, a "pure concept of reason", in its "regulative" employment. The third theme is the primacy of understanding in Kant's philosophy. I will suggest that we should understand the employment of the pure concepts of reason in general, and the idea of the soul in particular, as supported not by their role in enabling cognition but by their role in enabling explanatory understanding. It follows that Kant's claims about the mind and mental faculties are not grounded in cognition of the mind as an entity but in reason's demand for understanding that transcends the limits of cognition.

## The Objective Concept of Mind

I first introduced the problem of the mind in Kant as a metaphysical problem—What is the mind? One reason for doing this is that this is how the issue has been traditionally approached in the

literature (see Chapters 2 and 3). Another reason is that this is how the issue must have naturally struck the reader of the *Critique*. However, as I have pointed out, metaphysically substantive concepts, for Kant, can be used in ways that do not issue in claims to cognition of independent reality. Therefore, we must distinguish the question of what Kant's (or our) concept of mind represents the mind to be from the question of what the mind is apart from this concept. An answer to the first question entails an answer to the second only if we can legitimately employ the concept of mind to form a judgment that amounts to cognition. In this study, we will attend to the question of the concept of mind as prolegomena to a metaphysics of mind. What is the origin of this concept? What cognitive role does it fulfil? What justifies Kant's (and our) employment of it?

I wish to argue that thinking about these questions as they arise in Kant is not of a purely exegetical interest. Kant's account of the concept of mind, properly understood, offers a solution to an important philosophical problem: How can we explain the transition from having a first-person perspective to having a conception of our own mind as a part of objective reality? On the one hand, we appear to employ a concept of our own mind as a part of objective reality. We individuate our mind from the minds of others. We speak of our mental lives in terms of states, processes, events, and causation. Thus, it appears that we employ an objective concept of mind, a concept of some entity or other that can be in states, undergo processes, and causally interact with other objects. On the other hand, however, the awareness of ourselves as thinking subjects, from the first-person perspective, seems devoid of any metaphysical content. In particular, this self-awareness does not seem to reveal the metaphysical nature of the subject of consciousness nor even present that subject as an object or a feature of an object. This opens up a gap between the first-personal self-awareness and any objective conception we may have of our own mind. From the "point of view of consciousness" (to use Bernard Williams's term), we seem to have no basis for applying any objective concept of mind to ourselves

as subjects of consciousness. This problem, I will argue, parallels the problem raised by Kant's employment of an objective concept of mind in the critique.

Williams introduces this problem as a response to Descartes's *cogito* argument. In Descartes' view, the first-personal consciousness of the content of our own mind is sufficient for arriving at an objective conception of the mind as an immaterial thinking substance and, indeed, for knowledge that this substance is the referent of the 'I'. Therefore, from any conscious self-ascription of mental content, such as 'I am thinking *p*' or 'I am feeling pain', we can always infer 'I exist (as an immaterial thinking substance)'. Williams denies the validity of such inference.

> The thought-event formulation […] requires the notion of objectively existing thought-events, and in supposing that it can start out merely from the idea of thoughts as experienced, and from that achieve the third-personal perspective which is necessary if this notion is to apply, it shares a basic error with Descartes. There is nothing in the pure Cartesian reflection to give us that perspective. The Cartesian reflection merely presents, or rather invites us into, the perspective of consciousness. Descartes thinks that he can proceed from that to the existence of what is, from the third-personal perspective, a substantial fact, the existence of a thinker.[15]

Williams is pointing out that the problem is not simply that we are not immediately aware of the existence of a thinker from within the point of view of consciousness. This much would have been granted by Descartes. On the contrary, Descartes's "basic error" is to take *thought-events* to exist objectively and from there infer that the bearer or the thinker of these thoughts must exist objectively as well. However, the "idea of thoughts as experienced" provides no ground for taking them to have objective existence (as accidents inhering in a thinking substance). Therefore, Williams concludes that "starting solely from the point of view of consciousness one cannot gain any objective conception of

[15] Bernard Williams, *Descartes: The Project of Pure Enquiry* (London and New York: Routledge, 1978), 100. For a penetrative discussion of Williams's problematic, from a systematic standpoint, see Naomi Eilan, "The First Person Perspective," *Proceedings of the Aristotelian Society* 95, no. 1 (1995): 51–66.

there being several such selves—nor, consequently, can one gain an objective conception of there being even one".

A historical precedent of Williams's anti-Cartesian argument is Hume's attack on Descartes and others who believe that "we are every moment intimately conscious of what we call our *self*; we feel its existence and its continuing to exist, and are certain […] both of its perfect identity and of its simplicity" (*Treatise* I.iv.6). Hume observes:

> When I look inward at what I call *myself*, I always stumble on some particular perception of heat or cold, light or shade, love or hatred, pain or pleasure, or the like. I never catch myself without a perception, and never observe anything but the perception. (*Treatise* I.iv.6)

Thus, Hume agrees with Williams that the point of view of consciousness presents us with nothing besides "thoughts as experienced" ('perceptions', in Hume's terms). In particular, it does not present us with any entity that is the bearer of these thoughts. Further, Hume argues that it is impossible to identify a single perception from which the idea of the self is derived: all of our perceptions are transitory, whereas the self is supposed to persist through all these perceptions. From this, Hume concludes that the idea of the self is completely fictitious—"we don't so much as have an *idea* of self". Both Hume and Williams rebut Descartes's argument by appealing to the character of thoughts or perceptions as experienced by the thinker. Nevertheless, there is one important difference: while Hume attends to the *temporal* character of perceptions or thoughts as they are presented to our consciousness, Williams emphasizes the fact that they do not even have any *objective* character to begin with.

Before turning to Kant, let us sharpen the problem further by looking at another formulation of it from Ludwig Wittgenstein's *Tractatus*.[16] Wittgenstein observes that if we produce a description of

[16] Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, trans. C.K. Ogden (London and New York: Routledge, 1981).

the world as it appears from our point of view—if I wrote a book "The world as I found it"—the "thinking, presenting subject" would not be found in it (5.631). *This* subject does not find itself one among the items that it thinks about or represents to itself. In this sense, "the subject does not belong to the world but it is a limit of the world" (5.632). Wittgenstein goes on to introduce an analogy between the point of view of consciousness and the field of sight: just as the eye is not visible within the field of sight, so the subject of consciousness is not "visible" within the "field of consciousness".

> 5.633 *Where in* the world is a metaphysical subject to be noted?
>
> You say that this case is altogether like that of the eye and the field of sight. But you do *not* really see the eye.
>
> And from nothing *in the field of sight* can it be concluded that it is seen from an eye.

What Wittgenstein's analogy bring out, beyond what we have already learned from Hume and Williams, is that the elusiveness of the thinking subject is not a matter of contingency, having to do with the fact that we happen to lack a representational means to gain immediate access to ourselves as a thinking subject, as if this subject were in front of our eyes all along but undetectable by our sensorium. On the contrary, this limit appears to be a matter of necessity. What seems to be required for us to conceive of the subject of thinking as part of the objective world is to perform the conceptually impossible task of "stepping back" from the point of view of consciousness so as to view our cognitive situation as a whole from outside of it and, hence, to view ourselves as thinking subjects situated within. In Wittgenstein's analogy, this would be like viewing, with our eyes, our own visual field in its entirety from the outside, like a bubble, with our own eye sticking to its point of origin.

The basic point made by the three philosophers we have considered is this: what is presented to us from the first-person point of view or the point of view of consciousness does not provide us with a sufficient basis for conceiving of ourselves *qua* the subject of thinking as part of the objective world. In Kantian terms, the point of view of the 'I think' does not provide the transcendental subject

with any basis for conceiving of itself as a part of either the sensible world or the intelligible world. Therefore, it seems that from the first-person perspective alone we cannot gain an *objective* conception of our own mind. Hence, all three philosophers problematize the availability of such a conception of mind when assuming the first-person perspective as the starting point.

Although Kant never considers this problem explicitly in in his writings, I want to argue that it is immanent to his philosophical system. As noted earlier, the point of view of consciousness—the point of view of the 'I think'—is the standpoint from which the Critical investigation is conducted and, to that extent, it is constrained to the resource available from within that standpoint. It is relatively clear that our ability to reflect upon, and gain access to, cognitive activities investigated in the critique must be rooted in our capacity for pure apperception, our capacity to become aware of our representations or thoughts simply in virtue of being the subject who thinks them.[17] The point of departure of the Critical investigation, then, is thoughts or representations *qua attributable to the I think* and, hence, *qua* presented to consciousness from my point of view. And, as we will see below, Kant's own characterization of this point of view is in complete agreement with the basic point made by the other philosophers we have considered. However, Kant's claims about mental faculties seem to presuppose an objective concept of mind—they seem to be claims about the character of the mind as an object. Such claims seem impossible unless the subject of thinking, the subject who conducts a critique of her own rational faculties, can conceive of herself as an object among other objects in the world.

[17] More on this in Chapter 1. See also Dieter Henrich, "Kant's Notion of a Deduction and the Methodological Background of the First Critique," in *Kant's Transcendental Deductions: The Three Critiques and the Opus Postumum*, ed. Eckart Förster (Palo Alto: Stanford University Press, 1989), 29–46; and Houston Smit, "The Role of Reflection in Kant's Critique of Pure Reason," *Pacific Philosophical Quarterly* 80, no. 2 (1999): 203–23.

It is unlikely that Kant should be unaware of this problem. Arguably, Kant's own criticism of Descartes in the Paralogisms covers the same ground as Williams's (Williams explicitly acknowledges Kant as his source). What is more, Kant regards his own thinking as significantly shaped by his engagement with Hume's work, and it is commonly agreed that his account of the I think is in part a response to Hume's rejection of the idea of the self.[18] So, it would seem incomprehensible that Kant should fail to recognize that the problem that Descartes faces is as much a problem of his own. A more plausible hypothesis is that Kant believes he has all the resources to address the problem and trusts his readers to put the pieces together by themselves.

Kant's account of pure apperception, on which his criticism of the Cartesian Cogito rests, reveals Kant's agreement with the view on the character of the first-person perspective we have previously sketched. Pure apperception is to be distinguished from empirical apperception (through inner sense), which is the empirical consciousness of oneself as an object. In this sense, pure apperception is distinctively first-personal—it is the consciousness of oneself as the *subject* (as opposed to an object) of thinking. This self-consciousness grounds the representation 'I' or 'I think' (the *cogito*, for Descartes), which must be able to accompany all the representations of which I am conscious (cf. B131-32). Therefore, in Kant's view, the unity of this self-consciousness (the "transcendental unity of apperception") is what grounds the unity of my point of view of consciousness. The fact that a given representation or a thought belongs to my point of view just consists in the fact that I must be able to ascribe it to myself, by attaching the 'I think' to it.

[18] Kant was supposed to have access to Hume's passages on the self through a German translation of James Beattie's *Essay on the Nature and Immutability of Truth*. See Robert Wolff, "Kant's Debt to Hume Via Beattie," *Journal of the History of Ideas* 21, no. 1/4 (1960): 117. For discussion of the relation between Kant's view on the self and Hume's, see Patricia Kitcher, "Kant's Cognitive Self," in *Kant's Critique of Pure Reason: Critical Essays*, ed. Patricia Kitcher (Lanham: Rowman & Littlefield, n.d.), 59–84.

However, the possibility of ascribing all thoughts to the same 'I', from the first-person perspective, does not entail the objective existence of a thinker.

> [The representation I is a] simple and in content for itself wholly empty representation […], of which one cannot even say that it is a concept, but a mere consciousness that accompanies every concept. Through this I, or He, or It (the thing), which thinks, nothing further is represented than a transcendental subject of thoughts = x, which is recognized only through the thoughts that are its predicates, and about which, in abstraction, we can never have even the least concept; […] the consciousness in itself is not even a representation distinguishing a particular object, but rather a form of representation in general, insofar as it is to be called a cognition. (A346/B404, cf. A381-82)

The important point in this convoluted passage is that the 'I' is "empty" in the sense that that it is devoid of any objective or metaphysically significant content. It is also merely "formal" in that is only expresses the formal unity of consciousness that cannot be identified with or explained by the metaphysical unity of a substance. For this reason, the 'I' is not a concept or a "representation distinguishing a particular object". Hence, one cannot use this representation (as the rational psychologists have attempted) as a "sole text" from which to derive a conception of the mind as a thinking substance or, for that matter, any objective conception of the mind. This point can be put differently by saying that the apperceptive awareness of our own thoughts or representations does not take us beyond the point of view of consciousness. In being aware of our thoughts in an apperceptive way, we are aware of them from an "engaged" or "an inhabitant's" point of view.[19]

Kant sometimes points to the gap between the way we are conscious of ourselves as subjects of thinking and the conception of the same subject as an object by referring to the incongruity between

[19] For a recent articulation of this thought, see Matthew Boyle, "Transparency and Reflection," *Canadian Journal of Philosophy* 49, no. 7 (2019): 1012–39; and Matthew Boyle, *Transparency and Reflection: A Study of Self-Knowledge and the First Person Perspective* (Oxford: Oxford University Press, Forthcoming).

the determining [bestimmend] and the determinable [bestimmbar] self. The self of which I am apperceptively conscious is the determining self and as such it can never become something that I determine (i.e., think or represent to myself) as an object through the categories.

> The **I think** expresses the act of determining my existence. The existence is thereby already given […]. For that self-intuition is required, which is grounded in an *a priori* given form, i.e., time, which is sensible and belongs to the receptivity of the determinable. Now I do not have yet another self-intuition, which would give the **determining** in me, of the spontaneity of which alone I am conscious, even before the act of **determination** […]. (B157-58n)

> I cannot cognize as an object itself that which I must presuppose in order to cognize an object at all; and […] the determining Self (the thinking) is different from the determinable Self (the thinking subject) as cognition is different from its object. (A402, cf. B407)

Kant does not deny that we have ways of representing the determinable self, i.e., representing ourselves as an object that can be determined through the categories. According to Kant, there are at least two ways in which we can represent ourselves as a determinable object, as an object of inner intuition and as a human being. However, the point is that the thinking subject—the I of the I think—cannot be identified with the object of any of these ways of representing. As soon as I try to represent myself in one of these ways, I no longer represent myself as *determining* but as *determined*. If this is the case, how can the thinking subject identify herself as an object, as seemingly required for the critique of her own faculties to be possible?

This difficulty has not been sufficiently appreciated by Kant scholars who assume that the I of pure apperception can problematically be taken to be a bearer of mental faculties.[20] While they rightly suppose that the critique is conducted from within the point of view of the I think and, hence,

[20] For an explicit statement of this view, see Chapter 8 of Stefan Heßbrüggen-Walter, *Die Seele Und Ihre Vermögen: Kants Metaphysik Des Mentalen in Der Kritik Der Reinen Vernunft* (Paderborn: Mentis, 2004).

that the thinking subject must be taken as the primary locus of cognitive activities that the critique sets out to analyze, they fail to recognize the tension between the non-objective conception of mind that this point of view can offer and the objective conception of mind that seems to be required by the Critical account of mental faculties. I insist that these two distinct conceptions be kept apart. Although the latter conception must involve the former conception in some way (the faculty of thinking that is being "critiqued" must somehow belong to the I think), an account needs to be given of the relationship between the two conceptions and how we can make available to ourselves the objective conception given that we start out with the non-objective conception.

## Reason as a Source of Concepts

The heart of the problem of the objective concept of mind is the gap between this concept and the non-objective way in which we are conscious of ourselves as subjects of thinking. As we have seen, Kant agrees with Hume, Wittgenstein, and Williams (against Descartes) that the point of view of consciousness (the point of view of the I think) is in a relevant sense empty—it is devoid of metaphysical content that can become the content of any objective concept of mind. The proposition 'I think *p*', asserted from the point of view of consciousness, does not signify *p* as an accident inhering in a thinking substance to which 'I' refers. Rather, it merely designates *p* as belonging to *one and the same* point of view of consciousness to which all my other representations also belong. For Kant, the unity and identity of the point of view of consciousness (the transcendental unity of apperception) is not equivalent or reducible to the metaphysical unity of a substance. Thus, it seems that from within the point of view of consciousness I cannot obtain a concept of myself *qua* the subject of thinking as an object. But if the Critical investigation of mental faculties is to be possible, such a concept must be available to us. This raises the question what the source of this concept is. As Kant himself has

instructed us, sorting out the source of origin of a given concept—its "birth certificate" [Geburtsbrief] (A86/B119)—will also reveal the justificatory basis of its employment.[21]

The discussion of the different sources of concepts in Kant usually focuses on the sources of the two classes of concepts that figure prominently in the Transcendental Analytic: empirical concepts and pure concepts of the understanding. What is common between these two classes of concepts is that they contain a certain unity or synthesis that can be encountered in experience.

> A concept that includes a synthesis in it is to be held as empty, and does not relate to any object, if this synthesis does not belong to experience, either as borrowed from it, in which case it is an **empirical concept**, or as one on which, as *a priori* condition, experience in general (its form) rests, and then it is a **pure concept** [of understanding], which nevertheless belongs to experience, since its object can be encountered in the latter. (A220/B267)

Thus, an empirical concept (with respect to its content or matter) has its source in experience. It is "borrowed" from experience of objects that fall under it and, therefore, applicable to them. In contrast, pure concepts of the understanding (the categories) have their source in the understanding as the faculty that prescribes conceptual form to experience in general. Although these concepts are not acquired from experience, they do not lack prerequisite relation to it. Because these concepts constitute necessary unity of any experience, they are universally applicable to all objects of experience.

However, I wish to argue that neither experience nor the pure understanding can be the source of the objective concept of mind. Why is it implausible to claim that the understanding is the source of the concept of mind? The concept of mind does not seem to belong to the same kind of concepts as the pure concepts of the understanding. These concepts represent general ontological features that apply to any object—they are "concepts of an object in general" (e.g., B128)—and this is why it makes

[21] E.g., A65-6/B90-1, A86-7/B119. For a helpful discussion of the connection between the origin of a concept and its deductive ground, see Henrich, "Kant's Notion of a Deduction."

sense to claim that they constitute necessary unity of experience of any object. The concept of mind, in contrast, is a concept of either an individual thing or a particular kind of objects. Therefore, the concept of mind must contain marks through which the objects that fall under it are distinguished from the other objects.[22] It is relatively clear that, for Kant, this sort of marks can only be derived from experience and not from the understanding *a priori*. Although the categories do not exhaust all the pure concepts of the understanding but only the primitive ones, the other pure concepts of the understanding, i.e., those that are derived from the categories (what Kant calls 'predicables'), are likewise concepts that represent general ontological features of objects rather than marks that distinguish one kind of objects from the other. There is no reason to suppose that these concepts can constitute concepts of objects of a particular kind in the absence of any empirical content.

One reason why it is implausible to take Kant's concept of mind to be an empirical concept is that doing so would imply that the critique must be either an empirical investigation of the mind or an *a priori* natural science based on an empirical concept, exemplified by the *Metaphysical Foundations of Natural Science*. The first option seems to be ruled out by Kant's insistence that the critique draws nothing from experience and, therefore, it is to be sharply distinguished from empirical psychology. The second option seems equally implausible given Kant's explicit rejection of such an *a priori* science of the mind. The reasons for rejecting these readings will be discussed more extensively in Chapter 3. For now, I want to point to another problem that would arise if take Kant's concept of mind to be an empirical concept, which is immediately connected with our central problem. An empirical concept of mind would be a concept of an object that can be determined as appearance and, hence, fall short of representing the subject of thinking *qua* determining. Such a concept would be structurally

[22] This is not to deny that the categories contain marks. But the marks of the categories must be ones that pertain to all objects and, hence, not ones through which objects of one kind are distinguished from objects of another kind.

analogous with the concept of our own body. We would form this concept on the basis of the experience of our mental states just as we form the concept of our body on the basis of the experience of its physical properties. Accordingly, this concept would be applied on the basis of experience. On this reading, the object that falls under the concept of mind is on a par with other objects of experience, such as cups, tables, trees, and other human beings. Like these other objects, my conscious access to my mind depends on its being given to me in (inner) intuition. To that extent, however, my relation to this inner object is radically different from my relation to myself as the transcendental subject. How do *I* (in the sense of the I of the I think) identify myself with the object that is represented by this empirical concept?[23] As long as the subject of apperception cannot be unproblematically identified with the body or the human being, it cannot be unproblematically identified with an empirically cognizable mental entity, either.

In this study, I propose that the origin of the objective concept of mind is neither experience nor the understanding but a third, often neglected, source of concepts according to Kant, namely, *reason*: "reason itself contains the origin of certain concepts and principles, which it derives neither from the senses nor from the understanding" (A299/B355). Concepts originating in reason are called 'transcendental ideas' or 'pure concepts of reason'. These make up the third species of concepts in the *Stufenleiter*:

> A concept is either an **empirical** or a **pure concept**, and the pure concept, insofar as it has its origin solely in the understanding (not in a pure image of sensibility), is called

[23] Hannah Ginsborg has raised a similar problem about how I (understood apperceptively) identify myself with the human being that is an object of both inner and outer sense. See Hannah Ginsborg, "The Appearance of Spontaneity," in *Self, World, and Art: Metaphysical Topics in Kant and Hegel*, ed. Dina Emundts (Berlin/Boston: De Gruyter, 2013), 119–44. In this paragraph, I focus on the problem of the identity between the apperceptive I and the mind (rather than the human being) as a proper object of inner sense (on the assumption that we have a concept of such an object).

> *notio*. A concept made up of notions, which goes beyond the possibility of experience, is an **idea** or a concept of reason. (A320/B377, original emphasis)

Although transcendental ideas presuppose the categories ("notions"), one cannot say that they have their origin in the understanding because they contain a distinctive kind of unity "which goes beyond the possibility of experience" and "of which the understanding has no concept". Kant distinguished this kind of unity from the "unity of understanding" [Verstandeseinheit] and calls it the "unity of reason" [Vernunfteinheit]. The unity of reason consists in "the absolute totality in the use of concepts", which is obtained by "[carrying] the synthetic unity, which is thought in the categories, all the way to the absolutely unconditioned" (A326/B383). Thus, reason generates a transcendental idea by "free[ing] a concept of the understanding from the unavoidable limitations of a possible experience and thus seek[ing] to extend it beyond the boundaries of the empirical" (A409/B435). Although the categorial unity, indeed, is rooted in the functions of the understanding, the extension of the categorial unity to its absolute totality is the work of reason.

The task of the critique of pure reason as "reason's self-knowledge" (Axi), which presupposes nothing beyond reason itself (cf. Axiv). It follows that reason must draw upon resources available from within itself for a conceptual framework to be used in the investigation of its own faculties. Admittedly, 'reason' here is to be understood in broad sense that includes the entire higher faculty of cognition. This suggests two possible sources for this conceptual framework, namely, the understanding and reason (in the narrow sense). But if I am correct to claim that the understanding cannot be the source of an objective concept of mind, then we are left only with the pure concepts of reason, which have their origin in "the nature of reason itself", being "generated in an entirely necessary way by reason according to its original laws" (A327/B384, A338-39/396-97).

Do we find any plausible candidate among the pure concepts of reason? Indeed. Kant's list of the pure concepts of reason does include a concept that qualifies as an objective concept of mind,

namely, the idea of the soul, which contains "the absolute (unconditioned) unity of the thinking subject" (A334/B391). Kant also explicitly states that this idea represents the thinking subject not as "the logical unity of every thought, in which I abstract from every object" but as "an *object* that I think" (A398, italics added). Further, since the soul as represented in this idea is "a simple substance, unchangeable in itself (identical in personality), standing in community with other real things outside it—in a word, the concept of a simple self-sufficient intelligence" (A682/B710), the idea of the soul represents an entity that can intelligibly be a bearer of faculties. At the same time, the idea of the soul is also connected to the subject of thinking insofar as it is "inferred" from "the transcendental concept of a subject that contains nothing manifold" (A340/B398). This inferential connection can explain how Kant's objective concept of mind is related to the subject of thinking. The idea of the soul, as a pure concept of reason, emerges as a promising solution to our problem.

An obvious objection to my proposal appeals to the fact that Kant himself censures the use of the idea of the soul in rational psychology. How is it possible for him to consistently employ the same concept in his own investigation? Moreover, Kant's criticism of rational psychology seems to be directed not only at a specific use of the concept also at the concept itself: the idea of the soul is generated by *dialectical* inferences, which are fallacious.[24] These inferences (i.e., the paralogisms) give rise to the marks of the idea of the soul (substantiality, simplicity, identity), without which it would have no specific content.[25]

---

[24] E.g. A402, B411.

[25] "Now at least the transcendental (subjective) reality of pure concepts of reason rests on the fact that we are brought to such ideas by a necessary syllogism. Thus there will be syllogisms containing no empirical premises, by means of which we can infer from something with which we are acquainted to something of which we have no concept, and yet to which we nevertheless, by an unavoidable illusion, give objective reality. In respect of their result, such inferences are thus to be called sophistical rather than rational inferences; even though they might lay claim to the latter term on account of what occasions them, because they are not thought up, nor do they arise contingently, but have sprung from the nature of reason. They are sophistries not of human beings but of pure

In response to this objection, I want to note that, for Kant, transcendental illusion has a purpose that accords with the teleology of reason. While Kant acknowledges the illusory nature of the transcendental ideas, Kant argues that they must have a purposive use that accords with the end of reason. The reason is that the ideas are "just as natural to [reason] as the categories are to the understanding", and "everything grounded in the nature of our powers must be purposive and consistent with their correct use" (A462/B670, cf. A669/B697). The ideas have at least three purposive uses. First, these ideas have a legitimate theoretical use, which Kant calls the 'regulative use', in which they, "in a fundamental and unnoticed way, serve the understanding as a canon for its extended and self-consistent use" (A392/B387). Through the regulative use of the ideas, reason prescribes to the understanding "the direction toward a certain unity of which the understanding has no concept, proceeding to comprehend all the actions of the understanding in respect of every object into an absolute whole" (A326-27/B383). Second, in the practical context, these very same ideas serve as constituents of the postulates of pure practical reason, through whose validity the reality of the ideas receives practical confirmation.[26] Roughly speaking, these postulates are beliefs that are constitutive of practical reason as such: they are "theoretical propositions" that are nonetheless "inseparably connected with the moral law" (*KpV* 5:122). Third, because the transcendental ideas originate from theoretical reason and obtain their objective reality through practical reason, they "make possible a transition from concepts of nature to the practical" (A329/B386-87, cf. A798/B826ff.). In this way, these ideas manifest the practical end of reason even within the theoretical sphere and, hence, reveals the unity between theoretical and practical reason.

---

reason itself, and even the wisest of all human beings cannot get free of them; perhaps after much effort he may guard himself from error, but he can never be wholly rid of the illusion, which ceaselessly teases and mocks him" (A339/B397).

[26] On the identity of the theoretical and practical ideas, see B395n and *KpV* 5:134ff.

The key to resolving the tension between the dialectical origin of the transcendental ideas and their purposive use, I suggest, lies in Kant's distinction between illusion [Schein] and deception [Betrug]. Transcendental illusion has a peculiarity such that it "does not cease even though it is uncovered and its nullity is clearly seen into by transcendental criticism" (A297/B353). However, it does not necessarily follow from the fact that the *illusion* lingers after it has been exposed that we must be *deceived* by it. Although the pure concepts and principles of reason continue to "look entirely like objective principles", we need not *judge* that they represent "the determination of things in themselves". In Kant's analogy, the sea unavoidably appears to us higher than the shore; nonetheless, knowing the optical cause of this illusion, we can avoid judging and, hence, being deceived that the sea is actually higher than the shore. Thus, we can admit that the transcendental ideas are derived from an illusion and that whatever use they may have cannot be detached from this illusion because this is what gives them content. Nevertheless, as long as we are mindful of their illusory nature and not deceived by their objective look, we can put them to appropriate use. In other words, there is no contradiction in holding that purposive uses of the transcendental ideas involve *illusion* although not *deception*—none of these uses requires us to judge the ideas to be objectively valid as determination of things in themselves.[27] In this way, I argue, we allay the worry that Kant might be inconsistent in making use of the idea of the soul in his own philosophical inquiry.

[27] The case of the deployment of the ideas in the practical postulates is somewhat complicated. However, Kant is positive that "there was […] no extension of the cognition *of given supersensible objects*, but there was nevertheless an extension of theoretical reason and of its cognition with respect to the supersensible in general, inasmuch as theoretical reason was forced to grant *that there are such objects*, though it cannot determine them more closely and so cannot itself extend this cognition of the objects […]" (*KpV* 5:135). Cf. "For through [practical reason], which for the first time reveals to me the consciousness of the moral law, I would indeed have a principle for the determination of my existence that is purely intellectual; but through which predicates? Through none other than those that would have to be given to me in sensible intuition, and thus I would have landed right back where I was in rational psychology, namely in need of sensible intuitions in order to obtain significance for my concepts of the understanding, substance, cause, etc. […]" (B431).

As we have seen, Kant attaches tremendous significance to the ideas of reason within the Critical system as a whole, despite that this significance is often inadequately or obscurely articulated. In this study, I argue that the importance of the ideas of reason for Kant's philosophy reaches far beyond what is commonly understood. More specifically, the idea of the soul underlies the possibility of Critical philosophy itself. As I have been arguing, the idea of the soul *is* the concept of mind that underlies Kant's claims about mental faculties.

## The Primacy of Explanatory Understanding

The problem of the gap between the first-person perspective and the objective concept of mind has led us on a search for the source of the latter concept. I have argued that this source lies neither in experience nor in the understanding. However, reason, as the third source of concepts in Kant's philosophy, comes forward as a promising answer. In particular, the idea of the soul, one of the pure concepts of reason, stands out as a concept of an entity that is qualified as a possible bearer of faculties and appropriately related to the transcendental subject of thinking. To further demonstrate the plausibility of this reading, we must show that the role of the concept of mind in transcendental philosophy aligns itself with one of the positive uses Kant ascribes to the transcendental ideas. To this end, I will argue that the employment of the idea of the soul in transcendental philosophy corresponds to what Kant calls the regulative use. This calls for a revision of the standard interpretation of the regulative function of the idea of the soul, which I will offer in Chapters 4 and 5.

My revisionist account revolves around the distinction between two kinds or degrees of cognition, which Kant terms "understanding" [Verstehen] and "comprehension" [Begreifen] (cf. A311/367). The difference between the two degrees of cognition corresponds to the distinction between two kinds of unity we encountered earlier: the unity of understanding and the unity of reason. "Understanding", characterized by the unity of understanding, is what we might think of as a body of

cognition constituted by an aggregate of judgments. In order to become "comprehension", characterized by the unity of reason, cognitions must be systematically connected in such a way that they yield explanatory understanding—the understanding *why* something is the case. In this way, Kant distinction roughly corresponds to the distinction in contemporary epistemology between knowledge and understanding.[28]

My interpretation takes its cue from Kant's claim that whereas the categories are constitutive of the unity of understanding, the transcendental ideas are constitutive of the unity of reason. I understand this claim as saying that just as the categories dictate the forms of empirical judgment, so the transcendental ideas dictate the forms of explanatory understanding. This proposal will be discussed in greater detail in Chapter 4.

For the present purposes, it is important to note the following. First, the unity of reason is irreducible to the unity of understanding. Hence, for Kant, explanatory understanding is irreducible to mere cognition. Possessing cognition of many propositions about a given subject matter does not guarantee that one *understands* that subject matter. For Kant, comprehension involves a higher-order structural feature (captured by the notion of systematic unity) that can only be given from top down, i.e., through reason *a priori*, and not from bottom up, i.e., derived from experience. Second, although explanatory understanding presupposes lower-level cognition, the demand for explanatory understanding can outrun our capacity for cognition—reason can make explanatory demand that cannot be fulfilled given the limits of our cognition. This can be seen most clearly in the case of dogmatic metaphysics, in which reason is driven to overstep its own boundaries by an impulse toward explanatory understanding that cannot be met by cognition. Third, even when reason's explanatory demand cannot be met due the limit of our cognition, this demand itself is nonetheless justified, at

[2828] For discussion, see John Bengson, "The Unity of Understanding," in *Making Sense of the World: New Essays on the Philosophy of Understanding*, ed. Stephen Grimm (Oxford: Oxford University Press, 2017), 14–53.

least subjectively.[29] Kant thinks that this justified demand of reason goes hand in hand with an (at least subjectively) valid presupposition that the object complies with this demand. In other words, reason is justified in presupposing that reality is so constituted such that a *complete* explanation of any aspects of it that stands in need of explanation is possible (though, as I argue in Chapter 4, for rational beings whose intellect is not limited by the conditions of sensible intuition).

On the reading I propose, the regulative function of the transcendental ideas consists precisely in dictating what counts as an explanation for a given aspect of reality that stands in need of explanation. In this sense, the transcendental ideas dictate *a priori* the forms of explanatory understanding (hence, there are exactly three distinct forms of explanatory understanding corresponding to the three ideas). Therefore, the cognitive role of the ideas parallels that of the categories, which dictate *a priori* the forms of empirical judgment.

As I will argue in Chapter 1, Kant's Critical philosophy has an explanatory agenda, that is, to explain the *a priori* cognizable aspect of reality. I will contend that the form of explanatory understanding that the critique assumes is the one grounded in the idea of the soul: explanation in terms of the subject's representational faculties. This type of explanation is only possible under the presupposition of a certain metaphysical conception of the subject of thinking, which is precisely what is made possible by the idea of the soul. Hence, reason's explanatory demand with respect to the *a priori* cognizable aspect of reality is what entitles, indeed enjoins, the thinking subject to conceive of herself according to the idea of the soul—as a thinking substance endowed with representational faculties. On the basis of this self-conception, the thinking subject is in a position to make sense of the *a priori* features of the objects she experiences.

[29] I say 'at least' because Kant sometimes speaks (rather cryptically) of the "indeterminate objective validity" of the presuppositions of reason (cf. A663/B691, A669/B697, A680/B708).

An important upshot of my interpretation is that the general conception of mind that underlies Kant's Critical account of mental faculties do not rest on *cognition* of any entity. In this way, my reading departs from the realist readings (examined in Chapter 3), which holds that Kant's claims about the mind are claims to cognition of a noumenal or a phenomenal entity. My reading also diverges from the anti-metaphysical reading (discussed in Chapter 2) in that it acknowledges the metaphysical import of these claims, while denying them the status of cognition of an object.[30] Rather, Kant's employment of an objective concept of mind is justified by reason's demand for explanatory understanding. Unless we make use the idea of the soul and conceive of ourselves in accordance with it, we would not be in a position to make sense of the fact that objects owe certain *a priori* features to our way of representing them. This, I argue, is Kant's solution to the problem of the objective concept of mind. The origin of this concept lies in the Cartesian error of mistaking the formal unity of self-consciousness to be the objective or metaphysical unity of a thinking substance. This illusion is natural and unavoidable. However, there is also a point (a purpose) to it, since reason makes a demand for explanatory understanding the fulfilment of which is only possible (though never entirely) through the use of a concept originating from that illusion.

Strawson rightly observes that "whatever necessities Kant found in our conception of experience he ascribed to the nature of our faculties".[31] In other words, Kant *explains* the *a priori* elements in the general structure of experience in terms of the specific character of our faculties. However, for Strawson, it is a mistake to think that the *a priori* structural features of experience call for or admit of explanation in the first place: "there is no sense in the idea that we might look to facts

[30] Nevertheless, I allow that these claims may still be understood as what Kant calls "formal" cognition. See Chapter 2.

[31] Strawson, *The Bounds of Sense*, 19.

altogether outside our experience in order to find an explanation of there being this or that limit".[32] According to Strawson, we establish that experience necessarily exhibits these features by analyzing our conception of experience, which is self-validatory in the sense that no part in it can be intelligibly negated. But Kant clearly does not regard conceptual analysis as the basic philosophical method of the critique.[33] Moreover, he thinks it is conceivable *in general* that experience might be otherwise than it is and that things in themselves are different than experience presents them to be (although it might well be the case, as Strawson insists, that we are not in a position to conceive of experience in which a *particular* feature is absent). Precisely for this reason, the fact that experience has necessary features that we can cognize *a priori* stands in need of an explanation. Thus, for Kant, the necessary features of experience do not coincide with the limits of intelligibility (as they do for Strawson) but are *within* those limits as part of what is to be made intelligible.

In this way, transcendental idealism and the conceptual framework supplied by the idea of the soul co-operate to deliver Kant's explanation of the possibility of *a priori* cognition of necessary features of experience *a priori* as well as the limits of experience. Transcendental idealism says that certain aspects of reality are mind-dependent. This allows us to explain why these aspects of reality are cognizable *a priori* but also why we cannot cognize things in themselves. However, the idea that certain aspects of reality are dependent on our mind presupposes that we have a way to represent our mind as a part of the objective order. For how else can we understand this notion of dependence if not as *metaphysical* dependence? The difficulty, as we have seen, is that the conception that the subject of thinking can obtain of herself from within the first-person point of view does not present that subject as something that can enter any metaphysical relation whatsoever. The idea of the soul solves this problem by offering a way to conceive of our own mind that meets exactly those requirements.

[32] Strawson, 271.

[33] See the Transcendental Doctrine of Method, especially at A735-38/B763-66.

On this reading, then, the explanatory demand that motivates the critique is rooted in reason's global explanatory demand with regard to all aspects of reality that stand in need of explanation. In the case of the critique, this demand is directed at the aspect of reality that can be known to us *a priori*, giving rise to the "general problem of pure reason": How are synthetic *a priori* judgments possible? Reason's demand for explanatory understanding is what ultimately entitles us—makes it subjectively necessary for us—to conceive of ourselves according to the idea of the soul. Nevertheless, this subjectively necessary self-conception can never lay claim to cognition of what we are in ourselves.

# Chapter 1

# The Bounds of Reflection

## 1.1 Introduction

We might hope to gain some clarity on the question of Kant's concept of mind and what justifies him in employing this concept to make substantive claims about mental faculties by examining the methodology and the epistemology of the critique. As we have seen, Strawson's famous rejection of Kant's "transcendental psychology" is grounded in a methodological and epistemological consideration: Kant's propositions about mental faculties cannot be justified either by empirical psychology or by conceptual analysis. From this, he concludes that the whole domain of discourse on the mind which Critical philosophy engages in is completely "imaginary". One way to resist Strawson's objection on Kant's behalf is, therefore, to correct the methodological and epistemological assumption on which it is based.

Some commentators have sought to defend Kant along this line, arguing that the two methodological prongs of the objection do not exhaust the alternatives. Kant's method relies on neither empirical psychology and conceptual analysis. His claims about mental faculties are grounded, rather, in reflection or consciousness we have of our own rational faculties and their activity.[1] The

[1] Dieter Henrich, "Kant's Notion of a Deduction and the Methodological Background of the First Critique," in *Kant's Transcendental Deductions: The Three Critiques and the Opus Postumum*, ed. Eckart Förster (Palo Alto: Stanford University Press, 1989), 29–46; Houston Smit, "The Role of Reflection in Kant's Critique of Pure Reason," *Pacific Philosophical Quarterly* 80, no. 2 (1999): 203–23; Matthew Boyle, "Kant on Categories and the Activity of Reflection," in *Palgrave Handbook of German Idealism and Analytic Philosophy*, ed. James Conant and Jonas Held (London: Palgrave Macmillan, Forthcoming); Karl Schafer, "Kant: Constitutivism as Capacities-First Philosophy,"

cognitive access we have to these faculties is explained by the fact that *rational* faculties are inherently self-conscious. On this proposal, we have implicit awareness or knowledge of the laws governing the activity of our rational faculties simply in virtue of possessing and exercising them. The task of philosophy is to make those laws explicit. Call this account of Kant's basic philosophical method 'the reflectivist account'.

According to the reflectivists, consciousness of our own rational faculties explains the source of not only particular claims about the character of these faculty in the crtique but also its general framework of faculty psychology. This framework presupposes a non-trivial, metaphysically substantive conception of mind as the metaphysical bearer of faculties.[2] Hence, the reflective account can provide an adequate account of this general framework only if it can explain how we are justified in talking about our own mind in metaphysically substantive terms. In this chapter, the assumption that the reflectivist account can offer this explanation will be my target. While I do not deny that the reflective account offers a plausible account of how at least some of Kant's *particular* claims about mental faculties are justified, I do not think that it can explain the general framework underlying those particular claims. In this respect, the reflectivist account is severely incomplete.

But should this worry us? After all, the reflectivists might respond that Kant's goal is not at all to give a metaphysical account of the mind. Perhaps, all he intends to do is to make explicit the normative principles governing our faculties so that we do not overstep them (this is indeed what is suggested by Kant's description of the critique of pure reason as a "treatise on method", which sets

*Philosophical Explorations* 22, no. 2 (2019): 177–93; Karl Schafer, "Transcendental Philosophy As Capacities-First Philosophy," *Philosophy and Phenomenological Research* 103, no. 3 (2021): 661–86; Thomas Land, "Epistemic Agency and the Self-Knowledge of Reason: On the Contemporary Relevance of Kant's Method of Faculty Analysis," *Synthese* 198, no. 13 (Supplement) (2021): 3137–54.

[2] For further discussion, see Chapter 2.

down a "canon" for the correct use of reason).[3] If it is right to think of the critique as a "user guide", as it were, then, the reflectivists might argue further, there is no need to delve into the detail about the nature and the mechanics of the device (i.e., the faculties) more than necessary for operating it. In this sense, the metaphysical concepts are used exclusively as means for making pragmatically relevant distinctions, not for elucidating the metaphysical nature of reason.

In response, I will argue that this suggestion misses an important aspect of Kant's philosophical project. While the critique undeniably has prescriptive implications, it also has an explanatory agenda.[4] This can be discerned most clearly in its central problem: "How are synthetic *a priori* judgment possible?" In my view, this question can only be understood as a question demanding an explanation for the possibility of synthetic *a priori* judgment. Moreover, the laws governing our rational faculties can fulfil their explanatory role with respect to this question only against the background of a specific understanding of what these faculties *are*, metaphysically speaking. More specifically, the appeal to principles governing them can have the requisite explanatory power only if the faculties themselves are understood as law-governed grounding relations between a substance and its accidents. Such background understanding of our own faculties, however, cannot be delivered by the reflective consciousness we have of them, or so I will argue. I will turn first to the claim that the critique of pure reason has an explanatory agenda.

## 1.2 Critical Philosophy as an Explanatory Project

The explanatory agenda of the critique has not been sufficiently appreciated by commentators. I want to argue that this explanatory agenda is implicit in Kant's familiar characterization of the critique

[3] E.g. A12/B26, A83/B109.

[4] The relationship between the critique's prescriptive and explanatory agendas is an important and interesting question that cannot be addressed here.

as a *science* [Wissenschaft]. The critique of pure reason is supposed to be a science or at least a "propaedeutic", a beginning, of a new science that is yet to be fully realized through the completion of a "system of pure reason" (A841/B869) or a "future system of metaphysics" (Bxxxvi, cf. the full title of *Prolegomena*).[5] What is a science for Kant?

Kant's official definition of a science is a body of cognitions that constitutes a *system* as opposed to a mere *aggregate* (cf. *JL* 9:72): "I understand by a system […] the unity of the manifold cognitions under one idea. This is the rational concept of the form of a whole, insofar as through this the domain of the manifold as well as the position of the parts with respect to each other is determined *a priori*" (A832/B860). In the *Metaphysical Foundations*, Kant further distinguishes between a science in general and a "proper" science or science "in the strict sense".

> Any whole of cognition that is systematic can, for this reason, already be called *science*, and, if the connection of cognition in this system is an interconnection of grounds and consequences, even *rational* science. If, however, the grounds or principles themselves are still in the end merely empirical, as in chemistry, for example, and the laws from which the given facts are explained through reason are mere laws of experience, then they carry with them no consciousness of their *necessity* (they are not apodictally certain), and thus the whole of cognition does not deserve the name of a science in the strict sense. (*MAN* 4:468)

Thus, for a science to be a science properly so-called, it must, in addition to being systematic, satisfy two further conditions: (1) it must qualify as what Kant calls 'rational cognition' [rationale Erkenntnis, Vernunfterkenntnis], that is, cognition from grounds or principles.[6] What this means is that the connection of cognition contained in the system must be "interconnection of grounds and consequences"; (2) the grounds or principles themselves must be *a priori* and apodictically certain. It

[5] E.g. Bxxii, Bxxiv, Bxliii, B6, A11/B24-25, A14-16/B28-30, A21/B35-36, A57/B81, A758/B787, A839/B867, *Prol.* 4:325, *KpV* 5:8, *KpV* 5:89, R4148 17:434.

[6] Cf. A713/B714ff., A836/B864.

is relatively clear that Kant thinks Critical philosophy is a proper science that satisfies both conditions, in addition to being systematic.[7] Kant regards philosophy in general as a paradigm case of rational cognition, cognition from grounds or principles.[8] He also often stresses the absolutely *a priori* character of the Critical investigation, which implies that none of its principles is empirical. In this regard, the critique stands in contrast with the Lockean "physiology of the human understanding" (Axi), which proceeds from empirical principles.

As some commentators have noted, Kant understands 'ground' [Grund] or 'principle' [Prinzip] (as in 'cognition from grounds or principles') in the sense of an *explanatory* ground.[9] Accordingly, the connection between a ground and a consequence, in which the connection of cognition in a scientific system consists, is to be understood as an explanatory relation: a cognition that is a ground or principle must in some way explain another cognition that is its consequence. Note that Kant's use of these terms is often ambiguous: a ground or principle can be *either* a real entity that metaphysically grounds and explains another entity *or* a proposition stating a fact that explains another fact. Kant sometimes disambiguates the latter sense by opting for the term 'Grundsatz' instead of 'Grund'. In both cases, however, the relation that is at issue is, fundamentally, an explanatory relation. In this way, the Kantian notion of rational cognition coincides with the traditional Aristotelian notion of '*a priori* knowledge' as knowledge from explanatory grounds.[10] In this traditional sense, we know something *a priori* just in case we know *why* it exists or *why* it is the case through its cause or

[7] On the systematic character of Critical philosophy, see, e.g. A10-14/B23-28, A64-65/B89-90.

[8] A713/B741ff., A836-37/B864-65, *MAN* 4:469.

[9] See Houston Smit, "Kant on Apriority and the Spontaneity of Cognition," in *Metaphysics and the Good: Themes From the Philosophy of Robert Merrihew Adams*, ed. Samuel Newlands and Larry Jorgensen (Oxford: Oxford University Press, 2009), 188–251; Schafer, "Transcendental Philosophy," 3–4.

[10] On the persistence of the Aristotelian notion of the *a priori* in Kant, see Desmond Hogan, "Three Kinds of Rationalism and the Non-Spatiality of Things in Themselves," *Journal of the History of Philosophy* 47, no. 3 (2009): 355–82; Smit, "Kant on Apriority and the Spontaneity of Cognition."

metaphysical ground. In some metaphysics lecture notes from the Critical period, Kant describes philosophy as *a priori* cognition explicitly in this traditional sense.[11] Hence, for Kant, to characterize a proper science as a system of rational cognition, whose internal connection consists in the connection between grounds and consequences, amounts to saying that a science in the strict sense must be explanatory with respect to its subject matter.

Consequently, if Critical philosophy is rightly to be called a proper science as well as rational cognition, it must be explanatory or have an explanatory agenda. What, then, does the critique seek to explain? An answer is given by its central question: 'How are synthetic a priori judgments possible?', the "real" and the "general" problem of reason (B19; cf. A10). According to this problem, the overarching goal of the critique as an explanatory enterprise is to explain the possibility of synthetic *a priori* judgments. This problem is *general* because "one can bring a multitude of investigations" under it. In the first *Critique* alone, these investigations span three different domains of synthetic *a priori* judgments: mathematics, pure natural science, and metaphysics. The subsequent *Critiques* are also guided and unified by the same problem. Each Critique focuses on one of the three basic faculties of the mind, all of which are sources of representations that can relate to objects. Further, in virtue of being subordinated to reason, these faculties can all be sources of *a priori* judgments about objects (cf. *EE* 20:241; *KU* 5:168). This "discovery", as Kant puts it in a 1787 letter to Reinhold[12], gives rise to three different how-possible questions. The first *Critique* is dedicated to the question of the possibility of synthetic *a priori* cognitions of objects insofar as they are objects of our faculty of cognition. The second *Critique* considers the possibility of synthetic *a priori* cognitions of objects insofar as they are objects of our faculty of desire. In the third *Critique*, Kant turns to the possibility of synthetic *a priori* cognitions of objects insofar as they are objects of our faculty of feeling. In this sense, the task of

[11] Cf. *MM* 29:747-50, *ML*$_2$ 28:531.

[12] *BW* 10:514-15.

explaining the possibility of synthetic *a priori* cognitions is what unifies Kant's Critical philosophical project as a whole.

For Kant, what it means to explain the *possibility* of synthetic *a priori* judgments is to explain their *objective validity*—to explain how these judgments can be related to or be about an object (cf. A57/B81). Thus, we can trace the Critical formulation of the problem of pure reason in terms of synthetic a priori judgments back to its precursor in Kant's famous 1772 letter to Markus Herz:

> What is the ground of the relation of that in us which we call 'representation' to the object? […] The passive or sensuous representations have an **understandable** relationship to objects, and the principles that are derived from the nature of our soul have an **understandable** validity for all things insofar as those things are supposed to be objects of the senses. Similarly, […] if the object itself were created by the representation […], the conformity of these representations to their objects could also be **understood**. Thus the possibility of both an *intellectus archetypus* […] and an *intellectus ectypus* […] is at least **comprehensible**. However, our understanding, through its representations, is neither the cause of the object […] nor is the object the cause of our intellectual representations in the real sense (*BW* 10:130, boldface added).

The language Kant uses in this passage unmistakably indicates that the critique is motivated by an explanatory concern. The problem being raised here is the following. In contrast to representations that are passively caused by objects and representations that actively cause their objects, the relation of the *a priori* representations of the understanding to objects, which fall into neither of those categories, seems incomprehensible. In other words, it seems that we cannot *explain* how these representations can relate to objects. In the first *Critique*, Kant reformulates the same problem in broader terms: 'representations of the understanding' and 'intellectual representations' become synthetic *a priori* judgments. But the concern remains the same: How do we explain the relation of

synthetic *a priori* judgments to objects, given that these judgments neither are passively caused by objects nor actively bring objects into existence (as in the case of the divine or archetypal intellect)?[13]

The objective validity of synthetic *a priori* judgments is puzzling precisely because they are both 'synthetic' and '*a priori*'. Consider, first, how the question of objective validity is at stake in synthetic judgments in a way that is different from analytic judgments.[14] Both synthetic and analytic judgments can relate to an object. In the case of analytic judgment, however, the relation of the constituent concepts to an object is irrelevant for the connection of concepts in it. The connection between the subject concept and the predicate concept is asserted simply on the basis of the fact that the predicate concept is contained in the subject concept. In synthetic judgments, by contrast, since the connection between the subject concept and the predicate concept cannot be established by analyzing the subject concept, this connection must be grounded in and mediated by the relation of both concepts to the same object. Thus, the relation of constituent concepts to an identical object is a condition for the assertion of the subject-predicate connection in a synthetic judgment.

Now, concepts can only be related to an object that can be *given* to us, that is to say, an object with which we can be acquainted or immediately presented in consciousness.[15] Hence, in making a synthetic judgment, we effectively assert that an object can be *given* that jointly exemplifies the concepts combined in that judgment.[16]

---

[13] Cf. A92/B124-25.

[14] The difference between analytic and synthetic judgments with regard to their relation to an object has been emphasized by Henry Allison, *Kant's Transcendental Idealism*, 2nd ed. (New Haven: Yale University Press, 2004), 91–95.

[15] The significance of the 'givenness' of objects for the Kantian how-possible question is emphasized by Lucy Allais, *Manifest Reality: Kant's Idealism and His Realism* (Oxford: Oxford University Press, 2015), 190–94.

[16] This is of course true only if we make the judgment with the intent that it amounts to cognition. However, I take it that this is the only sense of "making a judgment" that Kant is interested in in raising the how-possible question.

Since our understanding is discursive rather than intuitive, an object can only be given to us through receptivity of the senses and, hence, not *a priori*: "An understanding, in which through self-consciousness all of the manifold would at the same time be given, would **intuit**; ours can only **think** and must seek the intuition in the senses" (B135). Kant defines an intuition as a representation that relates immediately to an object and, therefore, it is the species of representation through which objects are given to us. As Kant is pointing out, an intuitive understanding would actively bring about an intuition and thereby give itself an object. For a discursive understanding like ours, however, an object can only be given in an intuition passively by an object affecting our senses. Therefore, for us, no object can be given *a priori*. This is why the objective validity of synthetic judgments that are *a priori* seems inexplicable.

> It is thus experience on which the possibility of the synthesis of the predicate [concept] with the [subject concept] is grounded, since both concepts, though the one is not contained in the other, nevertheless belong together, though only contingently, as parts of a whole, namely experience, which is itself a synthetic combination of intuitions. But in synthetic *a priori* judgments this means of help is entirely lacking. If I am to go beyond the concept A in order to cognize another as combined with it, what is it on which I depend and by means of which the synthesis becomes possible, since I here do not have the advantage of looking around for it in the field of experience? (A8-9/B12-13)

Objects can only be given to us in experience and, hence, through sensibility. The puzzle here is that *prima facie* synthetic *a priori* judgments can relate to objects only if objects can be given to us independently of sensibility.[17] It is this puzzle that Copernican hypothesis responds to: "Let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of an *a priori*

[17] Recall that part of Kant's solution is to claim that sensibility has an *a priori* form, in virtue of which we can cognize some facts about objects independently of their being actually given to us.

cognition of them, which is to establish something about objects **before they are given to us**" (Bxvi, boldface added).

As we have seen, Kant's fleshes out this explanatory puzzle in terms of the objective validity of a certain class of judgments or, more specifically, cognitions. Since the objective validity of these judgments depends on *objects* being given to us that instantiate certain properties (properties corresponding to the connections of concepts asserted in the judgments), to explain how synthetic *a priori* judgments can be objectively valid amounts to explaining how given objects can instantiate properties that we can cognize *a priori* ("before they are given to us"). In this sense, the primary explanandum of the critique has a dual aspect. On the one hand, the critique seeks to explain the objective validity of synthetic *a priori* judgments, which are *representations* of objects. To do so is, on the other hand, to explain how *objects* can have properties that are attributable to them *a priori*. Kant is pointing to the dual aspect of the explanandum of the critique when he says the following.

> Synthetic *a priori* judgments are possible, if we relate the formal conditions of *a priori* intuition, the synthesis of the imagination, and its necessary unity in a transcendental apperception to a possible cognition of experience in general, and say: The conditions of the **possibility of experience** in general are at the same time conditions of the **possibility of the objects of experience**, and on this account have objective validity in a synthetic judgment *a priori*. (A158/B197, original emphasis)

Note that this dual characterization does not apply to synthetic *a priori* judgments in the theoretical domain only. Kant defines theoretical cognition as cognition of *what exists* and practical cognition as cognition of *what ought to exist* (e.g., A633/661). As a result, we may understand the theoretical part of Critical philosophy to be concerned with *a priori* judgments about properties objects of experience *have*, whereas the practical part is concerned with *a priori* judgments about properties objects of experience *ought to have*.

The dual aspect of Kant's explanatory agenda helps to elucidate the significance of transcendental idealism for the mode of explanation deployed in Kant's Critical philosophy. Transcendental idealism must be presupposed for this mode of explanation to be possible. Kant frequently stresses that we cannot *explain* the objective validity of synthetic *a priori* propositions, such as mathematical propositions and the principles of pure understanding, unless we assume that spatiotemporal objects are mere appearances or mind-dependent representations in us and not things in themselves.[18] Thus, one of Kant's arguments against transcendental realism is that it forecloses the possibility of explaining the possibility of synthetic *a priori* judgments: we could not explain the *a priori* claims that permeate our cognition if the objects of our cognition were things in themselves. In this way, the availability of the specific mode of explanation that Kant employs to explain the possibility of synthetic *a priori* judgments seems to presuppose transcendental idealism.

The presupposition that objects of experience, insofar as their sensible and intellectual forms are concerned, are mind-dependent representations opens the way for the explanatory strategy that Kant employs throughout the three *Critiques*: the *a priori* cognizable properties of objects are explained by appeal to the relation of metaphysical dependence between these properties and the subject's mental faculties. Thus, Kant's explanation why these properties are attributable to objects *a priori* is that these properties are metaphysically grounded in our mental faculties. In the first *Critique*, Kant argues that the spatiotemporality and the categorial properties of objects are grounded in our faculty of cognition and explained by the character of this faculty. In the second *Critique*, he argues that the properties that objects *ought* to have (the properties that moral maxims enjoin us to realize) are grounded in our higher faculty of desire and explained by its character. In the third *Critique*, Kant seeks

[18] E.g., A40-41/B57-58, A128-29, B159-60.

to explain the aesthetic and teleological features we ascribe to objects by appeal to the character of the higher faculty of feeling of pleasure and displeasure as the metaphysical ground of these features.

In this way, Kant's appeal to mental faculties as the metaphysical grounds of properties attributable to the objects in virtue of their transcendental ideality explains why the objects of our faculties instantiate those properties. However, this constitutes only one part of the explanation Kant needs. An adequate explanation of the possibility of synthetic *a priori* judgments must also explain how we can have *a priori* cognitive access to those faculties and the grounding relations they stand to objects. Kant's strategy shifts the explanatory focus from the *a priori* cognition of things to the *a priori* cognition of our own faculties. Therefore, in order to complete his explanation of the possibility of synthetic *a priori* judgments, Kant needs to explain how it is possible for us to cognize our faculties *a priori*. This is the issue that the reflectivist account aims to address.

## 1.3 The Explanatory Role of Mental Faculties

I have argued that the critique has an explanatory task, captured by the question of the possibility of synthetic *a priori* cognitions. The specific mode of explanation adopted by Kant involves appeal to mental faculties, which highlights the significance of transcendental idealism as a presupposition of this mode of explanation. Now, if Kant's appeal to mental faculties is to play an explanatory role with respect to the possibility of certain synthetic *a priori* judgments, they must fulfill at least two conditions. First, the appeal to faculties in general is explanatory insofar as it explains a certain effect as a specific case of a general rule. Therefore, the characteristic activity of the mental faculties that Kant appeals to must be characterizable in terms of general rules or principles. Call this the 'generalizability condition'. Second, these principles must be epistemically accessible to us in a manner that is independent of empirical observation. Call this the 'accessibility condition'. In this

section, I will elaborate on how mental faculties, according to Kant, fulfil the generalizability condition. The accessibility condition will be our focus in the next section.

The generalizability condition requires that the activity of an explanatorily relevant faculty be characterizable in terms of general rules. An appeal to a faculty would not be explanatory with respect to an effect it produces, if the faculty produces that effect in a one-off fashion. Unless it is a generalizable fact that a faculty *X* brings about a property *F* in an object *a* standing in a suitable relation to *X*, an appeal to *X* would not explain why *a* is *F*. Thus, an appeal to *X* is explanatory with respect to the fact *Fa* precisely because it picks out a generalizable dependence relation between *X* and *F*.

Kant often claims that faculties and powers in general operate according to rules, regardless of whether they belong to animate or inanimate, rational or non-rational beings: "Everything in nature, both in the lifeless and in the living world, takes place *according to rules*. […] The exercise of our powers also takes place according to certain rules […]" (*JL* 9:11, cf. A294/B350, *GMS* 4:412). But what is the reason for this claim? I will show that rule-governedness is already built into Kant's conception of a faculty or power as a species of ground-consequence relations. In this way, any faculty or power satisfies the generalizability condition *by definition* (it would not count as a faculty or power if it failed to do so). If the generalizability condition is a condition on something being a faculty at all, 'faculty' is an inherently explanatory notion.[19]

Kant's systematic account of faculties will be discussed more fully in Chapter 2. For present purposes, what is crucial is that Kant understands a faculty in *relational* terms: a faculty is defined as a *relation* between a substance and an accident insofar as the substance is the *ground* of the accident. When a substance is related to an accident in a way that qualifies as a faculty, the substance contains the ground in virtue of which that accident inheres in it. Now, Kant understands the relation between

[19] In the present discussion I will mostly ignore the metaphysical distinction between a faculty [Vermögen] and a power [Kraft].

a ground [Grund, *ratio*] and a consequence [Folge, *rationatum*] as a "connection according to a rule" (*MM* 29:747); a ground is "something upon which something else follows according to a rule" (*MM* 29:801). Because the ground-consequence relation is governed by a rule, the existence of the consequence is explained by the existence of the ground: "A is posited, another thing B which is posited **because** A has been posited <*ponitur, B aliud quod posito A ponitur*> […]. A is the ground <*est ratio*>, B the consequence <*rationatum*>" (*MM* 29:808, boldface added).

Kant borrows this generic definition of grounds and consequences from Baumgarten, whose textbook he uses as basis for the metaphysics lectures. However, Kant's explication of this definition in the lecture transcripts makes it clear that his own position departs from Baumgarten's in two respects.[20] First, Kant recognizes the difference between *real* and *logical* grounding relations. If the connection between a ground and a consequence is analytic, it is logical; if the connection is synthetic, it is real. For Baumgarten, all relations between grounds and consequences are analytic and, hence, logical. The notion of grounds that goes into Kant's conception of faculties and powers is that of real grounds. Second, while Baumgarten's definition entails that the ground-consequence relation is completely symmetrical, Kant refines this definition such that the relation is asymmetrical.

> The ground is that which, having been posited, another thing is posited determinately, the consequence is that which is not posited unless something else is posited […]; for if there is a consequence, there must likewise always be a ground, and if something is a ground, there must likewise always be a consequence, but in the first case it is indeterminate, in the other determinate (*MM* 29:808).

According to this refined definition, a ground can be distinguished from its consequence because the ground is *indeterminate* (with respect to the positing of the consequence) whereas the consequence is *determinate* (with respect to the positing of the ground). In other words, a given consequence can follow

[20] Here I follow Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2005), 257–58.

from any number of different grounds, while a ground is such that, when it is posited, only one type of consequence can follow. From this asymmetry it follows that a grounding relation is explanatory in one direction only: the ground explains the consequence but not *vice versa*.

The concept of a faculty is a means to capture the real ground-consequence relation that falls under the relational category of substance-accidence. As Kant puts it, "in relation to accidents <*respectu accidentium*> [a substance] has power insofar as it is the ground of their inherence" (*MM* 29:770). Accordingly, when we attribute a faculty to something, we do the following: (1) we identify it as something that can figure as a ground in a ground-consequence relation; (2) we identify it as a *substance* which through its activity can ground the inherence of an accident in it. Therefore, in attributing a certain faculty to the mind, Kant is in effect identifying the mind as a substance through whose activity accidents of a particular type can come to inhere in it. For instance, to attribute the faculty of thinking to the mind is to claim that the mind is a substance that can ground the inherence of thoughts *qua* accidents in it: "What then is the faculty of thinking? The relation of the soul to thought insofar as it contains the ground of its actuality" (*MM* 29:771). Like any faculty, the faculty of thinking is governed by rules. These rules are the laws of general and transcendental logic.[21] Thus, against the background understanding that these laws are laws governing the faculty of thinking (and, I should emphasize, *only* against that background understanding), the logical laws explain why thoughts have logical features they do.

As commentators have pointed out, Kant conceives the rules governing our rational faculties as at the same time *constitutive* and *normative*.[22] These rules are constitutive in that they adequately

[21] E.g., A52/B76ff.

[22] See Konstantin Pollok, *Kant's Theory of Normativity* (Cambridge: Cambridge University Press, 2017), 4ff.; Schafer, "Kant: Constitutivism," 186–87; Matthew Boyle, "Kant on Logic and the Laws of the Understanding," in *The Logical Alien*, ed. Sofia Miguens (Cambridge: Harvard University Press, 2020), 125ff.; Land, "Faculty Analysis," 7.

describe the characteristic activity of the respective faculty—how the relevant faculty as such functions, in the absence of external influences. In this sense, these rules genuinely describe what it is for an activity to be an exercise of a particular faculty. The activity of the faculty invariably conforms to them absent external influences. However, the activity of a faculty may be interfered or hindered by the influence of another faculty. When this happens, the activity of that faculty may fail to conform to the rules governing it. Consequently, if the faculty is so constituted such that its activity may be interfered at all, the rules governing it will admit of exceptions and hence expressible by an "ought". This picture of the faculty underlies Kant's explanation of the normativity of the moral law.

> A practical rule is always a product of reason because it prescribes action as a means to an effect, which is its purpose. But for a being in whom reason quite alone is not the determining ground of the will, this rule is an imperative, that is, a rule indicated by an "ought," which expresses objective necessitation to the action and signifies that if reason completely determined the will the action would without fail take place in accordance with this rule. (*KpV* 5:20; cf. *GMS* 4:412-13, *GMS* 4:449)

Thus, in the absence of the external influence from sensibility, the will would conform to the moral law without exception. However, the *human* will is imperfectly rational and sensibly conditioned. Accordingly, our will can be interfered by sensibility and, therefore, the moral law is imperative (i.e., normative) for finite rational beings like us. For a perfectly rational will, by contrast, the moral law would be exceptionless.

This conception of the human rational faculty applies not only to the practical faculty. Kant holds a similar view with respect to our cognitive faculty. Like the will, the understanding by itself would not depart from the laws of logic unless it is influenced by the sensibility.

> No power [Kraft] of nature can of itself depart from its own laws. Hence neither the understanding by itself (without the influence of another cause), nor the senses by themselves, can err; the first cannot, because while it acts merely according to its own laws, its effect (the judgment) must necessarily agree with these laws. […] Error is

> effected only through the unnoticed influence of sensibility on understanding, […] just as a moved body would of itself always stay in a straight line in the same direction, but starts off on a curved line if at the same time another force influences it in another direction. (A294-95/B350-51, translation amended)

Just as the will can be deflected from its proper course by sensibility, sensibility can also interfere with the activity of the understanding and cause deviation from its own laws. For this reason, although the laws of logic are constitutive of the activity of the understanding, it does not follow that the understanding cannot fail to comply to them.

Although facts about laws governing faculties are explanatory with respect to the inherence of a certain accident in a substance, these facts need not be explanatorily basic. Any faculty that we can cognize is always in principle reducible to a more basic faculty. In Kant's view, our explanation can never appeal to an *absolutely* fundamental faculty (of any given substance). The reason for this is that we cannot cognize an absolutely fundamental faculty or power of a substance.[23] Nevertheless, Kant allows that we may speak of fundamental faculties or powers in a *comparative* sense.[24] Comparatively fundamental faculties are those that cannot be derived from any more fundamental faculty that we are able to conceive. Hence, they are fundamental relative to the limits of our cognition. Once we have arrived at such faculties, we have reached an explanatory bedrock.

> All human insight is at an end as soon as we have arrived at basic powers [Grundkräfte] or basic faculties [Grundvermögen] for there is nothing through which their possibility can be conceived, and yet it may not be invented and assumed at one's discretion. (*KpV* 5:46-47)

It is in this comparative sense that we should understand Kant's references to fundamental powers throughout his Critical writings, for example, to the three fundamental powers of the mind or to

[23] Cf. *MVo* 28:431-32.

[24] Cf. A469/B677.

repulsive and attractive powers as fundamental powers of matter.[25] Repulsive and attractive powers are the most fundamental powers of matter that we can cognize but not its absolutely fundamental powers. Due to our cognitive limitation, we cannot even conceive of any more fundamental power from which these two can be derived (*MAN* 4:502, 4:513, 4:523-24). Only these two powers "can be thought, as forces to which all moving forces in material nature must be reduced" (*MAN* 4:499). In this sense, these fundamental powers mark the boundaries of our grasp of the nature or essence of matter "beyond which our reason cannot go" (*MAN* 4:534). Similarly, the three fundamental faculties of the mind are to be understood as comparatively fundamental. We cannot derive them from one another or from a common fundamental faculty but only because of the limits of our cognitive insight, though reason enjoins us to assume them to be "derived from one unique fundamental power" (A682/B710).

Note that an appeal to a faculty is explanatory only insofar as it explains the production of an effect as an instance of a general rule. However, it does not explain why the rule itself obtains, that is, why the activity of the substance in question characteristically produces this kind of effect.[26] What *would* explain the rule is the *nature* or *essence* of the substance to which the faculty belongs. However, Kant thinks that, because of the discursivity of our understanding, we can never cognize the essence of a substance apart from its accidents.[27]

> We have absolutely no acquaintance with the substantial, i.e., the subject, in which no accidents inhere, which must be necessarily distinguished from the accident, for if I

[25] The alternation between 'power' and 'force' merely reflects the inconsistency in English translations of the single German term 'Kraft'.

[26] I owe this way of formulating the issue to Daniel Warren, *Reality and Impenetrability in Kant's Philosophy of Nature* (New York: Routledge, 2001), 51.

[27] Cf. *Prol.* 4:333. For a thorough discussion of this point, see Andrew Pitel, "Kant's Critique of Rationalism: Substance, Essence, Noumenon" (Ph.D. Dissertation, The University of Chicago, 2021).

> cancel all positive predicates then I have no predicates and cannot think anything at all. (*MM* 29:771)

If it is impossible for us to grasp the nature of a substance independently of the effects it produces, we can cognize a faculty or power of a substance only through the accidents (the consequence) but never through the substance as such (the ground). For us, a faculty or power is always a faculty or power *to produce a certain kind of effect*. Again, the limits of our cognition constrain the limits of what we can explain by appeal to faculties in general. Kant explicitly says that the laws of the fundamental faculties of our mind cannot themselves be explained.

> For the peculiarity of our understanding, that it is able to bring about the unity of apperception *a priori* only by means of the categories and only through precisely this kind and number of them, a further ground may be offered just as little as one can be offered for why we have precisely these and no other functions for judgment or for why space and time are the sole forms of our possible intuition. (B145-46)

Thus, the appeal to a faculty in general has limited explanatory potency. While it explains the effect of a certain faculty as an instance of a general rule, it does not explain why the rule itself holds.

My main purpose in this section is to clarify how the appeal to faculties in general satisfies the generalizability condition and, therefore, how Kant's own appeal to mental faculties can be seen as responsive to the explanatory goal of the critique. We have seen that a faculty by definition is governed a rule because it is a species of ground-consequence relations, more specifically, one that falls under the category of substance. However, the appeal to laws governing our rational faculties is explanatory only against the background of the metaphysical understanding of these laws as laws governing the activity through which a substance grounds the inherence of an accident. Precisely this sort of metaphysical understanding is what, I will argue, cannot be derived from the awareness of our own rational faculties spotlighted by the reflectivist account.

## 1.4 Reflection, Rationality, and Teleology

The conception of a faculty that we have discussed so far applies equally to rational and non-rational faculties alike. As noted earlier, Kant employs the same framework of metaphysical concepts in both his account of the human mind and his account of material nature. According to the reflectivist account, the difference between rational and non-rational faculties emerges when we consider how they satisfy the *accessibility* condition. The way we come to cognize our own rational faculties differs from the way we can come to cognize the faculties of non-rational things in nature.[28]

In Kant's view, there are two ways we can gain cognition of powers in nature. First, the first way is through empirical observation. We observe effects produced by powers of matter (through changes in appearances) and posit powers that would explain these effects.[29] Further, we can progressively reduce the initial plethora of powers to a smaller number of more fundamental powers, by applying the logical principles of homogeneity, specification, and continuity, with a view to getting closer and closer to an absolutely fundamental power.[30] Second, we can also establish the existence of certain natural powers *a priori*. As Kant demonstrates in the *Metaphysical Foundations*, we can derive the fundamental powers of matter *a priori* from the empirical concept of matter by considering how matter (conceived initially as the movable in space) instantiates the categories.[31]

It is relatively clear that the critique employs neither of these methods to uncover the fundamental faculties of the human mind and the laws governing them. On the one hand, Kant constantly distances Critical philosophy from empirical psychology and Lockean "physiology of the understanding", which suggests that his own investigation of mental faculties does not rely on

[28] How we come to know about rational faculties of *other* beings (i.e. other minds) in Kant's view is a difficult question that I will not consider here.

[29] cf. A204-6/B249-51.

[30] A651/B679ff.

[31] *MAN* 4:474ff.

empirical observation of the working of these faculties or any antecedently given empirical principles. On the other hand, Kant's skeptical attitude toward the possibility of a "pure part" of empirical psychology (i.e., the possibility of deriving *a priori* principles from the empirical concept of a thinking being) speaks against taking Critical philosophy to be an *a priori* science based on an empirical concept of thinking nature analogous to the *a priori* science of corporeal nature of the *Metaphysical Foundations*.

On the contrary, Kant seems to indicate we can come to know about our own rational faculties and the laws governing their activity in a way that is radically different from the way come to know about objects distinct from us. Further, he suggests that this distinctive mode of knowing is what makes the systematicity of the critique possible.

> I have to do merely with reason itself and its pure thinking; to gain exhaustive acquaintance with them I need not seek far beyond myself, because it is in myself that I encounter them, and common logic already also gives me an example of how the simple acts of reason may be fully and systematically enumerated. (Axiv)
>
> Nothing here can escape us, because what reason brings forth entirely out of itself cannot be hidden, but is brought to light by reason itself as soon as reason's common principle has been discovered. (Axx)
>
> Pure speculative reason has this peculiarity about it, that it can and should measure its own faculty a according to the different ways for choosing the objects of its thinking, and also completely enumerate the manifold ways of putting tasks [Aufgaben] before itself, so as to catalog the entire preliminary sketch of a whole system of metaphysics; because […] in *a priori* cognition nothing can be ascribed to the objects except what the thinking subject takes out of itself […]. (Bxxiii, translation modified)
>
> That [a critique] should be possible, indeed that such a system should not be too great in scope for us to hope to be able entirely to complete it, can be assessed in advance from the fact that our object is not the nature of things, which is inexhaustible, but the understanding, which judges about the nature of things, and this in turn only in regard to its *a priori* cognition, the supply of which, since we do not need to search for it externally, cannot remain hidden from us. (A12-13/B26-27)

> [In the critique] reason has the sources of its cognition not in objects and their intuition (through which reason cannot be taught one thing more), but in itself, and, if reason has presented the fundamental laws of its faculty fully and determinately (against all misinterpretation), nothing else remains that pure reason could cognize *a priori*, or even about which it could have cause to ask. (*Prol.* 4:366)

In these passages, Kant consistently insists that it must be possible to us to uncover the laws governing the acts of our rational faculties because it is in ourselves that we encounter them. In appealing to the fact that the faculties are *internal* to us, Kant is implying that the mode of cognitive access we have to our faculties is fundamentally different from the mode of cognitive access we have to things external to us. Further, this distinctive mode of cognitive access makes it possible not only to discover the laws of rational faculties but to do so *completely* and *systematically*.

One way to understand this claim is this. We have immediate and non-empirical conscious access to our own rational faculties in virtue of their self-conscious character. This idea forms the core of the reflectivist account. On this account, the rational faculty essentially differs from the non-rational faculty in that its activity in some way involves an implicit awareness of that activity and its laws. In this way, we can gain cognitive access to the laws governing our rational faculties by making the content of such awareness explicit. The following passage from Schafer offers a representative statement of the reflectivist account.

> For Kant, certain basic faculties of the mind […] must be taken for granted by transcendental philosophy. As a result, the claim that these faculties exist in us can only be shown to be legitimate and non-arbitrary insofar as these faculties are essentially self-conscious. Of course, once we have these basic rational capacities in view, we can then explore their necessary conditions through further reflection upon them. More precisely, we can form an explicit conception of what these capacities involve through reflection upon these basic forms of self-consciousness, and can then investigate the necessary conditions on these capacities through an analysis of the resulting conception of them. But, for Kant at least, this process of reflection and

> analysis must begin with the self-consciousness involved in merely having rational capacities.[32]

On the reflectivist account, we are aware of our own rational faculties simply in virtue of possessing them and this awareness serves as the methodological and epistemological starting point of Kant's philosophy. This awareness supplies the "first principles" from which Kant derives his other philosophical theses. In this section, we will consider textual support for the reflectivist account with a view to finding a principled way to determine the scope and limits of what we can know on the basis of the immediate awareness of our own rational faculties. As we will see, while textual evidence does confirm that Kant thinks we have *some* sort of non-empirical cognitive access to our own rational faculties and that this plays a fundamental role in the critique, these texts offer little aid in constructing a positive epistemological account of reflective cognition of our faculties. More specifically, they do not help to answer what it could mean that we are *implicitly* aware of the laws governing our faculties and what is involved in making the transition from the implicit to the *explicit* awareness of these laws.[33]

The reflectivists draw textual support for their central thesis from various sources. One source is Kant's rather cryptic discussion of 'transcendental reflection' in the Amphiboly. There Kant characterizes transcendental reflection as "the consciousness of the relation of given representations to our various sources of cognition, through which alone their relation among themselves can be correctly determined" (A260/B316). He also claims that transcendental reflection is involved in every act of judgment in some way.[34] One key point of the Amphiboly passage is that reflection in general employs four pairs of opposed concepts called 'concepts of reflection', among them 'matter' and 'form', which are "inseparably bound up with every use of the understanding" (A266/B322).

---

[32] Schafer, "Transcendental Philosophy," 11.

[33] In identifying these issues, I have been helped by Chapter 7 of Matthew Boyle, *Transparency and Reflection: A Study of Self-Knowledge and the First Person Perspective* (Oxford: Oxford University Press, Forthcoming).

[34] Cf. A261/B317.

According to Houston Smit, the fact that these two concepts are enmeshed in every activity of the understanding explains how it is possible for us to identify the form of our cognition and form a conception of it.[35] What appears to be the same notion of reflection is also discussed in some of the logic lecture notes.[36] In these contexts, 'reflection' [Überlegung, *reflexio*] is contrasted with 'investigation' [Untersuchung, *examinatio*]. Reflection refers to implicit awareness of the laws of the understanding and reason or, more precisely, an implicit act of *comparing* representations with those laws. This implicit act necessarily accompanies every act of thinking as a condition for its compliance with the laws. On the contrary, by *investigation* Kant seems to mean the procedure by which we make explicit those laws. As Dieter Henrich argues, we should understand the critique as an investigation in this technical sense.[37]

As we saw earlier, Kant considers the science of general logic as a model for the critique—as "an example of how the simple acts of reason may be fully and systematically enumerated". Both logic and Critical philosophy have the acts of reason as their objects: "general logic, which sets forth the actions and rules of thinking in general, differs from transcendental philosophy, which sets forth the special actions and rules of **pure** thinking, that is, of thinking by which objects are cognized completely *a priori*" (*GMS* 4:390). By claiming that the critique can follow the example of logic, the method of logic can also be applied in the critique in order to exhaustively enumerate "the actions and rules of thinking". If this is correct, we can expect some of Kant's remarks on the method of general logic to illuminate the method of the critique.

Consider the following remark from the very beginning of *Jäsche Logik*.

> The exercise of our powers also takes place according to certain rules that we follow, *unconscious* of them at first, until we gradually arrive at cognition of them through

[35] Smit, "The Role of Reflection," 215ff.

[36] See, for example, *LB* 24:161ff., *LDW* 24:737ff., *JL* 9:75-76.

[37] Henrich, "Kant's Notion of a Deduction," 41ff.

> experiments and lengthy use of our powers, indeed, until we finally become so familiar with them that it costs us much effort to think them *in abstracto*. Thus universal grammar is the form of a language in general, for example. One speaks even without being acquainted with grammar, however; and he who speaks without being acquainted with it does actually have a grammar and speaks according to rules, but ones of which he is not himself conscious. (*JL* 9:11)

Here Kant contrasts two modes in which we can be cognizant with rules and implement them in our activity: *in abstracto* and *in concreto* (the term '*in concreto*' is mentioned later at *JL* 9:17). On the one hand, we can be explicitly conscious or thinking of rules while following them. This is what Kant means by consciousness of rules *in abstracto*. On the other hand, however, it is also possible for us to apply rules without being explicitly conscious of them or explicitly thinking about them. In Kant's example, native speakers speak their language perfectly without having explicit knowledge of its grammar. Nonetheless, there is a sense in which they *know* how to speak the language. The ordinary notion of *knowing* how to do something without explicit knowledge of its rules is captured by Kant's notion of rules beings "cognized only *in concreto*" (*JL* 9:17). Like language use, "the natural use of the understanding and of reason *in concreto*" also takes place independently of consciousness of the laws of logic *in abstracto*. The science of general logic aims to make explicit the rules of which we are implicitly aware in "knowing" how to think—to transform our consciousness *in concreto* of the rules of the understanding and reason into consciousness *in abstracto*. This corresponds to the technical notion of 'investigation', emphasized by Henrich, as the act through which we bring 'reflection' to explicit articulation. In the *Prolegomena*, Kant describes the task of the philosopher in strikingly similar terms: "Nothing can be more desirable to a philosopher than to be able to derive, *a priori* from one principle, the multiplicity of concepts or basic principles that previously had exhibited themselves to him piecemeal in the use he had made of them *in concreto*, and in this way to be able to unite them all in one cognition" (*Prol.* 4:322).

According to Kant, the availability of consciousness *in concreto* of the laws of thinking is what accounts for the possibility of the *completeness* of general logic. In general, Kant connects completeness of a body of cognition with the notion of systematic derivation *from principles*.[38] Hence, if the critique is to achieve its completeness, it must systematically derive cognition in it from principles as well. But what can we reasonably take to be the principles of the critique?

Recall that Kant understands the connection between principles and their consequences as an explanatory relation. Now, since the critique considers our rational faculties as the explanatory grounds of its primary explanandum, the possibility of synthetic *a priori* cognition, it seems plausible to think of those faculties as *themselves* the principles (in the real, metaphysical sense) of Critical philosophy. In fact, this is suggested by Kant's identification of the understanding itself as the principle of the Transcendental Analytic, which is the science that aims to "entirely exhaust the entire field of pure understanding" (A64/B89). The cornerstone of the completeness of the science of pure understanding is the completeness of its table of the categories, by which "the understanding is completely exhausted and its capacity entirely measured" (A79/B105). Kant argues that his table of the categories is complete on the ground that its "division is systematically generated from a **common principle**, namely the faculty for judging (which is the same as the faculty for thinking), and has not arisen rhapsodically from a haphazard search for pure concepts, of the completeness of which one could never be certain" (A80-81/B106, boldface added). Kant goes on to say that although Aristotle had made a significant contribution to the search for the fundamental concepts of the understanding, "his table still had holes" precisely because "he had no principle" (A81/B107). As Kant explicitly says in the passage just quoted, the principle needed here is the *faculty* of understanding itself. This is

[38] Cf. A64-65/B89-90.

consistent with Kant's remark at the beginning of the Analytic. Despite its name, the Analytic is not to be understood in terms of

> the usual procedure of philosophical investigations, that of analyzing the content of concepts that present themselves and bringing them to distinctness, but rather the much less frequently attempted **analysis of the faculty of understanding itself**, in order to research the possibility of *a priori* concepts by seeking them only in the understanding as their birthplace and analyzing its pure use in general; for this is the proper business of a transcendental philosophy. (A65-66/B90-91, boldface added)

This confirms that the faculty of the understanding itself is the fundamental principle of the Transcendental Analytic. But if it is possible for us to analyze this faculty and derive a system of cognition from it, it must first be possible for us to have cognitive access to this faculty—and this cognitive access must be different from our cognitive access to an external object, since this can never guarantee the completeness of our cognition of it.

The Transcendental Analytic is indeed just one part of Critical philosophy, concerned with one specific faculty. However, there seems to be no reason why the idea that a rational faculty can itself serve a principle of rational cognition should not generalize to the other rational faculties as well. If the three higher faculties of the mind are the principles from which cognitions in the three *Critiques* arise, then we must have a special mode of cognitive access to all of these faculties that allows us to discover their laws in entirety. This generalization is supported by Kant's remark that Critical philosophy in both theoretical and practical domains relies on the same kind of consciousness we have of our rational faculties.

> But how is consciousness of that moral law possible? We can become aware of pure practical laws just as we are aware of pure theoretical principles, by attending to the necessity with which reason prescribes them to us and to the setting aside of all empirical conditions to which reason directs us. The concept of a pure will arises from the first, as consciousness of a pure understanding arises from the latter. (*KpV* 5:30)

This remark makes it clear that we can become aware of the laws governing our practical faculty in the same way that we can become aware of the laws governing our theoretical faculty.

The final piece of evidence for the reflectivist account that we will consider comes from one way in which Kant's characterizes the essential difference between non-rational and rational beings. As we have seen, the activities of both rational and non-rational beings take place according to laws. However, Kant holds that the manners in which their activities are determined by laws are different.

> Everything in nature works in accordance with laws. Only a rational being has the capacity to act in accordance with the representation of laws, that is, in accordance with principles […]. (*GMS* 4:412, cf. *KpV* 5:32, 5:125)

Thus, a rational being differs from a non-rational being in that the compliance of their activity with laws is mediated by the *representation* of those laws. In this context, Kant has specifically in mind the will as the faculty for engaging in volitional actions in accordance with moral maxims. Nevertheless, given the generic character of the contrast (between rational beings and the rest of nature), there is no reason why this characterization of a rational faculty cannot be extended to other rational faculties as well. Accordingly, the essential difference between rational and non-rational beings turns on the fundamental role that the faculty of *representation* (i.e., the faculty of cognition) plays in the mental life of the rational being as a whole. This point is also revealed by Kant's identification of the higher parts of the fundamental faculties (cognition, feeling, desire) of the mind with the three higher faculties of cognition (understanding, judgment, reason).

What we might call 'representation-centered' conception of rationality agrees with Kant's characterization of rational faculties in teleological terms.[39] Kant consistently attributes an 'end'

[39] Cf. Schafer, "Kant: Constitutivism," 184–86. Claudi Brink recently argued that the spontaneity of the understanding in particular is to be understood in terms of teleology, which I take to be a congenial point. See Claudi Brink, "Spontaneity and Teleology in Kant's Theory of Apperception" (Ph.D. Dissertation, University of California, San Diego, 2022).

[Zweck] or an 'interest' [Interesse] (a closely related notion) to various rational faculties.[40] An end in the most general sense is defined as "the object of a concept insofar as the latter is regarded as the cause of the former (the real ground of its possibility)" (*KU* 5:220). Thus, an end is an object of conceptual causality, which Kant terms 'purposiveness' [Zweckmäßigkeit]—"the causality of a **concept** with regard to its **object**" (*KU* 5:220). Accordingly, to attribute an end to a faculty is to identify it as a *causality* of a specific kind, namely, one that is governed by the conceptual representation of the object it characteristically brings about.[41] As an example of such a causality, Kant mentions the will as "the faculty of desire, insofar as it is determinable only through concepts, i.e., to act in accordance with the representation of an end". This appears to recast the characterization of the will in the *Critique of Practical Reason* as "the ability to determine [one's] causality by the representation of rules" (*KpV* 5:32). A comparison of these two formulations suggests that the representation of an *end* of a faculty is the same thing as the representation of *laws* governing it. Schafer has proposed as way to understand this: the end of a rational faculty is its own proper exercise; a rational faculty is constitutively aimed at its own proper exercise. If this is correct, then to say that a faculty acts in accordance with the representation of its own end is equivalent to saying that its activity is guided by the representation of its proper exercise and, hence, of the laws governing it.

Schafer's own account of the connection between rationality and teleology of faculties, however, seems to confuse what I take to be two different senses in which our rational faculties are teleologically determined. Call these *internal* and *external* senses. In the internal sense, which we have

[40] E.g., A462/B490ff., A643/B671, A666/B694, A797/B825ff., *KU* 5:184, 5:187. For discussion, see Karl Schafer, "A System of Rational Faculties: Additive or Transformative?," *European Journal of Philosophy* 29, no. 4 (2021): 918–36; and Thomas Pendlebury, "The Shape of the Kantian Mind," *Philosophy and Phenomenological Research*, Forthcoming.

[41] For an insightful discussion of the basic notion of purposiveness as causality in Kant, see Thomas Teufel, "Kant's Non-Teleological Conception of Purposiveness," *Kant-Studien* 102, no. 2 (2011): 232–52.

just considered, each of our rational faculty is teleologically determined in that its activity is guided by an internal representation of its characteristic effect. In the external sense, the entire system of our rational faculties is teleologically structured akin to that of a living organism; the end of each faculty is subordinated to that of another faculty and ultimately to the ultimate end of the whole. I do not deny that Kant is committed to the claim that our rational faculties are teleologically determined in both senses. But whereas Schafer understands them to be two sides of the same coin, I take these to be two conceptually independent claims.[42] On the one hand, the claim that an individual faculty acts according to an internal representation of its characteristic effect does not (at least not obviously) entail that it is part of a teleologically organized manifold of faculties. On the other hand, a teleologically organized system of faculties need not have representation-guided (i.e., rational) faculties as its constituents (think of a living body).

In any case, Kant's teleological characterization of rational faculties adds a further dimension to the thought that some sort of implicit awareness of laws is constitutively involved in the activities of rational faculties. Hence, Kant appears to be characterizing different aspects of one and the same phenomenon in his discussion of (i) reflection, (ii) cognition of rules *in concreto*, (iii) the role of representation of laws in the activity of a rational faculty, and (iv) the fact that rational faculties are teleologically determined in the internal sense. However, these different expressions of essentially the same idea do not get us very far in trying to understand what it could mean for the activity of a rational faculty to constitutively involve some sort of consciousness of laws that is implicit or, paradoxically, unconscious (Kant's preferred term). In short, these expressions seem to be offer different ways to state the problem rather than a solution.

[42] Note that upholding this conceptual distinction does not foreclose the possibility that the specific way in which *our* (i.e., human) rational capacities are teleologically determined in the internal sense depends on their being teleologically determined in the external sense or vice versa.

Note that pre-philosophically we already make a distinction between rational and non-rational activity in terms of the presence or absence of some kind of implicit knowledge or understanding of the rules governing the activity (e.g., between parroting and genuine use of language). We ordinarily say of someone that she *knows* French or *knows* how to dance salsa. The task of philosophy, I take it, is to explicate what goes into the ordinary distinction between "blind" and intelligent rule-following.[43] More importantly for our purposes, the various distinctions that Kant makes do not seem to offer any help in explaining how the possession of a rational faculty enables a distinctive way to gain cognitive access to the laws governing them. Kant's various expressions of this very idea are simply different ways of saying *that* we can come to know these laws rather that explaining *how* we can do so.

Matthew Boyle's otherwise illuminating treatment of the issue suffers from a similar circularity. Boyle focuses on the understanding's formation of empirical concepts on the basis of recognition of similarities and differences among diverse objects as a case study of the exercise of a rational faculty.[44] Boyle proposes that "what distinguishes the kind of awareness of similarity of which our faculty of understanding is capable from a more primitive kind of awareness of similarity is precisely that the understanding classifies consciously, and hence is capable of bringing the principles governing its own classificatory activity to reflective self-consciousness".[45] To classify *consciously*, for Boyle, means to classify things on the basis of the representation of their "points of similarity or dissimilarity as such". The ability to represent similarity or difference *as such*, however, requires implicit awareness of the principles that govern our classificatory activity (this, I take it, is tantamount to saying that it requires *concepts* of similarity and difference). While non-rational animals can respond differentially to things

[43] For a classic discussion of this ordinary sense of 'knowing' and its relation to intelligent rule-following, see Chapter 2 in Gilbert Ryle, *The Concept of Mind* (London: Hutchinson, 1949).

[44] What Kant calls the "logical *actus* of comparison, reflection and abstraction", cf. *JL* 9:94.

[45] Boyle, "Kant on Categories."

that are similar and different, they do not recognize their similarities and differences as such and, hence, are not implicitly conscious of the principles that govern their own responses (they do not have *concepts* of similarity and difference). The problem with this proposal is that it explains the rationality of one activity (conceptual classification of objects as opposed to "blind" response to them) in terms of another activity that is equally rational (conceptual recognition of similarity and difference as opposed to "blind" response to them). In this way, it does not help to explain what it means for our classificatory activity to be rational and not "blind". The appeal to implicit awareness of rules, I have been contending, says nothing more than that the activity in question is, in some yet unspecified sense, not "blind".

Boyle's proposal, in fact, closely matches Kant's own account of the difference between rational and non-rational animals in the pre-Critical essay *The False Subtlety of the Four Syllogistic Figures*, from 1762.

> The higher faculty of cognition rests absolutely and simply on the capacity to judge. Accordingly, if a being can judge, then it possesses the higher faculty of cognition. […] It is one thing *to differentiate* things from each other [Dinge von einander unterscheiden], and quite another thing *to recognize* the difference between them [den Unterschied der Dinge erkennen]. The latter is only possible by means of judgments and cannot occur in the case of animals, who are not endowed with reason. […] *Differentiating logically* means recognizing that a thing A is not B […]. *Physically differentiating* means being driven to different actions by different representations. The dog differentiates the roast from the loaf, and it does so because the way in which it is affected by the roast is different from the way in which it is affected by the loaf (for different things cause different sensations); and the sensations caused by the roast are a ground of desire in the dog which differs from the desire caused by the loaf […]. (2:59-60)

According to Kant, what distinguishes rational from non-rational animals is the possession of the capacity to judge, which involves not only the ability to respond differentially to things that are

different but also to *recognize* the difference between them as such. In the same passage, Kant goes on to consider what these different ways of differentiating objects entails about the "essential difference" between rational and non-rational animal. However, instead of taking recourse to the notion of implicit awareness of rules, Kant suggests that this difference consists in the possession of the capacity to make "one's own representations the objects of one's thought".

> This consideration may induce us to think more carefully about the essential difference between animals endowed with reason and those not so endowed. If one succeeds in understanding what the mysterious power is which makes judging possible, one will have solved the problem. My present opinion tends to the view that this power or capacity is nothing other than the faculty of inner sense, that is to say, the faculty of making one's own representations the objects of one's thought. (2:60)

Kant's uncertainty reveals his recognition of the difficulty of the problem that we have been wrestling with. This problem will be solved if we understand "what the mysterious power is which makes judging possible". This understanding is what is currently missing, and Kant does not pretend to fully have it at the time. However, he believes he has a clue to its solution: the capacity to judge is to be explained in terms of the capacity to entertain thoughts about one's own representations.

In 1762, Kant identifies this capacity with inner sense, but we know that his view changes by the time of the publication of the first *Critique*. In the later work, the capacity has come to be identified with pure apperception, which Kant emphatically distinguishes from inner sense.[46] Nonetheless, Kant remains committed to the view that the capacity to become conscious of one's one representations in a special way, now called pure apperception, is what constitutes the essential mark of rationality. He reportedly endorses the view, for instance, in the *Metaphysik Mrongovius* from 1782-83.

> A concept is the consciousness that the same thing is contained in one representation as in another, or that in manifold representations common marks are contained. This

[46] On the development of Kant's view on self-consciousness, see Chapter 7 of Karl Ameriks, *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*, 2nd ed. (Oxford: Clarendon Press, 2000).

> thus presupposes consciousness or apperception. Mere animals do indeed compare representations with one another, but they are not conscious wherein they harmonize or fail to harmonize with each other. Therefore they have no concepts, and since the higher cognitive faculty consists of these, no higher cognitive faculty. This [faculty] is thus distinguished from the lower cognitive faculty by apperception. (*MM* 29:888)

Apperception is what constitutes the essential difference between the higher (i.e. rational) and the lower (i.e. non-rational) faculty. Further, apperception is described as a necessary condition for the capacity to recognize "common marks" among representations or the points of their similarity and difference and, hence, for possession of concepts.

Kant's suggestion seems to be that in order to explicate the essential difference between rational and non-rational faculties we need to attend to the distinctive way in which the rational subject can be conscious of her representations, that is, in such a way that "the I think must be able to accompany all [her] representations" (B131). Unfortunately, this is not the place to consider whether and in what way this suggestion may produce to a satisfactory account of what makes a faculty rational or in what sense its exercise can be seen as involving implicit cognitive grasp of rules. What I have tried to achieve is this section is to show that the reflectivists are right to attribute to Kant the view that some sort of implicit grasp of laws governing an activity is what essentially distinguishes the rational from the non-rational faculty. At the same time, it is also clear that Kant takes *explicit* cognition of the laws governing rational faculties to be the epistemological starting point of the critique. However, as I have argued, appealing to the distinction between the two modes of grasping rules alone is insufficient to explain how it is possible for us to acquire the special kind of knowledge of the faculties that the critique presupposes. As I have emphasized, as long as we can produce no positive account of what it means for implicit grasp of rules to be involved in the activity of a rational faculty, the appeal to this notion merely captures our pre-philosophical intuition that there is a cognitive

difference between rational and non-rational activity and thereby restates the philosophical puzzle rather than solving it.

## 1.5 Metaphysical Self-Understanding

We have considered textual evidence that supports the reflectivists' claim that Kant thinks of the implicit cognitive grasp of laws governing our rational faculties as the epistemological foundation of the critique. The question that interests us is whether this implicit cognitive grasp of our own faculties also equips us with the metaphysical understanding of these faculties that serves as the necessary background against which the appeal to them can have any explanatory force.

There seems be no reason, textual or philosophical, to assume that Kant thinks it does. Whatever it means to say to that we have an implicit grasp, cognition *in concreto*, of the rules governing our thinking, it is obvious that Kant point in saying this is to explain how we can follow these rules in thinking, while distinguishing this explanation (whatever it may be) from the kind of explanation one might give of why falling bodies follow the law of gravitation. However, it does not seem that to think coherently requires having any grasp of what thought or the activity of thinking is, metaphysically speaking. One can think coherently without having an implicit metaphysical theory of thinking. The same can be said about Kant's account of our practical faculty. To be able to judge whether an action is morally permissible and to act accordingly does not seem to require that one has an implicit metaphysical theory of action or agency.[47] Hence, the burden is on the reflectivists to show why it is plausible (at least as far as Kant's view is concerned) to assume that the implicit cognitive grasp of our

[47] Although Kant says that our action is accompanied by consciousness of our own freedom, he also says that this freedom is to be understood in the negative sense, i.e., freedom *from* the laws of nature and not freedom as causality. I take this to be compatible with the claim I am currently making. I will return to this point below.

rational faculties, which distinguishes the exercise of these faculties from that of non-rational faculties, include, or can in some way ground, the metaphysical understanding of these faculties.

If the reflectivists cannot meet this challenge, their claim that Kant's account of our rational faculties is grounded in our implicit cognitive grasp of these faculties does not explain whence the metaphysical framework underlying this account comes. At most the reflective account can explain the basis for Kant's specific claims about laws governing individual faculties. However, there seems to be no reason to assume that cognition *in concreto* of the laws governing our rational faculties would extend beyond rules which prescribe how those faculties are to be exercised. Further, there also seems to be no reason to assume that cognition *in concreto* of these laws needs to specify them in terms that make any reference to the metaphysical nature of the faculties, i.e. that our implicit grasp of these laws requires an implicit metaphysical understanding of the faculties. I conclude that absent a further argument showing that an implicit metaphysical understanding of the faculties is necessarily contained in our implicit cognitive grasp of them, the reflectivist account fails to explain the origin of the metaphysical framework necessary for the critique to fulfil its explanatory task.

Note that the reflectivists commonly take their own account to be in position to offer this explanation. This is already implicit in their claim that their account offers an adequate account of Kant's basic philosophical method as well as an adequate response to the Strawsonian objection, which is directed not only at the credentials of Kant's specific claims about mental faculties but also at the general framework underlying them. Schafer is especially explicit on this point:

> While this consciousness [our own rational capacities] does not provide us with cognition of the self, it does provide us with a set of basic concepts for thinking about our own capacities and their principles – concepts that the transcendental philosopher can then use as the basis for the development of their systematic philosophical project. So, while this consciousness does not allow us to cognize ourselves as determinate objects, it does allow us to form a basic conceptual framework for thinking about our

> rational capacities and to know that these concepts are instantiated by something (highly indeterminate) in us.[48]

The basic concepts Schafer is referring to must include such concepts as 'faculty', 'power', 'spontaneity', 'receptivity', which, as I have pointed out, must be taken in their technical, metaphysical senses if they are to constitute an explanatory framework for Critical philosophy. I agree with Schafer that whatever the basis may be on which we can legitimately apply such metaphysical concepts to ourselves as thinking subjects, we can only do so in a "highly indeterminate" way, since we are not acquainted with our own faculties and their activity through empirical or *a priori* intuition. But if my account of the character of the consciousness of the laws governing our own rational faculties is correct, this consciousness cannot support the application of any metaphysical concept, no matter how indeterminate, either to ourselves or our faculties.

Schafer seeks to make plausible the claim that the consciousness of our own rational faculties can underwrite some highly indeterminate application of metaphysical concepts to ourselves by assimilating this consciousness to the consciousness of the moral laws. Therefore, he in effect assimilates reflective cognition of our faculties in general to practical cognition, which in some sense reveals our nature as things in themselves, according to Kant. Since Kant's claims about mental faculties in the critique are not empirical claims, Schafer concludes that these claims must be claims about the mind as things in themselves but ones that are "highly indeterminate" or "highly abstract". The abstract or indeterminate character of these claims is supposed to consist in the fact that they do not express theoretical cognition but broadly practical cognition.

> It was never Kant's aim to deny that we are in position to make some non-trivial claims about things in themselves, even from a theoretical point of view. Rather, what Kant denies is that we are capable of achieving theoretical cognition (Erkenntnis) of things in themselves. And the impossibility of this is compatible with us having a

[48] Schafer, "Transcendental Philosophy," 22.

> consciousness of certain aspects of our nature as rational beings, provided this consciousness does not involve the features that are distinctive of "theoretical cognition" in Kant's sense. […] We can (at least in certain special cases) achieve a sort of practically-grounded cognition of the nature of things in themselves, according to Kant.[49]

I have previously indicated my agreement with the idea that the consciousness of our rational faculties has a broadly practical character in the sense that it is akin to the consciousness of rules implicit in the exercise of skills or knowledge-how. However, I have argued that precisely because the consciousness of laws governing our faculties is practical in this sense, it cannot support the application of metaphysical concepts to our faculties.

Neither does Kant hold that the consciousness of the moral laws alone suffices for the application of metaphysical concepts to our practical faculty. What Schafer overlooks is that self-consciousness of pure practical reason, what Kant calls the "fact of reason" [Faktum der Vernunft], does not by itself equip us with a metaphysical conception of our practical agency. As Kant emphasizes, we are *not* immediately conscious of freedom, understood as a positive metaphysical determination of our noumenal selves. Rather, what we are immediately conscious of is the moral law, which is the rule according to which we form maxims of the will.

> I ask instead from what our cognition of the unconditionally practical starts, whether from freedom or from the practical law. It cannot start from freedom, for we can neither be immediately conscious of this, since the first concept of it is negative […]. It is therefore the moral law, of which we become immediately conscious (as soon as we draw up maxims of the will for ourselves), that first offers itself to us. (*KpV* 5:29)

The immediate consciousness of the moral law goes together with the consciousness of freedom in the *negative* sense: part of what it means to be conscious of the moral laws as binding is to be conscious

[49] Ibid. at 20.

of the possibility of acting against the dictates of inclinations out of the respect for it. This, however, does not yet constitute consciousness of freedom in the metaphysically positive sense.

According to Kant, a metaphysical conception of our practical agency requires, in addition, the ideas of reason. Thus, in order to conceive of our freedom *positively* "as the causality of a being insofar as it belongs to the intelligible world" (*KpV* 5:132) we need "the *cosmological* idea of an intelligible world and consciousness of our existence in it" (*KpV* 5:133). Although pure practical reason, through the fact of reason, is the source of objective reality of this idea (and, ultimately, of all three transcendental ideas), its content originates from theoretical reason. As Kant puts it, its object has first to be *thought* by theoretical reason through the categories, for "every use of reason with respect to an object requires pure concepts of the understanding (categories), without which no object can be thought" (*KpV* 5:136). For this reason, Kant considers the postulates of pure practical reason, which result from practical validation of the transcendental ideas, to be *theoretical* propositions, yet "inseparably connected with the moral law" (*KpV* 5:122). Hence, the metaphysical scaffolding of practical cognition does not come from the immediate consciousness of our noumenal nature but rather from transcendental ideas.

## 1.6 Conclusion

I have argued that there is no reason to suppose, as the reflectivists suggest we should, that consciousness of our own rational faculties can be the source of metaphysical understanding of these faculties which undergirds the mode of explanation employed by the critique. Absent a further argument, the reflectivist account is inadequate not only as a response to the Strawsonian objection to Kant's faculty psychology but also as an account of his basic philosophical method.

We have also examined Kant's claim that the consciousness of the moral law, the law governing our faculty of practical reason, provides us with a grasp of our metaphysical nature as things

in themselves, in particular, our noumenal freedom. This may seem to suggest that Kant himself subscribes to the view that consciousness of our faculties can be a source of metaphysical understanding of these faculties. However, a closer look at Kant's view reveals that he does not think that the consciousness of practical reason alone suffices for having a positive metaphysical conception of our freedom. This requires, in addition, the transcendental ideas. Hence, the transcendental ideas are the source of the metaphysical scaffolding that complements our practical cognition. This gives us a clue about how we might explain the source of the metaphysical framework that underlies Kant's claims about mental faculties in general. The positive account of this framework that I will develop in the subsequent chapters will be an elaboration of this basic idea.

# Chapter 2

# The Anti-Metaphysical Reading and the Metaphysics of Faculties

## 2.1 Introduction

I have argued that there is no compelling reason to suppose that the implicit awareness of the laws governing our own rational faculties provides us with the *metaphysical* understanding of these faculties *as* faculties. It is more plausible that this awareness consists solely in cognition "*in concreto*" of rules that we implicitly follow *in* thinking (as something we *do*). This conclusion only seems appropriate given Kant's view, examined in the introductory chapter, that the consciousness of thoughts from the point of view of the transcendental subject of thinking (the I that thinks) is merely formal and does not present us with any metaphysical determinations of the self.

In this chapter, we will look at a popular interpretation of Kant's claims about the mind that perfectly squares with this understanding of our cognitive relationship to our own rational faculties. I call this the 'anti-metaphysical reading'. On this reading, Kant's claims about the mind have no ontological significance whatsoever. Despite their appearance to the contrary, these claims are not about the mind in any psychological or metaphysical sense. Rather, they are claims about quasi-logical structures and distinctions internal to cognition (or representation, more broadly). This passage from Robert Pippin provides a clear statement of the basic idea of the anti-metaphysical reading.

> [Kant's account of the form of cognitive activity] expresses a formal demand or a "supreme law" for what must be *done*, and, again, does not express any claim about the mind as some special, perhaps nonmaterial object. A claim about what, intellectually, we do is not the same as a claim about what the mind is, and the forms of such activity are thus not special mental objects, a special mental structure, or

> dispositions, or psychological laws, but rules, normative restrictions on various activities ("whatever" the mental subject of those activities is). Now, of course, from the Dissertation on, Kant does speak of laws of the mind (*Gemüt*), but in those contexts he is almost always thinking of the mind just as *Spontaneität*, as intuiting and judging activity. The Paralogisms supply ample evidence of this desire not to follow, let us say, the Cartesian route in philosophy, where a priori concern with what can and cannot be known ended up as a metaphysical theory about the mind (as mental substance) which knows and about "its" objects. Kant claims to be interested in the structure of knowing, not in the knowing subject.[1]

Note how Pippin's characterization of Kant's claims about our cognitive activity as claims about "rules, normative restrictions on various activities" or "what, intellectually, we do" aligns well with our diagnosis that consciousness of our rational faculties does not reach beyond rules we implicitly follow in exercising them. However, Pippin goes on to insist that this interpretation of Kant's claims about the mind in no way contradicts Kant's use of seemingly metaphysical notions, such as '*Gemüt*', 'spontaneity', and 'faculty'. It is possible to construe these notions in a non-Cartesian way—as presupposing no metaphysical theory about the mind. In this sense, the anti-metaphysical reading harbors a broadly Strawsonian attitude. While Strawson urges that we reject Kant's incoherent doctrine of "transcendental psychology" and reformulate his philosophically worthwhile theses into strictly formal or quasi-logical terms, the anti-metaphysical readers hold that no such reformulation is necessary; one misreads Kant if one takes his account of the mind to entail anything that goes beyond the claims about the *form* of our cognitive activity.

In maintaining that Kant's use of apparently metaphysical concepts in his account of the mind does not carry any metaphysical connotations that violate its presumptive formality, the anti-metaphysical reading rejects what I call the 'uniformity thesis'. The uniformity thesis says that the

[1] Robert Pippin, *Kant's Theory of Form: An Essay on the Critique of Pure Reason* (New Haven: Yale University Press, 1982), 219.

metaphysical concepts that appear in Kant's Critical account of the mind and in his accounts of systematic ontology and natural science have the same basic ontological meanings across all these contexts.[2] By 'basic' ontological meanings I mean the meanings ontological concepts (as opposed to logical functions) possess independently of their schematization in accordance with any specific form of sensibility. For example, the basic ontological meaning of 'substance' is 'that which exists without being the determination of another', whereas its schematized meaning (relative to our form of sensibility) is 'that which persists in time'. We will examine the basic meanings of Kant's key ontological concepts in detail in Section 2.3. In Kant's writings on natural science and his lectures on metaphysics, the meanings of such concepts as 'faculty', 'activity', and 'spontaneity' are grounded in their status as ontological concepts derived from the categories. As such, they are ontologically significant in that they have implications about what kind of entity (substance, accident, cause, and so on) the object to which they apply is. The anti-metaphysical interpretation denies that in the context of Kant's Critical account of the mind these concepts have such ontological significance: either the same terms do not signify the same concepts at all or they do signify the same concepts but in *logical* rather than *real* use. Unsurprisingly, the anti-metaphysical readers often insist that any use of the categories (and, by extension, any concepts derived from them) involved in Kant's claims about the mind is merely logical rather than real.

The appeal of the anti-metaphysical reading is obvious. It promises an easy way to salvage Kant from the charge of incoherence, and it also seems consistent with a plausible reading of his views on the awareness of our own rational faculties and the sort of knowledge we can have of the laws governing them. However, the anti-metaphysical reading faces the following problems. First, this reading cannot help to make sense how the critique can fulfil its explanatory task (see Chapter 1).

[2] For an explicit rejection of the uniformity thesis by an influential commentator, see Béatrice Longuenesse, *Kant and the Capacity to Judge* (Princeton: Princeton University Press, 1998), 7–8.

Recall that the explanatory strategy Kant deploys throughout the critique, is to appeal to mental faculties as law-governed explanatory relations that simultaneously explain the *a priori* features of our representations as well as the *a priori* cognizable features of real objects. However, as I have argued, the explanatory purchase of the appeal to faculties presupposes a certain metaphysical understanding of what faculties *are* (i.e., generalizable real ground-consequence relations). In the absence of this metaphysical understanding, the formal analysis of elements and structures of representations lacks explanatory significance. Formal analysis may describe *how* things are as represented by us but cannot explain *why* they are so represented. Further, and more importantly, unless our representational faculties are understood as real causal powers through which the mind relates to an object distinct from it, it remains unclear why the laws governing these faculties should entail any objectively valid claims (claims that are about an object in such a way that the laws governing our representational faculties are explanatory of some of its features) rather than being mere rules internal to our "game" of representing—the "blind" or "mere play of representations", as Kant sometimes puts it (cf. A101, A112, A194/B239). Note that we cannot solve the problem simply by showing that these rules are constitutive of a game *necessarily* played by subjects like us, since this would lend them only subjective necessity rather than objective validity.

Second, even if we set aside the explanatory concern of the critique, Kant's claims about the character of mental faculties, taken individually, must be understood as claims about *causal* relations the subject can enter to things in the world and, in this way, have irreducible metaphysical import. In the present chapter, we will consider how certain central claims Kant makes about our faculties can fulfil their intended systematic or argumentative roles only if they are understood as claims about causal relations between the mind and an object distinct from it. If it can be shown that at least *some* of Kant's claims about mental faculties must be understood as claims about real causal relations between the mind and an object distinct from it or about the mind's real causal powers (that is, about

the mind's *faculties* in the technical, ontologically significant sense), we have good reason to assume that *all* of Kant's claims about mental faculties in the critique are to be understood in the same way. One reason for this is that Kant constantly holds that the mind is an essential unity—the fundamental faculties of the mind belong to one and the same mind and form one unified system (we have looked at some passages that express this idea in Chapter 1). Such a claim would be unintelligible if the notion of a 'faculty' is to be interpreted differently in each case. Moreover, given Kant's commitment to architectonic unity of the Critical system, it is reasonable to suppose that its basic framework of concepts is applied uniformly throughout.

We begin, in Section 2.2, by looking more closely at the anti-metaphysical reading as articulated by its major proponents. Section 2.3 offers an overview of Kant's ontology based on the transcripts of his metaphysics lectures from the Critical period. Insofar as the uniformity thesis is true, this elucidates the meanings of the metaphysical concepts deployed in Kant's Critical account of the mind as well. However, the discussion in the lecture transcripts also reveals the meanings of these concepts to be ontologically significant and, hence, cannot be accommodated by the anti-metaphysical reading. In Section 2.4, I defend the uniformity thesis by demonstrating how Kant's central claims about mental faculties must be understood as claims about causal relations between the mind and an object distinct from it.

## 2.2 The Anti-Metaphysical Reading

The anti-metaphysical interpretation denies that Kant's claims about the mind have any ontological significance and, correspondingly, that the mind these claims are about can be understood as either a noumenal or a phenomenal entity. The anti-metaphysical interpretation was predominant during the first few decades following the 1966 publication of Strawson's *Bounds of Sense*, being a reaction against that influential work. As Günter Zöller notes in his 1993 review article, scholarship

on Kant from those decades "shows a remarkable agreement in their understanding of Kant's thinking self as a form or structure that eludes any attempt at reification".[3] Pippin, one of the leading anti-metaphysical interpreters, declares that the originality of Kant's philosophical approach lies in "his inauguration of a purely formal epistemology, one not tied to empirical psychology or to rationalist theories of mind".[4] This interpretive tendency clearly takes its cue from Kant's anti-Cartesian account of self-consciousness in the Transcendental Deduction and his critique of rationalist metaphysics of the soul in the Paralogisms.

The anti-metaphysical interpretation is influentially articulated in the works of Henry Allison and Robert Pippin.[5] Their denial that Kant's claims about the mind have any ontological significance involves two steps. First, they insist that we must distinguish between the subject of cognition and the object of cognition as mutually exclusive categories. The subject of cognition is what Kant refers to by the notion of a formal and empty transcendental subject or the I of apperception, and the subject in this formal sense is the ultimate subject matter of Kant's claims about the mind. Second, they argue that an ontological standing (e.g., noumenality, phenomenality) can meaningfully be ascribed only to an *object* of cognition:

> [T]he question of whether the self of Kant's theoretical philosophy is phenomenal or noumenal is based on the mistaken assumption that it is to be conceived as an object of some sort. […] [T]he essential point is that the phenomenal-noumenal distinction is meant to apply to objects of cognition […].[6]

---

[3] Günter Zöller, "Main Developments in Recent Scholarship on the Critique of Pure Reason," *Philosophy and Phenomenological Research* 53, no. 2 (1993): 460.

[4] Pippin, *Kant's Theory of Form*, 7.

[5] Pippin, *Kant's Theory of Form*; Robert Pippin, "Kant on the Spontaneity of Mind," *Canadian Journal of Philosophy* 17, no. 2 (1987): 449–75; Henry Allison, "Kant's Refutation of Materialism," *The Monist* 72, no. 2 (1989): 190–208; Henry Allison, "On Naturalizing Kant's Transcendental Psychology," *Dialectica* 49, no. 2/4 (1995): 335–51.

[6] Allison, "On Naturalizing Kant," 350; cf. Henry Allison, *Kant's Transcendental Idealism: An Interpretation and Defense*, 1st ed. (New Haven: Yale University Press, 1983), 283–93; Pippin, "Kant on the Spontaneity of Mind," 469.

Similarly, Pippin insists that "transcendentally, thoughts and the subjects of thoughts are miscategorized if thought to raise any metaphysical problems. They are just 'formal conditions for the possibility of experience'; e.g., the logical subject of experience is just 'that which experiences' and should not be construed as a type of being".[7] Hence, Kant's claims about the mind, being claims about the *subject* of cognition, are not about any object, noumenal or phenomenal. Rather, they are to be understood in terms of "a non-noumenal relativization […] to 'subject qua possible knowing subject,' and so can be said not to violate that restriction on knowledge of the subject simply as it is in itself […]".[8] Such claims are claims "about 'subjectivity,' or what it is to be a subject, and not as a claim about a thinking being or substance". This also has a further implication about how to understand Kant's apparent application of ontological concept (the categories and their derivatives) to the mind. Since these concepts are supposed to apply in an ontologically significant way only to *objects*, any use Kant makes of them with regard to the *subject* must be interpreted non-ontologically (either as a merely logical use or as involving different, non-ontological concepts altogether).

But if Kant's claims concerning mental faculties are not about any objects, what are they about? According to the anti-metaphysical interpretation, when Kant speaks of faculties of the mind, what he is really talking about is the *form* or formal features of cognition, and this is different from making substantive claims about the mind as an object in the metaphysical or psychological sense. This seems to be suggested by Kant's occasional remarks that the theoretical part of the critique is concerned with the "formal" conditions of experience; its goal is to produce "*a priori* formal cognition of all objects in general" (A129-30). Correspondingly, Kant calls the form of idealism he propounds 'formal idealism' (B518n, *Prol.* 4:337). Similarly, Kant often says that the practical part of Critical philosophy is

[7] Pippin, *Kant's Theory of Form*, 450.

[8] Pippin (1987, 468-9).

concerned with the "form" of the will, the "form" of a practical principle, or a "formal" practical principle (cf. *GMS* 4:436-37, *KpV* 5:27).

This preoccupation with form, formality, and formal conditions seems to imply that Critical cognition in both theoretical and practical domains falls on the formal side of the distinction Kant draws between 'formal' and 'material' cognition: "All rational cognition is either material and concerned with some object, or formal and occupied only with the form of the understanding and of reason itself and with the universal rules of thinking in general, without distinction of objects" (*GMS* 4:387). Although Kant has general logic specifically in mind when he refers to formal cognition in this passage, it is reasonable to suppose that transcendental logic or Critical cognition more generally also qualifies as formal in the same sense. The theoretical part of Critical philosophy is called transcendental *logic* because it, too, is concerned with "the form of the understanding and of reason" and "the universal rules of thinking" insofar as this can relate to an object *a priori*. Similarly, Kant claims that the critique in general is concerned no with a specific domain of objects but with our rational faculties insofar as these can be a source of *a priori* cognition of objects.[9] This well-known passage from a later work supports the reading according to which Critical cognition is formal rather than material.

> In form resides the essence of the thing [Sache] (*forma dat esse rei*, as the schoolmen said), so far as this is to be known by reason. If this thing [Sache] be an object of the senses, then it is the form of things [Dinge] in intuition (as appearances), and even pure mathematics is nothing else but a doctrine of form [Formenlehre] of pure *intuition*; just as metaphysics, *qua* pure philosophy, founds its knowledge at the highest level on *forms of thought*, under which every object (matter [Materie] of knowledge) may thereafter be subsumed. Upon these forms depends the possibility of all synthetic knowledge *a priori*, which we cannot, of course, deny that we possess. (8:404, translation modified)

[9] Cf. *KU* 5:176, 20:195, 20:202, *MM* 29:751-2.

The anti-metaphysical readers conclude that the subject matters of Critical philosophy are "the necessary and universal 'ways' in which the content of knowledge is known, 'modes' of knowledge, or 'forms of experience'".[10] Kant's appeal to faculties just is a way of referring to "ways", "modes" or "forms" of representing. Allison famously coins the term 'epistemic conditions' to refer to such forms. An epistemic or "objectivating" condition is "a necessary condition for the representation of objects, that is, a condition without which our representations would not relate to objects or, equivalently, possess objective reality".[11] Such is to be contrasted, on the one hand, with a psychological condition, which is "a propensity or mechanism of the mind", and, on the other hand, with an ontological condition, which is "a condition of the possibility of the existence of things, which conditions these things quite independently of their relation to the human (or any other) mind". Allison cites the Humean principles of association (one of which is causation) as an example of psychological conditions and the Newtonian absolute space and time as an example of ontological conditions. On Kant's account, however, both space and time and (the category of) causation are epistemic conditions.

On the anti-metaphysical interpretation, then, Kant's claims about mental faculties are fundamentally claims about objectivating conditions. While Allison introduces this notion in the context of the specifically cognitive relation between a representation and objects, it can be broadened to cover the varieties of relation of a representation to objects treated in the subsequent *Critiques*. Note that objectivating conditions (such as space, time, and the categories) are internal to representation. Put differently, they are themselves (formal) representations that are necessarily involved in any

[10] Pippin, *Kant's Theory of Form*, 16. Cf. Karl Schafer, "Transcendental Philosophy As Capacities-First Philosophy," *Philosophy and Phenomenological Research* 103, no. 3 (2021): 2.

[11] Henry Allison, *Kant's Transcendental Idealism: An Interpretation and Defense*, 2nd ed. (New Haven: Yale University Press, 2004), 11.

representation that relates to an object—they articulate the formal features that a given representation must exhibit in order to have an object-directed content. For this reason, although the anti-metaphysical readers usually do not deny that Kant's claims about epistemic conditions involve the categories and derivative concepts, they insist that these concepts are somehow to be interpreted not as applied to any objects but to our ways of representing them. In this sense, when these concepts occur in the Critical context, they are to be understood as technical terms in a "metalanguage" devised specifically for second-order philosophical investigation of the conditions of our representation of objects.[12] This metalinguistic transposition explains why Kant's apparent application of ontological concepts to the mind does not give rise to any ontologically significant claims about the mind as an object.

The anti-metaphysical readers are often aware of certain limits of their interpretation but usually take them to indicate tensions within to Kant's own account rather than interpretive failures. For instance, Pippin notes that Kant's claim that the mind is *spontaneous* is metaphysically substantive and thus goes beyond the "strict divide of formal-logical vs. substantive".[13] To say that the mind is spontaneous entails both its causal efficacity and its independence from an external cause. Moreover, since spontaneity, for Kant, cannot be attributed to any appearance, his spontaneity thesis stands in tension with his noumenal ignorance thesis. However, Pippin does not consider this to be a problem for the anti-metaphysical interpretation. Rather, he suggests that this is a tension endemic to Kant's system, which later becomes a problematic the post-Kantian "philosophy of subjectivity" seeks to resolve.[14] Allison also recognizes that Kant's spontaneity thesis (in the theoretical case) as well as his doctrine of transcendental freedom (in the practical case) present a threat to the supposed

[12] See Pippin, *Kant's Theory of Form*, 16–17; Allison, *Kant's Transcendental Idealism, 2nd Ed.*, 62–63, 72–73.

[13] Pippin, "Kant on the Spontaneity of Mind," 454.

[14] Pippin, *Kant's Theory of Form*, 220, 451–52, 469, 475.

metaphysical neutrality of Kant's philosophy. While Allison admits that in view of their *content* these claims are metaphysically substantive (and, hence, cannot be simply transposed to a metalinguistic register), he thinks that the *assertion* of these claims is an expression of *normative* rather than *ontological* commitment. Allison rejects the assumption that there must be "some standpoint-independent fact of matter" to which these claims are answerable. On the contrary, each possesses what he calls "warranted assertability relativized to a point of view".[15] What this means is that "each point of view (the theoretical and the practical) has its own set of norms on the basis of which assertions are justified" independently of context-independent truth or fact of the matter. The assertion of freedom is rationally warranted only from a practical point of view, that is, "only in connection with our conception of ourselves as accountable moral agents". Similarly, the assertion of spontaneity in the theoretical case reflects "the way in which the thinking subject must be conceived (or conceive itself) qua engaged in cognition".[16] Indeed, for Allison, there is an analogy between the assertability of our conception of ourselves as free moral agents and that of our conception of ourselves as spontaneous thinking subjects: "Kant's position is that, just as we can act only under the idea of freedom, so we can think only under the idea of spontaneity."[17]

In the remainder of this chapter, my aim is to argue that the claims of freedom and spontaneity are not the only aspects in Kant's Critical philosophy that defy the assumption of strict formality and metaphysical neutrality. In fact, Kant's Critical account of the mind is thoroughly and fundamentally pervaded by metaphysically substantive concepts, whose use, I argue, can only be construed as having irreducible metaphysical significance. These concepts include not only 'freedom' and 'spontaneity' but also 'faculty', 'power', 'action' or 'activity', 'affection', 'passivity', and 'receptivity'. Crucially, I contend

[15] Allison, *Kant's Transcendental Idealism, 2nd Ed.*, 48.

[16] Allison, "On Naturalizing Kant," 349.

[17] Allison, "Kant's Refutation of Materialism," 202; cf. Allison, "On Naturalizing Kant," 349.

that the ontological significance of these concepts cannot be eliminated by treating them as being employed merely logically or as technical terms in the metalanguage of transcendental philosophy, detached from their "object-level" use elsewhere in Kant's philosophy. On the contrary, the ontological concepts that figure in Kant's claims about the mind carry the same basic ontological meanings they do in the contexts of ontology and natural science (the uniformity thesis). This is not to deny that Kant is interested in the forms or formal conditions of our representation of objects or in the rules governing our cognitive activity. What I want to resist is the assumption that the notion of such "forms" or "rules" that "elude any attempt at reification" adequately defines the scope of Kant's philosophical concern in the critique. While Kant does intend to provide an account of forms of our representation and rules governing our cognitive activity, he also understands these forms and rules always at the same time as forms of our representational *faculty* (in an ontological significant sense) and, hence, as laws governing real causal powers of the mind. In this way, Kant's claims about forms or rules cannot be fully understood apart from the metaphysical scaffolding that supports it, or so I will argue.

## 2.3 Substance, Powers, and Activity in Kant's Critical Ontology

In his account of our rational faculties and their activities, Kant consistently employs a battery of interrelated concepts that appear to be anchored in his systematic ontology. In this section, we will try to situate these metaphysical concepts in Kant's ontology as recorded in student notes from his metaphysics lectures in the Critical period in order to understand their precise meaning and interrelation. However, to show that these concepts stem from in Kant's systematic ontology and that Kant employs them in accordance with their metaphysical meanings is not yet sufficient to rule out the anti-metaphysical interpretation, since it could still be argued that the claims in which these concepts figure can somehow be recovered as merely logical claims. Nonetheless, it does give us a

reason to doubt whether the anti-metaphysical interpretation is adequate. If Kant has no intention to make any metaphysically significant claims about the mind, why does he choose to frame his claims using metaphysical concepts? As we will see, these concepts have highly specific, technical meanings, tied to their roles in systematic ontology, whose basic tenets would have been familiar to his intended readers. Moreover, Kant elsewhere, such as in the *Metaphysical Foundations*, self-consciously deploys the same concepts in their metaphysical meanings to explicate his own philosophical views. Therefore, it is hard to believe that in his Critical account of the mind Kant should have employed these concepts "casually", while ignoring or being oblivious about their metaphysical implications.

The fact that the ontological doctrine underlying these concepts largely adheres to traditional rationalist metaphysics might arouse suspicion toward its relevance for understanding Kant's Critical philosophy or even Kant's consistency in framing his Critical views in terms of these concepts. Therefore, it is important to be clear on the extent to which Kant agrees or disagrees with his rationalist predecessors on the nature of ontology. His disagreement does not consist in a wholesale rejection of the business of ontology. Rather, the difference between Kant's Critical view on ontology and the traditional view concerns the nature of ontological concepts. Both views subscribe to the conception of ontology as "the science of the properties of all things in general, […] the science that deals with the general predicates of all things" (*MM* 29:784).[18] However, while the rationalists conceive of these predicates or concepts as grounded in the way things are in themselves, the Critical Kant regards them as pure concepts originating in our faculty of understanding. The "pure concepts of the understanding" act as rules according to which the understanding synthesizes a manifold of representations into one that represents an object in general. A complete ontology would present a system of the ancestral concepts of pure understanding (the categories) together with all their

[18] Compare this to §4 in Alexander Baumgarten, *Metaphysics: A Critical Translation with Kant's Elucidations, Selected Notes, and Related Materials*, trans. Courtney Fugate and John Hymers (London: Bloomsbury, 2014).

derivative concepts (what he calls 'predicables' [Prädikabilien]) in some sort of hierarchy ("the family tree of pure understanding", A82/B108). As Kant memorably declares in the first *Critique*, "the proud name of an ontology […] must give way to the modest one of a mere analytic of the pure understanding" (A247/B303). Accordingly, "with regard to content [ontology] is a self-cognition of the understanding or of reason, just as logic is a self-cognition of the understanding and of reason with regard to form" (*MM* 29:784). From the Critical point of view, then, systematic ontology is part of the unfolding of the "system of pure reason" upon the foundation provided by the critique and, hence, part of a mature science of transcendental philosophy.

Despite his disagreement with the rationalists on the nature of ontology, Kant's view when it comes to the specific content of Critically reformed ontology is surprisingly non-revisionist. Analogous to his attitude toward Aristotelian logic, Kant's attitude toward traditional rationalist ontology is that it already presents us with a largely accurate enumeration of pure concepts of the understanding, although, like Aristotelian logic, in an unsystematic and incomplete manner. This is due to the fact that traditional ontology lacks principles—it is a "hodgepodge" (*MM* 29:785).[19] Thus, one can turn traditional ontology into properly Critical ontology simply by taking the content of the former and imposing upon it the systematic organization grounded in the complete table of the categories.[20] Kant's non-revisionist attitude toward the content of ontology is reflected, on the one hand, in his remarks in the *Critique* that it is unnecessary to provide a complete catalogue of ontological concepts beyond the categories because traditional ontology textbooks can serve as a reliable guide

[19] As seen in Chapter 1, the critique remedies this by deriving a complete table of the categories from the faculty of understanding itself as a principle.

[20] "The headings already exist; it is merely necessary to fill them out, and a systematic topic, such as the present one, will make it easy not to miss the place where every concept properly belongs and at the same time will make it easy to notice any that is still empty" (A83/B109).

for that purpose.[21] On the other hand, it is manifested by the fact that Kant continues to use Baumgarten's *Metaphysics* (one example of such textbooks) as the coursebook for his metaphysics lectures throughout his career. Nonetheless, as the student notes attest, Kant does not shy away from expressing his disagreement with specific views presented by Baumgarten and providing corrections of them. Moreover, Kant very often goes to great lengths to explicate views that are clearly not Baumgarten's but recognizable as his own from the *Critique*. This can be gleaned from the general introduction and the introductory remarks on ontology in the lecture notes from the Critical period.[22]

Before we consider how these lecture notes help elucidate the metaphysical concepts involved in Kant's claims about the mind, let us turn first to what Kant has to say about these concepts in the *Critique* itself. As mentioned earlier, such concepts as 'faculty' and 'spontaneity' are instances of what Kant calls "predicables of the pure understanding" (A82/B108). They are "pure but derivative concepts", which is to say they are pure concepts derived from the categories as "ancestral concepts [Stammbegriffe] of pure understanding". More specifically, the predicables are the products of combining the categories either with each other or with the *modis* of sensibility, i.e. the pure concepts of spatial and temporal location and of various spatial and temporal relations.[23] For example, "under the category of causality, [one can subordinate] the predicables of power, action, and passion; under that of community, those of presence and resistance; under the predicaments of modality those of generation, corruption, alteration, and so on" (A82/B108, translation modified). As already noted, Kant refrains from giving a complete catalogue of the predicables because he thinks "one could readily

[21] A82/B108, A204/B249.

[22] E.g., *MM* 29:747-848, *MV₀* 28:355-440, *ML₂* 28: 531-77, *MV* 29:945-1006.

[23] Cf. "*Modi* of pure sensibility [include] *quando, ubi, situs*, as well as *prius, simul*" (A81/B107). Predicables are "derivable from [the categories] […] by connecting them with one another, or with the pure form of appearance (space and time) or its matter, provided the latter is not yet determined empirically (the object of sensation in general)" (*Prol.* 4:324).

reach this aim if one took the ontological textbooks in hand". The *Critique* contains a few indications in passing of how certain predicables are related to the categories. Here is one example: "Among the different kinds of unity according to concepts of the understanding belongs the causality of a substance, which is called 'power'" (A648/B676). Further, "causality leads to the concept of action, this to the concept of power, and thereby to the concept of substance. […] Where there is action, consequently activity and power, there is also substance […]" (A204/B249-50, translation modified).

As one would expect, these ontological concepts are treated much more extensively in the transcripts of Kant's metaphysics lectures from the Critical period, to which we now turn. For our purposes, it will suffice to focus on Kant's discussion of three pairs of complementary concepts: (i) substance and accidence, (ii) faculties and powers, (iii) activity and passivity.

*Substance and Accident.* 'Substance' and 'accident' are the concepts of entities that stand in a relation which falls under the category of "relation of inherence and subsistence (*substantia et accidens*)" (A80/B106). Accordingly, to apply the concept 'substance' or 'accident' to something is to identify it as a *relatum* in a relation falling under that category; because a given relation of inherence is asymmetric (an accident inheres in a substance but not vice versa), we can distinguish its two relata from one another. In the *Critique*, Kant does not provide the definitions of the categories after presenting their table. Instead, he excuses himself from this task—"I deliberately spare myself the definitions of these categories in this treatise, although I should like to be in possession of them" (A82-83/B108)—and defers it, once again, to a forthcoming "system of pure reason", in which "all the requisite definitions should be not only possible but even easy to produce". Such a system can be generated by systematic regimentation of traditional ontology following the guiding thread of the table of the categories.

Given this remark, it makes sense to look to Kant's explication of 'substance' and 'accident' in his lectures, in which he not only expounds standard teachings of traditional metaphysics but offers corrections from the Critical point of view. Following Baumgarten, Kant introduces the distinction

between a substance and an accident in terms of 'determination' [Bestimmung]: "That which exists without being the determination of another is substance; that which exists only as determination is accident" (*MM* 29:770).[24] In this context, determination means a particular way in which a thing exists. For Kant, every existing thing must exist in some particular way—nothing exists without being some way or other. This follows from the principle of thoroughgoing determination, which says that every existing thing is completely determined with respect to every pair of opposed predicates (cf. B573/B601).[25] That is, for every existing thing and for every pair of opposed predicates, one (and only one) of the two predicates must apply to the thing. Accordingly, determination is "a predicate of a thing by which the opposite is excluded" (*MM* 29:819). A substance is a thing which exists without being a determination of another thing, and an accident is a determination of a substance, the way in which the substance exists (more precisely, only a *positive* determination of a substance counts as an accident). Consequently, accidents cannot exist independently of a substance: "Accidents are modes <*modi*> of the existence of substance and these cannot be apart from that substance; for they exist as predicates and these cannot be apart from the subject" (*MM* 29:769). The Analogies of Experience presuppose the very same conception of substance and accidents: "The determinations of a substance that are nothing other than particular ways for it to exist are called **accidents**" (A186/B229).

'Inherence' refers to the relation between a substance and its accident. However, the talk of an accident inhering *in* a substance could mislead us to think of them as two separate entities standing in a relation of which each is independent. Kant warns us against this misunderstanding.

> With the expression inherence one imagines the substance carrying the accidents, as if they were separate existences, but requiring a basis; however that is simply a sheer misuse of speech; they are simply manners in which things exist. – Insofar as a thing is determined positively, accidents <*accidentia*> inhere in it; insofar as it is negatively

[24] Cf. §191 in Baumgarten, *Metaphysics*.

[25] Cf. §54 in Baumgarten.

> determined, they do not inhere in it. They do not exist for themselves and are not merely supported by the substance like a book in a bookcase. (*MM* 29:770)

As Kant here emphasizes, accidents are not something that exists over and above the substances in which they inhere; they are "simply manners in which things exist".

The particularity of Kant's view on substance can be discerned more clearly if we contrast it with two prevalent views in early modern philosophy.[26] The first view is the view of substance as "bare substrate" (commonly, but perhaps mistakenly, attributed to Locke). The core idea of this view is that if all accidents are set aside, then a substance is a featureless substrate. It should be clear from the foregoing discussion how Kant's view excludes the possibility of such featureless substrate. If every existing thing must be some way, then nothing can exist without being some way or having some particular feature or other. However, Kant's view should not be understood as the view that substance is a bundle of accidents, either—this obviously would nullify the asymmetric dependence between substance and accidents. On the contrary, while it may be the case that no substance exists without any accidents whatsoever, accidents are not all there is to a substance. In Kant's view, a substance has a *real essence* or *nature*; it contains the real *ground* of all of its essential determinations (cf. *MM* 29:820). We have considered Kant's notion of real grounds in Chapter 1. Recall that to say that a substance contains the ground of its accidents is to say that the substance stands in a rule-governed dependence relation to its accidents. It is in terms of such grounding relation between substance and accidents that Kant defines 'activity' and 'power'.

> Substance acts, insofar as it contains not merely the ground of the accidents, but rather also determines the existence of the accidents; or substance, insofar as its accidents

[26] For a helpful and thorough comparison of Kant's view with these two alternative views, see Chapter 1 of Andrew Pitel, "Kant's Critique of Rationalism: Substance, Essence, Noumenon" (Ph.D. Dissertation, The University of Chicago, 2021).

> inhere, is in action, and it acts insofar as it is the ground of the actuality of the accidents. (*MM* 29:822)

Hence, the notion of activity applies to a substance insofar as certain accidents *actually* inheres in it: to the extent that accidents inhere in it, a substance is in action. As we have seen earlier, the *relation* between a substance and its accident insofar as it contains the ground for the actuality of that accident is called a power.

The relational conception of power points toward the second view of substance which Kant's view stands in contrast with. Both Leibniz and the German rationalists equate a substance with a power.[27] Thus, Baumgarten writes: "Power in the stricter sense is either a substance or an accident. Now, it is not an accident, since it is the sufficient ground of all accidents. Therefore, it is a substance, and to the extent that accidents can inhere in it as in a subject, it is substantial".[28] Kant explicitly rejects this view in his lectures.

> Concerning power, it is to be noted: the author [i.e., Baumgarten] defines it as that which contains the ground of the inherence of the accidents; since accidents inhere in each substance, he concludes that every substance is a power. That is contrary to all rules of usage: I do not say that substance is a power, but rather that it has power, power is the relation <*respectus*> of the substance to the accidents, insofar as it contains the ground of their actuality, e.g.: I cannot say that the faculty of thinking within us is the substance itself—the faculty belongs to it—nor even [that] an accident of the thoughts is the accident. We thus have something that is not substance, yet also not accident. What then is the faculty of thinking? The relation of the soul to thought insofar as it contains the ground of its actuality. We have absolutely no acquaintance with the substantial, i.e., the subject, in which no accidents inhere, which must be

[27] For a detailed account of Kant's dispute with his predecessors on the distinctness of power from substance, see Dieter Henrich, "On the Unity of Subjectivity," in *The Unity of Reason: Essays on Kant's Philosophy*, ed. Richard Velkley (Cambridge: Harvard University Press, 1994), 17–54.

[28] §198 in Baumgarten, *Metaphysics*.

> necessarily distinguished from the accident, for if I cancel all positive predicates then I have no predicates and cannot think anything at all. (*MM* 29:771; cf. *MV₀* 28:431)

This dispute might seem to be merely verbal, concerning how we use the term 'power'. After all, Kant's basic picture appears to be the same as Baumgarten's: substance contains the ground of the inherence of its accidents. Why does it matter whether we treat 'power' as terminologically equivalent to 'substance'? The philosophical motivation behind Kant's separation of power from substance is indicated in the last sentence just quoted. The point of insisting on the distinction between a substance and its power is epistemological rather metaphysical; it is to maintain that we cannot cognize the *ground* of inherence of accidents (what Kant calls, following Baumgarten, the substantial) and, hence, the essence or nature of a substance as such. For Kant, we can cognize powers (construed relationally) to the extent and only to the extent that we can cognize their effects. In this sense, we have no insight into a substance considered apart from its accidents (i.e., its effects) at all.[29] The reason for this is that our understanding is discursive: "we cannot comprehend the substantial, but rather merely the accidents […] because the understanding can think only through concepts, and concepts are nothing more than predicates" (*ML₂* 28:563; cf. *MV₀* 28:429). In contrast, Kant is implying that Leibniz and the German rationalists are committed to the view that in cognizing powers (or forces) we are already cognizing substance itself (its real essence) and not merely the relation between substance and effects it produces.

*Faculty and Power.* For Kant as for the German rationalists, 'faculty' [Vermögen] and 'power' [Kraft] are closely related to 'substance'. Therefore, these metaphysical concepts cannot be understood independently of a specific conception of substance. Since Kant conceives of a faculty and a power in relational terms (as a real ground-consequence relation), he regards them as belonging to the category

[29] Cf. *ML₂* 28:563.

of causality rather than substance.[30] Kant treats 'cause' as another term for 'real ground' or '[real] principle'.[31] Hence, cause-effect relation is synonymous with real ground-consequence relation. Note that the fact that substance and causality are two distinct relational categories implies that inherence relation and causal relation are both basic metaphysical relations and irreducible to one another.

Accordingly, we can consider an accident in two ways according to two different relations it can stand to a substance: as a determination to a subject of inherence (according to the category of substance) or as a consequence to the *ground* of their existence (according to the category of causality): "The accident can be considered in two-fold relation; first, as a predicate to a subject, consequently, as a determination, on the other hand, as the consequence to the ground, where the substance is understood as ground and the accidence is understood as consequence" (*MvS* 28:514, my translation; Cf. *MVo* 28:431, *ML₂* 28:564). In the first case, the substance is an existing (or subsisting) thing to which the accident belongs as its determination, i.e., its way of existing. In the second case, the substance is that whose existence grounds the existence of the accident.

What is potentially confusing is that the two relations are often, but not always, co-instantiated by a given pair of a substance and an accident. In Kant's view, insofar as an accident inheres in a substance, the substance is always its (sufficient or insufficient) ground, in which case both relations obtain. However, a substance can also cause an accident to inhere in *another* substance, in which case an accident does not inhere in the substance that causes it.

> Above we have discussed grounds and consequences. However, the relation of substance to accidence is not the same as the relation between grounds and consequences, although they are closely related. A builder builds a house. So, he is the ground of the existence of the house. But the house is not his accident. It requires a special proof to show that substance is the ground of all its inhering accidents. […]

[30] Cf. *MVo* 28:431.

[31] Cf. *MM* 29:769, *MVo* 28:431, *MvS* 28:513, *ML₂* 28:548-49, 28:562, 28:564.

> Every thing is an *ens per se* but not *a se*. Pertaining to the *ens a se* is the relation of ground to consequence. An *ens per se* is a substance […]. If I ask if something is a substance, I do not thereby inquire about the ground but rather whether it is a thing for itself [i.e., an *ens per se*, PC] at all. (*MvS* 28:510-11, my translation)

Kant's point is that a substance can be a ground of accidents inhering in it as well as those inhering in another substance. Therefore, inherence relation is not the same as ground-consequence relation. In the same passage, Kant goes on to claim that Spinoza fails to distinguish inherence from causation, which leads him to substance monism: if every causal relation were an inherence relation, then everything that has a cause would ultimately inhere in a single substance that is the first cause of all things.

Both 'faculty' and 'power' express the generalizable relation between a substance and an accident of a particular type insofar as the substance is a ground of the accident (as opposed to a subject of its inherence). What is the difference between them? A power is the relation of a substance to an accident of a certain type insofar as it contains the ground or the cause of its *actuality*. By contrast, a faculty is the relation of a substance to an accident of a certain type insofar as it contains the ground or the cause of its *possibility*. Kant sometimes describes this difference as one between *sufficient* and *insufficient* ground of an accident or an action that brings it about: "By faculty we think of the possibility of an action, which does not contain the sufficient ground of the action, which is power, but rather the mere possibility of it" (*MVo* 29:434, my translation); a power, however, is "a faculty insofar as it suffices for the actuality of an accident" (*MM* 29:823). As we will see below, what it means for a faculty or power to be insufficient in this sense is that it requires determination by an external power, a power belonging to another substance, in order to actualize an accident. The point of the faculty-power distinction seems to be this: we can attribute a disposition to act in such a way as to produce a certain effect to a substance, due to its nature, even when the substance is not so acting; insofar as we abstract

from the actuality of action and its effect, we say that a substance has a *faculty* for producing a certain effect.

*Activity and Passivity.* A substance acts insofar as it has a power and not merely a faculty: "Substance acts [handelt], insofar as it contains not merely the ground of the accidents, but rather also determines the existence of the accidents" (*MM* 29:822). The relation of an action [Handlung] to a power is, roughly speaking, that of a token to its type: to attribute a certain power to a substance is to say that it can, in general, engage in a certain type of action that brings about a certain type of accidents. In contrast, to say that it is acting though that power is to say that it is engaging in an instance of that type of action and bringing about a particular accident of that type. For example, different thoughts that I entertain at different times are different actions, but they are instances of the same type of actions (thinking) and, hence, of the exercise of one and the same power (the power of thinking). This seems to be what Kant means when he reportedly says that a power is "the relation [Respectus] of a substance to the accidents […] insofar as it contains the *general* [allgemein] ground of a certain *kind* [Art] of accidents" (*MVo* 28:431, my translation, italics added), whereas action is "the determination of the power of a substance as a cause of a certain accident" (*MVo* 28:433, my translation).

An action of a substance, as we have seen, can be internal or external—in Kant's terms, 'immanent' [*immanens*] or 'transeunt' [*transiens*]—depending on whether its effect inheres in the substance itself or in another substance.[32] Immanent actions are those through which a substance produces accidents that inhere in it. Transeunt actions are those through which a substance produces accidents that inhere in another substance.

Now, a substance is said to suffer [leidet] or be passive insofar as the ground of an accident inhering in it is contained in another substance, that is, insofar as it is that which is acted upon in a

[32] Cf. *MM* 29:823, *MVo* 28:433, *MvS* 28:513, *ML*$_2$ 28:565.

transeunt action. Importantly, there is no pure passivity, for Kant: "A substance can never be passive without also being active, i.e., every substance must act in order to receive [empfangen]" (*MvS* 28:513, my translation). Kant often illustrates this point with the example of a representational power.

> Every substance is active insofar as its accidents inhere, but also passive, insofar as they inhere through an external power, this is not self-contradictory. E.g., a representation of a trumpet sound inheres in me through an external power, but not alone, for had I no power of representation <*vim repraesentativam*>, then it could be sounded forever and I could not have a representation. […] We can never be merely passive, but rather every passion is at the same time action. (*MM* 29:823)

Thus, the inherence of an accident in a substance always presupposes that the substance in which the accident inheres possesses a certain power and acts through it. However, the power need not by itself be sufficient to produce the accident, in which case the power requires determination by an external power: "The acting substance <*substantia agens*> determines the power of the substance being acted upon <*substantiae patientis*> in order to produce this accident, therefore all passivity <*passio*> is nothing more than the determination of the power of the suffering substance by an outer power" (*MM* 29:823). In this way, a substance is always active, i.e., acting through its own power, insofar as an accident inheres in it, though its own power might not be a sufficient ground of that accident. Being passive does not entail the absence of activity or power. Passivity is, as it were, a mode of activity.

This understanding of passivity helps to further elucidate the distinction between a faculty and a power. This distinction has consequential application (one which indicates a real ontological difference) only to powers whose exercise consists in passivity, i.e., powers which do not by themselves suffice to actualize the effect and whose exercise requires external determination. We can refer to any power as a faculty insofar as we abstract from its actualization (actuality implies possibility). But only a passive power can become a *mere* faculty. In the absence of external determination, a passive power becomes insufficient to produce its effect: "Faculty, insofar as it is determined with respect to

an effect, is power, and insofar as it is undetermined, *becomes* faculty" (*MM* 29:823-24); "The internal sufficient ground of an action of a substance is power. It is distinguished from being acted upon <*patientia*>, to which something more must be added if it is to become power" (*MM* 29:824). To return to Kant's example, if I possess a faculty of representing a trumpet sound, I also possess a power of representing a trumpet sound as long as that faculty is determined to represent a trumpet sound by (the power of) a sounding trumpet. Taking away the sounding trumpet, however, my power of representing a trumpet sound ceases to suffice to represent that sound and, thus, becomes a mere faculty.

By 'receptivity' [Receptivität] or 'ability' [Fähigkeit] Kant understands "the possibility of suffering" (*MM* 29:772), that is, "the possibility to receive something through the action of another being" (*MVo* 28:434, my translation). Although, following Baumgarten, receptivity is usually mentioned in pair with faculty, the possibility of acting, they are not opposites but rather flipsides of one another—the possibility of suffering just means the possibility of acting through the determination of an external power. On the contrary, the proper opposite of receptivity and passivity is 'spontaneity' or 'self-activity' [Selbsttätigkeit], which refers to activity that proceeds from an "inner principle" alone as opposed to being determined by an "outer principle" (cf. $ML_1$ 28:267-68, *MH* 28:96). Here the term 'principle' is clearly used in the sense of real metaphysical principle or ground. Hence, a spontaneous activity is one in which the substance contains a sufficient ground of the accident being actualized, and a substance is said to be spontaneous insofar as it can act or exercise its power independently of determination by an external power. Note how Kant's familiar characterization of the distinction between sensibility and understanding (and, more generally, between the lower and the higher faculties of the mind) in terms of receptivity and spontaneity aligns itself perfectly with the opposition between these terms in his ontology. In the next section, I will

argue that this is not a merely terminological coincidence but that Kant deliberately employs these ontological concepts in order to give a metaphysical characterization of the mind's causal powers.

## 2.4 Mental Faculties as Causal Relations

This brief overview of Kant's ontology gives us a grasp of the range of metaphysical implications of Kant's claims about the mind, which are framed in terms of ontological concepts, *if* these concepts are to be interpreted in the same way in the context of his Critical account of the mind as in the context of ontology and natural science. If this is so, Kant's claims about the mind would seem to imply a broadly Cartesian picture of the mind as a substance endowed with diverse representational powers. In exercising these powers, the mind grounds representations as accidents inhering in it. Some of these powers (e.g., sensibility) are receptive or passive powers, whose exercise depends on external determination. Some other powers of the mind (e.g., the understanding) are spontaneous and, hence, can bring about representations independently of external determination. However, as I pointed out earlier, to show that it is possible to interpret Kant's application of metaphysical concepts to the mind according to their ontological meanings is not yet to show that they retain these meanings in the present context. Therefore, we have not yet refuted the anti-metaphysical reading. My aim in this section is to argue that Kant's claims about the mind in the critique can only be understood as having irreducible metaphysical significance.

Although the systematic correspondence between the concepts in which Kant formulates his Critical account of the mind and the basic concepts of his ontology does not decisively establish their semantic identity, it does make it plausible to interpret these concepts as having the same meanings across these contexts, especially considering that in the critique there is no indication that these concepts are to be understood otherwise. The continuity between Kant's use of these concepts across these contexts is suggested not only by mere terminological correspondence. It is striking that Kant

frequently mentions mental faculties as examples in his lectures on ontology. We have already seen some of these instances: the faculty of thinking is discussed in Kant's presentation of the relational account of a faculty (*MM* 29:771) and the power of hearing or representing a sound serves as an example of receptivity or passive power (*MM* 29:823). Similar examples are found in all of the lecture transcripts from the Critical period. To mention just one other example, in the section on substance and accidents from *Metaphysik Volckmann* (1784-85) a series of consecutive examples uses mental faculties to illustrate various basic principles of ontology.

> No negative predicates are accidents; these must be affirmative, i.e., they all have reality in them, e.g. […] thoughts are accidents, for they are something real, positive, which contains existence. […] We represent the substance as the *portitor* (bearer) of all accidents, e.g., I am the substance, the desires, thoughts etc. the accidents; therefore, the I is, as it were, the bearer of these desires, thoughts etc. […] I notice, e.g., that I think, feel, perceive emotions etc. These are accidents; but now I wanted to set aside all accidents and cognize only the subject, the substratum – But we cannot cognize the substantial at all because this is to be regarded only as a correlate of all accidents or concept of Something of which all accidents are predicates. […] We can have no clear insight *a priori* how something can exist only as subject without being a predicate, and how this by itself can exist without subject, e.g., as thoughts exist merely as predicates of the human being, one cannot show *a priori* whether it exists as subject or only as predicate […]. (*MVo* 28:428-30)

These examples indicate that, for Kant, we can unproblematically characterize mental faculties—the very same ones that are examined in the critique—as faculties in the ontologically significant sense (and representations, correspondingly, as accidents inhering in the subject *qua* substance). Moreover, our own mental faculties are treated as paradigm cases of what we understand to be faculties in the ontologically significant sense (in these lectures Kant rarely mentions the powers of material substance as his examples). Thus, Kant's treatment of the mind as paradigmatically exemplifying the basic

principles of his ontology seems to indicate his commitment to a broadly Cartesian conception of the mind.

Let us now return to Kant's claims about mental faculties in the critique. I want to argue that these claims have irreducible ontological significance and, therefore, the metaphysical concepts that occur in them are to be interpreted in the same way as in the contexts of ontology and natural science. The basic point which I will try to argue for is this: Kant's purpose in appealing to various faculties of the mind is not simply to identify the types and characterize the formal structure of representations involved in our mental activity; rather, Kant refers to various mental faculties as ways to pick out different *causal* relations that the mind can enter with respect to *objects* that exist and are distinct from it. In this sense, Kant's claims about mental faculties are claims about the mind's causal properties and, hence, its faculties in the technical, ontologically significant sense. Further, I take it that the mind can causally interact with another existing object only if it is itself an existing object—in effect, a substance endowed with faculties that enable it to act or be acted upon by another substance.[33]

Indeed, in making claims about the basic faculties of the mind, Kant intends to say something about the character of our representations. The basic faculties of the mind are *representational* faculties in virtue of being rational or having a "higher" part. What this means is that, unlike the non-rational being's relations to its environment, the rational subject's cognitive, conative, and affective relations to the world (captured by the three basic faculties of the mind) are mediated by representations that

[33] Cf. "A composite <*compositum*> is classified into real, i.e., composite from substances <*compositum ex substantiis*>, and into ideal, i.e., into composite <*compositum*> that is composed from something that is not substance, or from accidents" (*MM* 29:825); "All substances, insofar as they stand in interaction <*commercio*>, constitute a real composite <*compositum reale*>. Interaction <*commercium*> is reciprocal influence <*influxus mutuus*>, for how else is the interaction <*commercium*> of different substances possible than by one determining something in the other, for the substances have an effect in each other" (*MM* 29:825); "[In a] transeunt <*transientes*> [action] [a substance] acts upon another substance or [has] influence <*influxus*>" (*MM* 29:823).

are conscious in the manner that entails the presence of a unified I (see Chapter 1). However, it is important to note that the representations which are said to belong to the basic faculties are not characterized by Kant merely as items internally accessible to the subject (the I) from within the point of view of consciousness (i.e. as Cartesian thoughts; see the introductory chapter), bearing no relation to anything in the world beyond.[34] On the contrary, these representations are normally involved in *causal* relations between the mind and external objects, being either caused by those objects or causally efficacious with regard to them. Kant's characterization of individual types of representations (e.g. intuition, desire, pleasure) as causally operative provides independent arguments for my claim that his concepts of individual mental faculties are not simply ways of referring to kinds of representations (construed formally or abstractly); rather, they refer to real causal relations (*powers* in the technical Kantian sense) which the representing subject can enter into vis-à-vis an external object. Thus, Kant's claims about mental faculties, I argue, are not claims about formal or abstract structures of representations but claims about the character of causal powers that are themselves representational powers. I will illustrate this point with two specific cases.

*Givenness and Cognition.* The first illustration comes from theoretical cognition, whose "conditions for the possibility" are the central concern of the first *Critique*. Cognition in this sense (sometimes called cognition 'in the narrow sense') is a conscious representation of a particular object together with some of its general properties.[35] This sort of objective representation results from

[34] To clarify the dialectic here, the main problem that this study seeks to solve is how Kant can bridge the gap between the conception of representations as Cartesian thoughts and the conception of representations as metaphysical entities (accidents of a thinking substance), since he subscribes to both. The main point of the present chapter is to argue that the anti-metaphysical reading is untenable precisely because Kant is committed to the latter conception.

[35] See Eric Watkins and Marcus Willaschek, "Kant's Account of Cognition," *Journal of the History of Philosophy* 55, no. 1 (2017): 83–112.

subsuming an object given in an intuition under a concept. According to Kant, cognition must satisfy the so-called givenness condition: the object of cognition must be *given* in order to be thought under a concept. An intuition is defined as the representation through which an object is given. Kant never explicitly defines what it means for an object to be given, and it is a matter of considerable dispute how the givenness condition is best formulated.[36] The basic idea, however, is this: in order to cognize an object, the subject must be conscious of the object in such a way that guarantees its existence.[37] Now, throughout his career, Kant is known to reject both the doctrine of preestablished harmony (Leibniz) and occasionalism (Malebranche) as adequate explanation of the reliability of our representations as indicators of the existence of the objects they represent.[38] Therefore, for Kant, the givenness condition can be satisfied in only one way, namely, if there is a relation of causal dependence between cognition and its object: either cognition causally depends on its object or *vice versa*.

In the case of the intuitive intellect, the givenness condition would be satisfied through a causal connection in the object-to-cognition direction of dependence. The objects of intellectual intuition would depend on representation for their existence: "through [its] representation the objects of this representation would at the same time exist" (B138-39; cf. B135, B145). However, our understanding is finite and discursive and thereby cannot produce the object it cognizes. Therefore, in the human case the givenness condition must be satisfied through causal dependence in the cognition-to-object direction; our cognition must causally depend on its object. The contrast between these two ways in

---

[36] For some divergent views, see Watkins and Willaschek; Stefanie Grüne, "Givenness, Objective Reality, and A Priori Intuitions," *Journal of the History of Philosophy* 55, no. 1 (2017): 113–30; Andrew Chignell, "Kant on Cognition, Givenness, and Ignorance," *Journal of the History of Philosophy* 55, no. 1 (2017): 131–42.

[37] Cf. Stephen Engstrom, "Understanding and Sensibility," *Inquiry* 49, no. 1 (2006): 2–25.

[38] He calls both of these views the "*deus ex machina*", which is "the greatest absurdity one could hit upon in the determination of the origin and validity of our cognitions" (10:131). On Kant's rationale for rejecting preestablished harmony and occasionalism in favor of interactionism, see Desmond Hogan, "Noumenal Affection," *Philosophical Review* 118, no. 4 (2009): 501–32.

which a causal connection can hold between a representation and its object is discussed in Kant's famous 1772 letter to Marcus Herz.

> If a representation comprises only the manner in which the subject is affected by the object, then it is easy to see how it is in conformity with this object, namely, as an effect accords with its cause, and it is easy to see how this modification of our mind can *represent* something, that is, have an object. Thus the passive or sensuous representations have an understandable relationship to objects, and the principles that are derived from the nature of our soul have an understandable validity for all things insofar as those things are supposed to be objects of the senses. Similarly, if that in us which we call "representation" were active with regard to the object, that is, if the object itself were created by the representation (as when divine cognitions are conceived as the archetypes of things), the conformity of these representations to their objects could also be understood. Thus the possibility of both an *intellectus archetypus* (an intellect whose intuition is itself the ground of things) and an *intellectus ectypus*, an intellect which would derive the data for its logical procedure from the sensuous intuition of things, is at least comprehensible. (10:130, cf. A92/B124-25)

In this letter, Kant refers to the finite discursive intellect as *intellectus ectypus*, whose intuitive representation "comprises only the manner in which the subject is affected by the object" and reliably indicates the existence of the object just "as an effect accords with its cause". This view is retained in the *Critique*: "The ability [Fähigkeit] (receptivity) to acquire representations through the way in which we are affected by objects is called **sensibility**. Objects are therefore **given** to us by means of sensibility, and it alone affords us **intuitions**" (A19/B33, translation modified).

Because the discursive understanding is unable to produce its object through its intellectual act, the causal connection between cognition and its object in the human case can only be secured by a representation that is acquired through our being causally acted upon (or "affected") by an object. Therefore, our cognitive faculty must have "two stems" (A5/B29): sensibility and understanding. The understanding captures the aspect in which our faculty of cognition is *spontaneous* or *self-active*—in which

it is the source of the form of synthetic unity of cognition. In contrast, sensibility captures the *passive* aspect of our faculty of cognition, which enables the mind to be acted upon by an object distinct from it and thereby produce a representation that causally depends on that object. This makes it clear that 'faculty' as in the 'faculty of sensibility' and the 'faculty of cognition', whose lower part it constitutes, is to be understood squarely in accordance with its use in the ontological context. Moreover, 'faculty' in this case can also be understood in a contrastive sense to 'power': the relation between a substance and an accident of a certain type insofar as it contains an internally *insufficient* ground of the latter. Accordingly, sensibility refers to our faculty of cognition insofar as it is an internally insufficient ground of representations and hence needs to be determined by an object distinct from it in order to produce representations as accidents inhering in the mind. Kant's language unmistakably confirms this reading: "the objects as things-in-themselves give the matter to empirical intuitions (they contain the ground by which to determine the faculty of representation in accordance with its sensibility)" (*UE* 8:215). Empirical intuition causally depends on its object, and, therefore, cognition satisfies the givenness condition insofar as it involves empirical intuition.[39]

This first illustration supports my claim that the notion of a faculty in Kant's Critical account of the mind is not merely a way to refer to a class of representations. Rather, its function is to pick out a generalizable causal relation between the mind and an object distinct from it. In the case of sensibility (and by extension the faculty of cognition as a whole), this relation is one in which the mind is passive vis-à-vis an object whose existence it cognizes, while being at the same time self-active with respect to the inherence of the representation of that object in it. Note that while Kant uses the same generic term 'faculty' to refer to both the sensibility and the understanding, he exclusively refers to the

[39] I leave out of consideration the case of mathematical cognition, in which the givenness condition is satisfied through construction in pure intuition.

sensibility by the more specific terms 'receptivity' [Receptivität] and 'ability' [Fähigkeit],[40] both of which, as we have seen, stand for passive powers. In contrast, the understanding is described by the opposite terms 'spontaneity' and 'self-activity'. This suggests that the understanding, too, is to be understood as a causal power though one that is internally sufficient and requires no external determination. Kant's systematic use of interrelated terminology can only be read as an indication that he was attentive to the metaphysical connotations these concepts have in accordance with his ontology.

If my interpretation of sensibility as a receptivity in the technical Kantian sense is correct, we must reject the anti-metaphysical reading, according to which the distinction between the sensibly given and the spontaneity of concepts is "a distinction internal to epistemology".[41] Allison has shown how the notion of the sensibly given can be construed as a purely formal or epistemological notion. First, he offers a non-causal interpretation of "affection". According to Allison, the object that is said to "affect" the mind and "ground" empirical intuition (with respect to its matter) can be identified with the "transcendental object", which Kant says "must be thought of only as something in general = X" (A104).[42] Now, on Allison's reading, the transcendental object is not a supersensible entity ontologically distinct from the empirical object. Rather, it is the object *immanent* to sensible representation (something like an intentional object) which, however, is considered in abstraction from the conditions of sensibility (i.e. space and time); the very same transcendental object, considered under the conditions of sensibility ("qua sensibly represented"), is an appearance. Allison concludes that affection is "an epistemic rather than a causal relation"[43]—it is "an epistemic relation between a

[40] E.g. A19/B33, A26/B42, A42/B59, A44/B61, A50/B74, B150.

[41] Karl Ameriks, "Kantian Apperception and the Non-Cartesian Subject," in *Kant and the Historical Turn: Philosophy as Critical Interpretation* (Oxford: Oxford University Press, 2006), 66.

[42] Allison, *Kant's Transcendental Idealism, 2nd Ed.*, 68.

[43] Allison, 64.

discursive intelligence and the source of the matter or content of its sensible intuition".[44] This relation is entirely internal to epistemology since the "discursive intelligence" is here to be understood as "the understanding qua engaged in transcendental reflection"[45], whereas "the source of the matter or content of its sensible intuition" is to be understood as the transcendental object immanent to representation, a "correlate of the unity of apperception" (A250, quoted by Allison).

Allison concedes that Kant's formulation of these epistemological claims involves the categories, especially causality. Nonetheless, he insists that "the function of the categories in these transcendental contexts is purely logical, and does not carry with it any assumptions about their objective reality with respect to some empirically inaccessible realm of being".[46] Basically, Allison holds that Kant's use of metaphysical concepts, including both the categories and their derivative concepts, in general within his transcendental account of cognition must be *reinterpreted* as terms of the "metalanguage" of transcendental philosophy, as opposed to the object-language of ontology and natural science.[47] In this sense, these concepts are not used to formulate any metaphysical description or explanation—they do not pick out any metaphysical entities or relations as they do in the first-order discourse of ontology and natural science. On the contrary, their function lies solely within "a second-order philosophical consideration of objects and the conditions of their cognition".[48] Allison's interpretation stands in sharp contrast with the reading I have been arguing for, which insists on the semantic uniformity of the metaphysical concepts across these contexts. I have shown, in particular, that 'faculty' and 'affection' in Kant's doctrine of sensibility must be taken to refer not to a purely epistemic relation but to a causal relation between the mind and an object ontologically distinct from

[44] Allison, 67.

[45] Allison, 58.

[46] Allison, 72–73.

[47] Allison, 73.

[48] Allison, 63.

it. Consequently, Kant's claims about the mind involve a real (and ontologically significant) rather than merely logical use of the categories.

*The Will as Causality.* Along the same line as Allison's anti-metaphysical interpretation of Kant's account of the cognitive faculty, one might be tempted to read his account of the will in the same way. It can seem plausible to interpret Kant's moral philosophy as offering not a *metaphysics* of morality but a *logic* of practical thinking—an account of rules of practical reasoning or what commentators have called the "categorical imperative procedure".[49] Accordingly, any occurrences of metaphysical concepts, such as 'faculty' and 'spontaneity', in Kant's account are to be interpreted as standing for some normative notions or else as belonging to the content of the postulates of pure practical reason, which are theoretical assertions justified on moral grounds. On such an anti-metaphysical reading, Kant's moral philosophy does not present us with any metaphysical explanation of practical normativity or agency but only an account of normative requirements on practical deliberation, considered from the point of view of the agent who engages in it. Accordingly, Kant's appeal to the higher faculty of desire or the will carries no metaphysical connotation: the faculty of desire does not refer to a real causal power of the mind but is simply a shorthand for a class of representations (understood abstractly or formally) involved in practical deliberation.

This anti-metaphysical reading might seem to find its basis in Kant's statement that the fundamental principle of morality, whose articulation and justification is the main concern of his practical philosophy, is in some sense "merely formal" (cf. *KpV* 5:22). While his doctrine of freedom can be read as specifying metaphysical conditions of morality, Kant also equates the will with practical reason and the free will with a will determined by "the mere form of a law" (cf. *KpV* 5:28-29). The

[49] For an influential example of the proceduralist reading, see John Rawls, "The Categorical Imperative: The First Formulation," in *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge: Harvard University Press, 2000), 162–80.

emphasis on the form of practical reason or the formality of the moral law could easily be understood as suggesting that the primary or even exclusive concern of Kant's practical philosophy is the formal aspect of practical reasoning. In other words, its aim is to articulate the formal normative principles according to which we deliberate or reason about what to do, analogous to the formal normative principles of theoretical reasoning, which logic aims to articulate. Accordingly, the anti-metaphysical readers may attempt to argue that the apparently metaphysical notions, such as 'faculty (of desire)', 'causality (of the will)', 'freedom', are in fact notions internal to normative ethics or moral epistemology (just as the anti-metaphysical interpretation of the Kantian account of cognition takes notions like 'faculty (of cognition)', 'affection' and 'spontaneity' to be internal to epistemology).

This reading, however, can be dismissed on a similar ground as its theoretical counterpart. Like his notion of the faculty of sensibility or of cognition, Kant's notion of the faculty of desire (and, more specifically, the will) refers to a generalizable causal relation between the mind and an object distinct from it. Therefore, this notion is not merely a way to single out a class of representations. Again, this is not to deny that the faculty of desire is a representational faculty and, thus, there is a class of representations which properly belongs to it. The point is that Kant considers these representations (e.g, practical principles, inclinations, pleasure) not only as items that figure in the deliberative process from the agent's point of view (cf. Cartesian thoughts in the case of cognition) but also as real entities that play a causal role in relation to a causal power of the mind. Accordingly, Kant does not understand the moral law merely as a formal rule governing the agent's practical deliberation but also as a law according to which a real causal power of the mind operates.[50]

[50] This claim has also been argued for in Robert Johnson, "The Moral Law as Causal Law," in *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*, ed. Jens Timmermann (Cambridge: Cambridge University Press, 2009), 82–101.

In the first *Critique*, Kant mentions the will, in parallel to the sensibility, as another faculty of the mind that instantiates a causal connection between representation and an object. The exercise of the will in intentional action provides another case in which a necessary relation obtains between representation and its object.

> There are only two possible cases in which synthetic representation and its objects can come together, necessarily relate to each other, and, as it were, meet each other: Either if the object alone makes the representation possible, or if the representation alone makes the object possible. If it is the first, then this relation is only empirical, and the representation is never possible *a priori*. And this is the case with appearance in respect of that in it which belongs to sensation. But if it is the second, then since representation in itself (for we are not here talking about its causality by means of the will) does not produce its object as far as its **existence** is concerned, the representation is still determinant of the object a priori if it is possible through it alone to **cognize something as an object**. (A92/B124-25)

The first case is obviously the affection of sensibility, which gives rise to sensation; the representation thereby produced causally depends on its object—"the object alone makes the representation possible". With respect to the second case, however, there are two possible ways in which "the representation alone makes the object possible": *either* the representation constitutes a condition for the possibility of cognizing an object as such *or* the representation "produce[s] its object as far as its **existence** is concerned". The latter, Kant says, takes place in the case of the "causality by means of the will". The conception of the will as a causality is also presupposed by the Third Antinomy, whose Thesis argument characterizes freedom of the will as "a causality […] through which something happens without its cause being further determined by another previous cause" (A446/B474). The apparent conflict between freedom of the will and determinism would not arise unless the will is conceived as a causality.

The same conception of the will becomes central to Kant's practical philosophy. The faculty of desire is explicated as "a being's faculty to be by means of its representations the cause of the reality of the objects of these representations" (*KpV* 5:9n; cf. *KU* 5:220). The will "the faculty to determine [one's] causality by the representation of rules" (*KpV* 5:32); "In the concept of a will […] the concept of causality is already contained, and thus in the concept of a pure will there is contained the concept of a causality with freedom" (*KpV* 5:55). This conception of the will as causality is also crucial for Kant's argument for the claim that free rational agents are subject to universal laws. The argument begins with a definition of the will as "a kind of causality of living beings insofar as they are rational, and *freedom* would be that property of such causality that it can be efficient independently of alien causes *determining* it" (*GMS* 4:446). Kant then argues that since "the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited" (*GMS* 4:446), a free will cannot be lawless but must operate according to "immutable laws" that must nonetheless be distinct from natural laws. In the second *Critique*, this step in the argument is not made explicit. However, the claim that the will is a causality and, hence, must be governed by a universal law is implicit in Kant's argument, in Theorem III, that a practical law is one that determines the will through its mere form alone. This argument shows why the *matter* of a practical principle (an empirically given object of interest) cannot be the source of the normative necessity required of a practical law. However, it would not follow that the *form* of a law can provide the requisite necessity unless we assume that the will is a causality and as such cannot operate without a law.[51]

[51] For a different reading of this argument, see Andrews Reath, "Formal Principles and the Form of a Law," in *Kant's Critique of Practical Reason: A Critical Guide*, ed. Andrews Reath and Jens Timmermann (Cambridge: Cambridge University Press, 2010), 31–54.

Given the indispensable role of the conception of the will as a causality in Kant's argument for the universality of the moral law, it only makes sense that Kant should describe the goal of the *Critique of Practical Reason* in terms of the discovery of the determining ground of the free will *qua* causality:

> In [the practical use of reason], reason is concerned with the determining grounds of the will, which is a faculty either of producing objects corresponding to representations or of determining itself to effect such objects (whether the physical power is sufficient or not), that is, of determining its causality. (*KpV* 5:15)

Note that the question of the "determining ground" of the will is tantamount to the question of the *internal sufficiency* of the will as a faculty in the technical sense. Thus, to show, as Kant intends to, that the determining ground of the will may exclude all empirically conditions and, hence, all conditions *external* to the will amounts to claiming that the will contains a sufficient ground for its own action—the will, in other words, is a *spontaneous* power, which does not depend on any determination by an external power. Accordingly, Kant describes the moral law, in accordance with which the will operates insofar as it is free, as "the law of causality from freedom" (*KpV* 5:16).

## 2.5 Conclusion

In the two central cases we have considered the notion of a faculty that figures in Kant's respective claims about the mind must be understood in its technical, ontological sense rather than as a mere shorthand for a class of representations (à la anti-metaphysical reading). Accordingly, the forms and rules of the relevant faculties cannot be construed exclusively as forms and rules internal to representations (understood abstractly or formally). Rather, as we have seen, the forms and rules in each case are at the same time forms and rules that govern the activity of real causal powers of the mind. Space and time are not just forms of our intuition but also the way in which we are affected by

objects. Similarly, the moral law is not just a rule according to which we deliberate about what to do but also a law governing our causality from freedom.

However, as I pointed out earlier, it is reasonable to suppose that this interpretive thesis can be generalized to all other faculties of the mind investigated in the critique. Kant holds that all these diverse faculties form a unified system attributable to a single mind, a claim which would be unintelligible if each instance of 'faculty' is interpreted differently. Moreover, Kant's commitment to architectonic unity makes it even more likely that his basic framework of concepts applies uniformly across different domains of the Critical system. Hence, we have good reason to conclude that Kant's claims about the mind in the critique are, in general, claims about the mind's real causal powers or generalizable causal relations either between the mind and one of its accidents or between the mind and objects distinct from it. Consequently, we must reject the fundamental thesis of the anti-metaphysical interpretation, namely, that these claims are claims about representations and their forms or formal conditions, which are to be construed as having different, non-metaphysical meanings.

But how is the critique in a position to sustain such metaphysical significant claims about the mind? My answer, which I will develop in later chapters, is in some way akin to Allison's proposed solution to the problem of Kant's claims of freedom and spontaneity, which seem to defy anti-metaphysical interpretation. Recall that Allison's strategy is to construe these claims as having their basis not in the answerability of these claims to independent "facts of the matter" but rather in their "warranted assertability" according to certain rational norms of assertation—"just as we can act only under the idea of freedom, so we can think only under the idea of spontaneity".[52] This implies that these metaphysically significant assertations do not meet the requirements of cognition in narrow theoretical sense.[53] In a similar vein, I will propose that Kant's metaphysically substantive claims about

[52] Allison, "Kant's Refutation of Materialism," 202.; cf. Allison, "On Naturalizing Kant," 349.

[53] On the question whether these claims can be interpreted simply as claims of *practical* cognition, see Chapter 1.6.

the mind are supported by warranted application of the transcendental idea of the soul to ourselves, which is not grounded in theoretical cognition of our noumenal souls but in reason's demand for explanatory understanding. In brief, our use of the idea of the soul is warranted neither because we have (theoretical) cognitive insight into our noumenal selves nor because this idea is a condition for the possibility of cognition (like the categories) but because it is a condition for the possibility of explanatory understanding (what Kant terms *Begreifen*).

Note that this is perfectly compatible with Kant's anti-Cartesian conception of self-consciousness as well as his critique of rationalist doctrine of the soul in the Paralogisms. Although Kant denies the possibility of theoretical cognition of whether we noumenally are immaterial thinking substance (as the idea of the soul represents us), he concedes that "everyone must necessarily regard himself as a substance, but regard his thinking only as accidents of his existence and determinations of his state" (A349). My interpretation can be considered as an attempt to provide a positive account of the necessity attached to this self-conception as thinking substance. On the reading I propose, this necessity is not a matter of subjective necessity (the point is not that we simply cannot conceive of ourselves otherwise) but a matter of normative necessity: we *ought* to regard ourselves as a substance but regard our thinking only as accidents of this substance, according to a certain requirement of rationality connected to the positive theoretical use of reason.

Although my proposed reading rests on the same basic idea as Allison's, it differs from his in the following respects. First, while Allison introduces the notion of context- or standpoint-dependent warranted assertability to handle recalcitrant cases of metaphysical claims that cannot be recovered as metalinguistic claims about epistemic conditions, my reading generalizes this notion to cover *all* of Kant's Critical claims about the mind insofar as these claims involve the use of metaphysically substantive concepts. Second, Allison does not provide a concrete account of what the norms are that justify the metaphysical assertions in question and why it is plausible to suppose that rationality

involves these norms. To that extent, his view is vulnerable to the criticism of being *ad hoc*. My reading, in contrast, is not vulnerable to such criticism because it grounds the justificatory basis of Kant's metaphysical conception of mind in his independent account of the rational demand for explanatory understanding. Finally, my position and Allison's appeal to two different textual provenances. The main textual reference of Allison's proposal is Kant's doctrine of the practical standpoint (which may seem to obscure rather than support his position). My proposal, as we will see, is based on Kant's account of the regulative use of the transcendental ideas.

Before I articulate and defend this reading in more detail, let us turn first to two other competing interpretations, which appear to be consistent with the diagnosis of Kant's claims about the mind put forward in this chapter.

# Chapter 3

# Two Realisms about the Mind

## 3.1 Introduction

In the last chapter, I argued that Kant's claims about the mind's faculties are to be understood as metaphysically substantive claims—they are claims about the mind's real causal powers. Accordingly, the ontological concepts that figure in them are to be interpreted as having the same basic ontological meanings as in the contexts of ontology and natural science. For this reason, we must reject the anti-metaphysical interpretation, which holds that these claims are claims about forms of representation or epistemic conditions, understood in a strictly formal and ontologically neutral sense. Given this diagnosis of the character of Kant's claims about the mind, it seems natural to take recourse to a realist interpretation. A realist interpretation is one that takes Kant's metaphysically substantive claims about the mind to entail his commitment to the mind's real existence as entity of some sort. This would provide a straightforward solution to the problem of how it is possible for Kant to sustain such metaphysically significant claims about the mind. On the realist reading, these claims are simply epistemic claims about an existing entity.

Now, for Kant, things exist either as an appearance or a thing in itself. This gives rise to two branches of the realist interpretation, both of which have been endorsed by influential commentators. On the *noumenalist* interpretation, Kant's claims about the mind are epistemic claims about a thing in itself. His account of the mind's faculties and activities, therefore, describes the mind's noumenal or "intelligible" character. On the *phenomenalist* interpretation, Kant's claims about the mind are epistemic claims about an appearance, that is, our own mind as it appears to us in inner sense.

In this chapter, we will consider the philosophical and textual rationales for each branch of the realist interpretation. My aim will be to argue that, for varying reasons, both of these interpretations fail to offer an adequate account of the status of Kant's claims about the mind in the critique.

## 3.2 The Phenomenalist Reading

The most influential defenders of the phenomenalist reading are Patricia Kitcher and Andrew Brook.[1] Both commentators maintain that the object of Kant's positive account of the mind can only be the phenomenal or empirical self, the thinking subject as she appears to herself in inner sense. Their basic argument is as follows: the phenomenal-noumenal distinction is exclusive and exhaustive, and the noumenal self is completely unknowable.[2] Given their implicit assumption that Kant's claims about the mind fully qualify as claims of objective knowledge, they conclude that the self or the mind which these claims are about is a phenomenal entity.

On the phenomenalist reading, Kant's claims about the mind ultimately rest on observable, empirical facts about the phenomenal self. More specifically, the starting point of Kant's analysis are empirical facts about what the cognitive subject can *do*. Kitcher equates these facts with facts about "experience in general" (as opposed to particular experiences), from which one can derive the "repertoire of cognitive tasks" that the thinking subject can perform. From empirical facts about cognitive tasks, we can then infer the existence of various mental faculties required to perform those

---

[1] Patricia Kitcher, *Kant's Transcendental Psychology* (Oxford: Oxford University Press, 1990); and Andrew Brook, *Kant and the Mind* (Cambridge: Cambridge University Press, 1994). Other proponents of the phenomenalist interpretation include Sellars, "...This I or He or It (The Thing) Which Thinks..."; Ralf Meerbote, "Kant's Functionalism," in *Historical Foundations of Cognitive Science*, ed. John-Christian Smith (Dordrecht: Springer Netherlands, 1991), 161–87; Paul Guyer, "Psychology and the Transcendental Deduction," in *Kant's Transcendental Deductions: The Three Critiques and the Opus Postumum*, ed. Eckart Förster (Palo Alto: Stanford University Press, 1989), 47–68; and C. Thomas Powell, *Kant's Theory of Self-Consciousness* (Oxford: Clarendon Press, 1990).

[2] Cf. Kitcher, *Kant's Transcendental Psychology*, 22.

tasks. The result of this analysis (the account of the mind presented in the critique) is a highly abstract, functional description of cognitive processes necessary for the range of experience that we as a matter of fact enjoy. On the one hand, despite being highly abstract, Kant's claims about the mind remain empirical inasmuch as they are based on empirical facts about cognitive tasks. They describes, in functional terms, cognitive processes (such as synthesis) that conjoin one mental state with a successive one according to a natural causal law.[3] On the other hand, despite being empirical, these claims are general and necessary, relative to minds with functional profiles like ours—they necessarily hold true of any cognitive system that are capable of the same set of cognitive tasks as our minds.[4] These relatively necessary claims state empirical laws governing causal relations connecting states of a cognitive system under a functional description.

In this way, the phenomenalist interpretation construes Kant's view on the mind as a forerunner of contemporary functionalism. On Brook's functionalist reading, which emphasizes this connection, Kant's account of the mind aims to explain mental phenomena (i.e. representations and their contents) by specifying the "unobservable psychological antecedents" of the observable mental phenomena.[5] While these antecedents are causal in character, their description is given in psychological terms (i.e., in terms of what the mind does in order to transform input states into output states) and may not be reducible to non-psychological description. Thus, the functional description remains highly abstract in the sense that it states what must be true of any possible cognitive system in which the phenomenon in question could occur, regardless of its concrete (physical or non-physical) constitution.[6] In an analogy commonly used by contemporary functionalists, a functional description

[3] Kitcher, 139–40.

[4] Brook, *Kant and the Mind*, 6–7.

[5] Brook, 12.

[6] Brook, 13–14; cf. Guyer, "Psychology and the Transcendental Deduction," 65–68.

specifies the "program" or "software" that can in principle be realized in a range of possible machines or "hardware" with totally different intrinsic constitutions. In other words, a mental state of a certain type is individuated not by its intrinsic constitution (which can vary from one cognitive system to another) but by its causal profile involving its relations to input sensory stimuli, other internal mental states, and output behavior. We say that a mental state of a certain type and content can be *multiply realized*. For example, while the state of pain in human beings consists in undergoing C-fiber stimulation, pain in other creatures may be realized by states with totally different physical constitution (e.g., states of a physiologically different neural unit or silicon-based states). What determines the identity of all these states as pain states, however, is their being causally related in the right way to the same sensory stimuli, the same set of mental states, and the same types of behavioral response (e.g., such that they tend to be caused by physical injury, give rise to the belief that something is awry with the body, cause anxiety, and so on).[7]

Kitcher's interpretation of the notion of synthesis in the Transcendental Deduction serves as a good example of how specific tenets of the Kantian critique can be construed along the line of functionalism. This quote nicely summarizes Kitcher's account of synthesis.

> A synthesis is an act, or to be more neutral, a process that produces a representation, by adding or combining diverse elements contained in different cognitive states in a further state that contains elements from these states. The easiest way to think about syntheses may be to regard them as processes that realize (mathematical) functions. Given a set of input states, a synthesis produces a certain output state. Thus, Kant's talk of *rules* of synthesis, or *functions* of synthesis, is, as it often seems in the text, pleonastic. The domain of synthesis comprises cognitive states (including representations); their range, cognitive states and representations. Kant provides no

[7] For the characterization of contemporary functionalism in this paragraph, I have drawn upon Janet Levin, "Functionalism," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2018 (Metaphysics Research Lab, Stanford University, 2018), https://plato.stanford.edu/archives/fall2018/entries/functionalism.

> account of the intrinsic nature of cognitive states representations, or synthesis. […] Syntheses might be processes that produce physical states from sets of physical states, that produce immaterial states from sets of immaterial states, or that produce symbols from symbolic inputs.[8]

Thus, Kant's account of synthesis, as Kitcher understands it, is perfectly consistent with the functionalist thesis of multiple realizability. What determines whether a process counts as a synthesis is not the intrinsic constitution of that process or of its input and output states but the causal relations between the input states and the output states, which, however, can be abstractly described in terms of mathematical functions.

On the basis of the multiple realizability thesis, the functionalist readers often hold that Kant's account of the mind is ontologically neutral. Note, however, that the sense in which Kant's account is ontologically neutral according to the functionalist reading is different from the sense in which it is said to be ontologically neutral according to the anti-metaphysical reading (see Chapter 2). On the anti-metaphysical reading, Kant's claims about the mind are ontologically neutral in that they are not claims about any entities whatsoever—forms of representations or epistemic conditions are not to be identified with any existing entities (doing so is to commit a category mistake). In contrast, what the functionalist readers mean when they say that these claims are ontologically neutral is that they have no implication about the intrinsic *constitution* of the entity that the mind is and (at least for Brook) not even about whether the mind is material or immaterial (thus, both materialism and phenomenal dualism are compatible with Kant's claims about the mind).[9] However, the functionalist readers do take Kant's claim about the mind to be claims about an existing entity and, more specifically, a

[8] Kitcher, *Kant's Transcendental Psychology*, 74–75.

[9] Textual evidence seems to suggest that Kant espouses phenomenal dualism. For discussion on this point, see Karl Ameriks, *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*, 2nd ed. (Oxford: Clarendon Press, 2000), 42–47.

phenomenal entity. As Brook explicitly says, Kant's claims about the mind are empirical claims about "actual minds as we observe them".

The phenomenalist interpretation may appear to be in a position to account for the metaphysical substantiveness of Kant's claims about the mind. As we have seen, a functional description of mental processes does presuppose the real existence of a concrete substratum underlying those processes, i.e., a phenomenal substance (material or immaterial) that undergoes the successive series of functional states constituting those processes. Hence, we can make sense of the ontological concepts that figure in these claims as applied to the mind *qua* phenomenal substance. Further, on the phenomenalist interpretation, these claims also qualify as genuine cognition, grounded in the application of the categories to an object of experience—they express empirical cognition of the subject as an object of inner sense. In this sense, insofar as the epistemic warrant for applying the categories (and derivative concepts) to the mind is concerned, Kant's claims about the mind *qua* phenomenal substance are on a par with epistemic claims we make about other objects of experience.

## 3.3 Higher Faculties and Non-Empirical Laws

The phenomenalist interpretation is, nonetheless, not free from difficulties. One of the main difficulties with this interpretation is that it cannot be easily reconciled with the supposedly non-empirical or *a priori* status of both the method and the results of Critical philosophy. As we have seen in Chapter 1, Kant maintains that the Critical investigation does not employ an empirical method—it draws nothing from experience—and, hence, must be sharply distinguished from empirical psychology. This goes hand in hand with the idea that the critique is supposed to be a science properly so-called or the foundation for such a science. According to Kant, a proper science must derive its doctrine solely from *a priori* principles. Further, Kant also suggests that we have available to us a non-empirical mode of cognitive access to the laws governing our rational faculties, whose systematic

articulation the critique seeks to provide. The non-empirical character of the Critical method and the *a priori* status of the cognition it aims to produce raise an issue for the phenomenalist interpretation. As we have seen, this interpretation holds that the Critical investigation proceeds from empirical observation (of cognitive tasks) and aims to discover empirical (albeit relatively general and necessary) facts about the mind.

The non-empirical character of the method of the critique is necessitated by the nature of its object of inquiry. The critique is concerned not only with the fundamental faculties of the mind but more specifically with its *higher* faculties.[10] Recall that three fundamental faculties of the mind are the faculty of cognition, the faculty of desire, and the faculty of feeling of pleasure and displeasure, each of which is divided into a higher and a lower faculty. A faculty possesses a higher part in virtue of its being subject to the legislation of reason (i.e., the higher faculty of cognition). The legislative sovereignty of reason over all the faculties of the mind is captured by Kant's claim that each of the three sub-faculties of the higher faculty of cognition (reason in the broad sense) doubles as the higher faculty of each of the three fundamental faculties of the mind: the understanding acts as the higher faculty of cognition, reason (in the narrow sense) acts as the higher faculty of desire, and the power of judgment acts as the higher faculty of feeling.[11]

This picture of the mind is what underlies the tripartite division of the critique of pure reason. Each of the three *Critiques* aims to uncover and justify synthetic *a priori* judgments originating from each of the three fundamental faculties of the mind. However, Kant also thinks that these faculties can serve as a source of synthetic *a priori* judgments only if they are governed by *a priori* principles—by laws legislated by reason which can be discovered *a priori*—and, hence, insofar as they contain a higher faculty. This seems to contradict the basic tenet of the phenomenalist interpretation that the

[10] Cf. *EE* 20:201, *EE* 20:245, *KU* 5:195.

[11] Cf. *KU* 5:168.

laws governing mental activity in the critique are empirical causal laws established on the basis of observation.

The difference between a higher and a lower faculty is essentially the difference between a spontaneous and a passive faculty (see Chapter 2). Thus, to claim that each of the fundamental faculties of the human mind has a higher and a lower part is to say that the human mind, with respect to each of its fundamental faculties, has both a spontaneous and a passive aspect—an aspect in which it is self-active or active "through an inner principle" as well as an aspect in which it is active through an external determination. In one of the metaphysics lecture transcripts, Kant explicates the way in which each of the fundamental faculties of the mind is spontaneous and the way in which it is passive.

> What belongs to my faculty so far as I am passive belongs to my lower faculty. What belongs to my faculty so far as I am [self-]active belongs to my higher faculty. […] The *lower* faculty of cognition is a power to have representations so far as we are affected by objects. The *higher* faculty of cognition is a power to have representations from ourselves. […] The *lower* faculty of desire is a power to desire something so far as we are affected by objects. The *higher* faculty of desire is a power to desire something from ourselves independently of objects. […] The *lower* faculty of pleasure and displeasure is a power to find satisfaction or dissatisfaction in the objects which affect us. The *higher* faculty of pleasure and displeasure is a power to sense a pleasure and displeasure in ourselves, independently of objects. (*ML*$_1$ 28:228-29)

This passage makes it clear that the distinction between the higher and the lower faculties of the mind concern the two different ways in which each faculty *qua* representational faculty can be the ground of their respective kind of representations (cognition, desire, pleasure and displeasure). On the one hand, insofar as they are lower powers, they ground the actuality of representations through being

affected by objects. On the other hand, insofar as they are higher powers, they ground the actuality of representations independently of affection by objects.[12]

Accordingly, our representations can be sorted into two broad categories according to their origin, either in a lower or a higher faculty. Kant calls the representations that originate from a lower faculty 'sensuous' [sensuell] representations and those that originate from a higher faculty 'intellectual' [intellectuell] representations.

> All of our cognitions, pleasure, etc., and desires are either sensuous or intellectual. So there are also sensible and intellectual representations. […] Sensible representations are representations according to the manner in which I am affected by things; intellectual representations are ones that are independent of that. (*MM* 29:880)
>
> All our representations have a twofold origin; they arise (1) from sensibility and (2) from the understanding. The first is called the lower, and the other the higher cognitive faculty. The first belongs to sensuality and the other to intellectuality. Everything that is sensible rests on receptivity; but what belongs to spontaneity belongs to the higher powers. We will have sensible cognitions, sensible pleasure and displeasure, and sensible desires. All three of these powers can be sensible. (*ML*$_2$ 28:584)

Thus, the point of the distinction between the lower and the higher faculties is to mark two different sources from which our representations originate. They can originate in the affection of our mental faculties by an object or in the self-activity of our mental faculties alone. In the first case, the ground or the cause of the representation lies both in our power *and* the power of the object affecting it. In the second case, the ground or the cause of the representation lies solely in our power.

In this way, we might think of the distinction between the lower and the higher faculties of the mind as the *metaphysical* counterpart of Kant's *epistemological* distinction between empirical and *a*

[12] Kant sometimes explicates the notion of a higher faculty in terms of its autonomy, which makes some sense given that, for him, spontaneity and autonomy (i.e., self-legislation) are closely related: "the faculties of the soul in general, insofar as they are considered as higher faculties […] contain an autonomy (*KU* 5:196).

*priori* representations.[13] The notion of a lower power refers to the causal origin of empirical representations; the notion of a higher power refers to the causal origin of *a priori* representations. This explains why the critique is concerned specifically with the higher faculties of the mind. As I pointed out in Chapter 1, the explanatory goal of the critique is to explain the possibility of synthetic *a priori* judgments—to explain how such judgments can relate to objects despite being *a priori*. Now, *a priori* representations are ones whose causal origin lies in the spontaneous activity of a *higher* mental faculty: "The higher [faculty] has spontaneity in its representations. Consequently, we view ourselves as the compelling cause for it" (*MM* 29:888). Therefore, if we can provide an account of what feature of a given higher mental faculty makes it the case that the *a priori* representations arising from it can be related to an object, we will also be able to explain how *a priori* judgments made on the basis of those representations can themselves be related to an object. As we have seen, such an account will detail how certain *a priori* cognizable properties of the objects of a specific kind of representations are grounded in the law-governed spontaneous activity of their faculty of origin.

For our present purposes, it is crucial to note that, in Kant's view, the laws governing the higher faculties of the mind are non-empirical laws in both a metaphysical and an epistemological sense. From a metaphysical point of view, they are not natural laws or laws governing causal relation among appearances. From an epistemological point of view, these laws cannot be established on the basis of observation. Therefore, the phenomenalist interpretation is false because it takes Kant's account of mental activity to consist in an empirical description of causal relations that hold among mental states and between mental states and stimulations and behavior, while taking mental states to be states of an appearance and their description to be derived from observation.

[13] As I argued in Section 2.4, Kant always understand mental faculties and related notions in dual senses—in the formal-epistemological sense as well as in the metaphysical sense.

The paradigm example of non-empirical laws is the moral laws or the law governing the will *qua* the higher faculty of desire. Kant often says that that the moral law is the *a priori* principle of the faculty of desire and, hence, cannot be "extracted from an empirical proposition" or derived from the observation of the actual, empirically conditioned operation of the will.[14] Further, as we saw in Section 2.4, Kant also positively characterizes the moral law as the law governing a noumenal causality, which he explicitly opposes to the causality of nature. Thus, the law of the higher faculty of desire is non-empirical in both the metaphysical and the epistemological sense.

However, the non-empiricality claim is not limited to only the laws governing the higher faculty of desire but extends to as well to the other two higher faculties of mind—cognition and the feeling of pleasure and displeasure. The generality of this claim is what Kant suggests in the following passage from the *Groundwork*.

> A rational being must regard himself *as intelligence* (hence not from the side of his lower powers) as belonging not to the world of sense but to the world of understanding; hence he has two standpoints from which he can regard himself and cognize laws for the use of his powers and consequently for all his actions; *first*, insofar as he belongs to the world of sense, under laws of nature (heteronomy); *second*, as belonging to the intelligible world, under laws which, being independent of nature, are not empirical but grounded merely in reason. (*GMS* 4:452)

This seems to imply that not only a rational being's higher power of desire but *all* of his higher powers (note that 'powers' is in the plural form throughout) are subject to laws of the "world of understanding". These laws are distinct and independent from laws of nature or "the world of sense", which govern appearances and are empirical. Laws of nature are characterized as "laws in accordance

[14] *KpV* 5:11ff.

with which everything happens" as opposed to "laws in accordance with which everything ought to happen" (*GMS* 4:388).[15]

Strikingly, in the *Jäsche Logik*, an analogy is drawn between the moral law and the laws of logic as laws of the faculty of understanding to indicate that the laws of logic are non-empirical in the same sense.

> Some logicians, to be sure, do presuppose *psychological* principles in logic. But to bring such principles into logic is just as absurd as to derive morals from life. If we were to take principles from psychology, i.e., from observations concerning our understanding, we would merely see *how* thinking does take place and *how* it *is* under various subjective obstacles and conditions; this would lead then to cognition of merely *contingent* laws. In logic, however, the question is not about *contingent* but about *necessary* rules; not how we do think, but how we ought to think. The rules of logic must thus be derived not from the *contingent* but from the *necessary* use of the understanding, which one finds in oneself apart from all psychology. (*JL* 9:14)

This remark obviously presupposes the picture of faculties according to which the operation of a higher power (here the understanding) can be interfered by a lower power (the sensibility) such that it deviates from the laws constitutive of its proper activity (See Section 1.3). The observation of actual operation of the understand can only inform us about how it takes place "under various subjective obstacles and conditions"—that is, under potential influence of the sensibility—rather than how it "ought" to take place. The point is so much that we cannot "derive an ought from an is" as that observation does not equip us with any basis for distinguishing defective from non-defective cases of the operation of the understanding. Therefore, we cannot rely on an empirical means if our aim is to discover the laws governing the proper exercise of the understanding.

[15] Cf. *GMS* 4:427, *GMS* 4:452, A547/B575.

Although the focus of the *Jäsche Logik* is the laws of the understanding articulated by general logic, the same characterization carries over to the laws of the understanding and reason articulated in the first *Critique* as well. Kant classify these laws as laws of "transcendental logic", "laws of the understanding and reason […] insofar as they are related to objects *a priori*" (A57/B81). Like the laws of general logic, the laws of transcendental logic also belongs as part of the "canon"[16]—"the sum total of the *a priori* principles of the correct use" (A796/B824)—of the understanding and reason. As Kant notes, such laws cannot serve as a canon if they are derived from experience: "Logic can have no empirical part, that is, no part in which the universal and necessary laws of thinking would rest on grounds taken from experience; for in that case it would not be logic, that is, a canon for the understanding or for reason" (*GMS* 4:387). Therefore, both general and transcendental logic "has no empirical principles [and] draws nothing from psychology" (A54/B78).

Kant also remarks that his derivation of the categories, which are the laws of transcendental synthesis of the understanding, differs from Locke's "physiological derivation" in that it is "entirely independent of experience" (A86/B119; cf. Aix). Correspondingly, transcendental synthesis itself is to be distinguished from empirical synthesis, which is "subject solely to empirical laws, namely those of association, and that therefore contributes nothing to the explanation of the possibility of cognition *a priori*, and on that account belongs not in transcendental philosophy but in psychology" (B152). Since transcendental philosophy (i.e., the critique) aims to explain the possibility of *a priori* cognition associated with the understanding, it must attend to transcendental rather than empirical synthesis. This illustrates the close connection, already pointed out, between the non-empirical character of the laws of faculties with which the critique is concerned and its explanatory goal, which is to explain the possibility of synthetic *a priori* cognition.

[16] A12/B26. Cf. A53/B77f; A131/B170.

Finally, in both Introductions to the *Critique of Judgment*, Kant states that the higher faculty of feeling is likewise governed by non-empirical laws, invoking the same reasoning.

> [I]f one wants to give the origin of these fundamental principles and attempts to do so in a psychological way, this is entirely contrary to their sense. For they do not say what happens, i.e., in accordance with which rule our powers of cognition actually perform their role and how things are judged, but rather how they ought to be judged and this logical objective necessity is not forthcoming if the principles are merely empirical. (*KU* 5:182)

> [A]esthetic judgments [...] lay claim to necessity and say, not that everyone does so judge – that would make their explanation a task for empirical psychology – but that everyone ought to so judge [...]. (*EE* 20:238-9)

To conclude, all the three higher faculties of the mind, in Kant's view, are governed by non-empirical laws. These laws are non-empirical both in the sense that they cannot be established on the basis of experience and in the sense that they are not laws governing natural causation. If these laws are the primary concern of the critique, it cannot be the case that Kant's claims about the mind are empirical propositions about natural causal relation between states of the mind. Hence, we must dismiss the phenomenalist interpretation.

## 3.4 Mind as Intelligence

The phenomenalist interpretation's claim about the epistemic status of Kant's claims about the mind goes hand in hand with its claim about the ontological status of the mind as the subject matter of these claims. If Kant's claims about the mind are empirical propositions, justified on the basis of experience, it is natural to take the mind, the object of these propositions, to be an object of experience or an appearance (i.e., of inner sense). I will now argue that we have reason to reject this latter claim as well.

We have seen that Kant holds, on the one hand, that the laws of the higher faculties, which the critique seeks to articulate, are non-empirical not only in the sense that they cannot be derived from experience but also in the sense that they are not laws of nature. On the other hand, Kant also emphasizes that mental activity that takes place according to laws of nature (such as reproductive synthesis) cannot contribute to the explanation of *a priori* representations and cognitions, which is the goal of the critique. It is not difficult to see why Kant thinks this is the case. *A priori* representations are those that arise independently of affection by objects; therefore, they must originate in the activity of a spontaneous power, which produces representations independently of such affection. But, according to Kant, spontaneity cannot be a property of appearance, and only appearances are subject to laws of nature. It follows that spontaneous powers cannot be attributed to appearances, which alone are subject to laws of nature.[17] In this way, the fact that *a priori* representations can only be explained in terms of spontaneous activity of the mind precludes that possibility of explaining them in terms of mental activity that takes place according to laws of nature.

[17] Indeed, this only follows if 'spontaneity' is understood in the *absolute* sense. Kant distinguishes between absolute and relative spontaneity, only the first of which is excluded from appearance (cf. *KpV* 5:96, A447/B475, A533/B561). However, it is reasonable to assume that higher faculties of the mind are spontaneous in the absolute sense. Recall that the point of Kant's distinction between the higher and the lower faculties is to distinguish spontaneous from *passive* powers of the mind: whereas the activity of passive powers depends on external determination, the activity of a self-active power proceeds solely from an "inner principle". However, Kant makes it clear that a relatively spontaneous power is not independent from external determination: "Spontaneity in some respect <*spontaneitas secundum quid*> is when something acts spontaneously *under a condition*. So, e.g., a body which is shot off moves spontaneously […]. This spontaneity <*spontaneitas*> is also called automatic spontaneity <*spontaneitas automatica*>, namely when a machine moves itself according to an inner principle, e.g., a watch, a turnspit. […] Here the inner principle <*principium*> was determined by an external principle <*principium externum*>" (*ML*$_1$ 28:267). Thus, strictly speaking, a relatively spontaneous power is not a self-active but a passive power. For a helpful discussion of the distinction between relative and absolute spontaneity, see Stefanie Grüne, "Kant and the Spontaneity of the Understanding," in *Self, World, and Art: Metaphysical Topics in Kant and Hegel*, ed. Dina Emundts (Berlin/Boston: De Gruyter, 2013), 145–76.

Now, in Kant's view, there are only two kinds of explanatory grounds we can appeal to in order to explain any given effect: nature and freedom, understood as two kinds of causality, each operating according to its own laws. In the *Metaphysik L*$_1$, Kant remarks that "nature and freedom […] are the two explanatory grounds [Erklärungsgründe] of the understanding" and that explanations that appeal to grounds other than these two "serve only as a cushion for ignorance and deprive the understanding of all use" (*ML*$_1$ 28:200).[18] The same idea is evinced in the following passage from the *Metaphysik Mrongovius*.

> There is thus in the world, on the one hand, nature and on the other freedom. What happens can also be viewed as an effect that does not have its ground in the series of appearances, which are connected according to the general laws of nature, but rather occurs spontaneously <*sponte*> according to laws of freedom; e.g., if I do something according to the guidance of my understanding then this is freedom, insofar as it springs from the understanding; […] In the world as a series of appearances, we cannot and must not explain any event from spontaneity <*ex spontaneitate*>, only the reason [Vernunft] of human beings is exempted from this. (*MM* 29:861-62)

Thus, there are only two ways in which a given effect can be explained—by appeal to laws of nature or by appeal to laws of freedom. Since Kant denies that *a priori* cognitions can be explained by appeal to laws of nature, it seems to follow that, if they can be explained at all, their explanation must be given in terms of laws of freedom.

It is obvious that freedom *qua* an explanatory ground is to be understood as freedom in the *positive* sense.[19] Freedom in the negative sense only indicates independence from causality according to the laws of nature. Thus, to appeal to freedom in the negative sense is not so much to specify an explanatory ground of a given effect as to state what its explanatory ground is not. In contrast, to

[18] For discussion of this idea, see Eric Watkins, "Kant on Rational Cosmology," in *Kant and the Sciences*, ed. Eric Watkins (Oxford: Oxford University Press, 2001), 70–89.

[19] Cf. *GMS* 4:446, *KpV* 5:132.

appeal to freedom in the positive sense is to specify an explanatory ground of an effect, namely, a noumenal causality—"the causality of a being insofar as it belongs to the intelligible world" (*KpV* 5:132). If this is correct, to explain *a priori* representations through freedom rather than through nature amounts to explaining them as effects of noumenal causality. Therefore, insofar as the mind possesses higher faculties, which produce *a priori* representations as their effects, it must be considered a noumenal entity.

There is ample textual evidence that confirms the conclusion of our long-winded inference. In a passage from *Groundwork* quoted above, Kant remarks that a rational being must regard herself as "intelligence" insofar as she possesses higher faculties. As we will see shortly, the concept of an intelligence, for Kant, is a concept of a noumenal being. The characterization of the human being as an intelligence insofar as she is endowed with a will is ubiquitous in Kant's works on practical philosophy. Here is another example.

> The human being, who […] regards himself as an intelligence, thereby puts himself in a different order of things and in a relation to determining grounds of an altogether different kind when he thinks of himself as an intelligence endowed with a will, and consequently with causality, than when he perceives himself as a phenomenon in the world of sense (as he also really is) and subjects his causality to external determination in accordance with laws of nature. Now he soon becomes aware that both can take place at the same time, and indeed must do so. For, that a *thing in appearance* (belonging to the world of sense) is subject to certain laws from which *as a thing* or a being *in itself* it is independent contains not the least contradiction; that he must represent and think of himself in this twofold way, however, rests as regards the first on consciousness of himself as an object affected through the senses and as regards the second on consciousness of himself as an intelligence, that is, as independent of sensible impressions in the use of reason (hence as belonging to the world of understanding). (*GMS* 4:457)[20]

[20] Cf. *GMS* 4:457f.; *GMS* 4:453; *KpV* 5:114.

In this passage, Kant makes it clear that to regard oneself as an intelligence and a member of the world of understanding is to regard oneself as a noumenon ("*a thing* or a being *in itself*"). This claim is made specifically with regard to the human being as a possessor of a will. However, I wish to argue that the same claim applies to the human being as a bearer of higher faculties more generally. Consider this well-known passage from the first *Critique*.

> Yet the human being, who is otherwise acquainted with the whole of nature solely through sense, knows [erkennt] himself also through pure apperception, and indeed in actions and inner determinations which cannot be accounted at all among impressions of sense; he obviously is in one part phenomenon, but in another part, namely in regard to certain faculties, he is a merely intelligible object, because the actions of this object cannot at all be ascribed to the receptivity of sensibility. We call these faculties understanding and reason […]. (A546-7/B574-5)

Kant seems to be asserting that insofar as the human being possesses higher faculties ("in regard to certain faculties", "understanding and reason") he is independent from nature, the sensible world, and phenomena. In addition, he characterizes the human being in this regard as a "merely intelligible object", i.e., an intelligence.

In the *Metaphysik Mrongovius*, the term 'intelligence' is likewise introduced in the context of the distinction between the two ways in which the human being can regard herself.

> Consciousness is the principle of the possibility of the understanding, but not of sensibility. […] The self underlies consciousness and is what is peculiar to spirit [Geist]. we can consider this self in three ways: I think as intelligence, i.e., the subject of thinking is intelligence. I think as subject which has sensibility, and am soul [Seele]. I think as intelligence and soul, and am a human being. (*MM* 29:878)

This passage begins with a restatement of a familiar claim that the capacity for non-empirical self-consciousness or pure apperception is the distinguishing mark of a rational being. It then goes on to contrast two different ways of in which we can consider this self (of which we are conscious in a non-empirical manner): as a bearer of the understanding or the higher faculties, the self is called

'intelligence'; as a bearer of sensibility or the lower faculties, the self is called 'soul'. Here a third way of considering the self is added, namely, the 'human being', which is a way to refer to the self as both an intelligence and a soul. In the same lecture transcript, however, we also find unequivocal evidence that, in Kant's view, intelligences or bearers of higher faculties are noumenal beings.

> With respect to the powers of the mind [Gemüt], a human being belongs to the noumenal world <*mundo noumeno*>, for through the understanding he can cognize things as they are, as e.g., his moral relations, truth, etc., and in this regard his actions are free, as well as the phenomenal world <*mundo phaenomeno*>, insofar as through his actions he belongs to the chain of appearances. (*MM* 29:862)

> When we do something, insofar as it proceeds from physical causes, we must explain it from the laws of nature and not from spontaneity, otherwise we would come to intelligible grounds which belong to the noumenal world <*mundo noumenon*> […]. The intelligences <*intelligentia*> (beings of understanding and free beings, and thus also human beings) belong to the noumenal world <*mundo noumeno*>. (*MM* 29:926)

These passages confirm that the distinction between the world of sense and the world of understanding in terms of which Kant characterizes the two ways of regarding oneself in his published works indeed map onto the distinction between phenomena and noumena and that insofar as it possesses higher faculties, the mind must be considered as a member of the noumenal world.

The latter passage just quoted also provides a further context to the notion of an intelligence. To call something an intelligence is to identify it as an explanatory ground that does not belong to the phenomenal world, governed by the laws of nature, but rather an "intelligible" ground which belongs to the noumenal world. Thus, the notion of an intelligence appears to be tantamount to the notion of an explanatory ground according to laws of freedom. Like his distinction between the higher and the lower faculties of the mind, Kant's characterization of the mind as an intelligence also has a special significance with regard to the explanatory goal of the critique. As we have seen, to name a certain faculty a higher faculty is to characterize it as a self-active power, a power that does not depend on

affection by objects and, hence, as a potential explanatory ground of *a priori* representations. Similarly, to characterize the mind as an intelligence is not simply to characterize it as a noumenon but also to identify it as an explanatory ground of an action (physical or mental), one that is independent from laws of nature.

## 3.5 The Noumenalist Reading

The textual materials we have just considered agree on this point: to the extent that the mind is a bearer of higher faculties, it is to be regarded as a noumenal entity. Since the higher faculties of the mind are the primary focus of the critique (because its aim is to explain the possibility of *a priori* cognition), Kant's central claims about the mind cannot be claims about an appearance, as the phenomenalist interpretation take them to be. The considerations that speak against the phenomenalist interpretation seem to speak in favor of the other branch of the realist interpretation, according to which Kant's claims about the mind are claims about a noumenal entity. I call this the noumenalist interpretation.

Note that the noumenalist interpretation is not just the uncontroversial view that, for Kant, there must be some noumenal thing in itself or other that underlies our conscious representations, considered as inner appearances or objects of inner sense. What I call the noumenalist interpretation is a view concerning the status of Kant's claims about the mind in the critique. It holds that these claims express *cognition* of the way the mind is in itself. On the noumenalist reading, the laws governing the mental faculties that the critique seeks to articulate are the laws of the faculties of the noumenal entity that the mind is. This, as we have seen, seems to be suggested by Kant's claim that these laws are non-empirical and that the faculties they govern belong to an intelligence.

Traditionally, the noumenalist interpretation is endorsed by unsympathetic readers of Kant and underpins their criticism of his so-called transcendental psychology. Their basic point of criticism

is that in making claims about the noumenal character of the mind, Kant thereby contradicts his own doctrine of noumenal ignorance—the doctrine that we have no cognitive access to things in themselves. An influential figure in this tradition is Strawson, whose criticism of Kant we have previously considered. On Strawson's reading, Kant's account of mental faculties consists in a description of "quasi-causal" relations in the realm of things in themselves, to which belong both our mind and things that affect it. Therefore, in his attempt to explain limits of experience in terms of the limits of our noumenal faculties, Kant transgresses what he himself takes to be a requirement of intelligible use of concepts: that concepts be related to empirical conditions of their meaningful use, i.e., spatiotemporal objects instantiating them.[21] Strawson sees this error as resulting from Kant's internalization of space and time to the mind itself. Since space and time, for Kant, are the mind's capacity to be affected in a certain way by objects, the mind itself cannot be in space and time. Yet, Kant allows himself to make substantive claims about the nature of mental faculties. Strawson concludes: "Kant's arguments for these limiting conclusions are developed within the framework of a set of doctrines which themselves appear to violate his own critical principles. He seeks to draw the bounds of sense from the point outside of them, a point which, if they are rightly drawn, cannot exist".[22] Following in Strawson's footsteps, Paul Guyer more recently complains that "the problem with [Kant's transcendental psychology] is not that it is psychological, nor even that it is transcendental, but that it is transcendent—it is based on a claim to certitude in *a priori* knowledge, which Kant does not and cannot sustain".[23]

We can put this criticism in Kant's own terms as follows. Kant's claims about the mind involve a real, ontologically significant use of the categories and derivative concepts. However, according to

[21] Peter Strawson, *The Bounds of Sense* (London: Methuen, 1966), 41; cf. 254-6.

[22] Strawson, 12; cf. 254-6.

[23] Guyer, "Psychology and the Transcendental Deduction," 56–57.

him, intuition is required for any application of the categories to an object that can issue in cognition (in the narrow sense). The reason is that cognition requires that concepts be related to an object that is given to us in intuition (the givenness condition).[24] Without "the possibility of giving it an object to which it is to be related", a concept "has no sense, and is entirely empty of content" (A239/B298). However, due to the discursivity of our understanding, an object can only be given to us through the receptivity of the senses: our understanding "can only think and must seek the intuition in the senses" (B135). Kant sometimes puts the same point by saying that "all intuition that is possible for us is sensible" (B146). Because all our intuition is sensible as opposed to intellectual, objects can only be given to us passively through their affection of our sensibility rather than spontaneously through a creative act of the understanding. However, our sensibility is limited by the condition of its forms, space and time. Therefore, the only objects that can be given to us in intuition and to which the categories can be related in cognition are spatiotemporal objects: "The category has no other use for the cognition of things than its application to objects of experience" (B146). According to Kant, spatiotemporal properties are not features of things in themselves. Hence, the fact that our intuition is sensible and limited by our forms of sensibility entails that we have neither intuition of things in themselves nor, consequently, cognition of them that is grounded in the use of the categories.

The phenomenalist interpretation subscribes to three theses: (i) Kant's claims about the mind are claims about a thing in itself, (ii) these claims involve real, ontologically significant application of the categories, and (iii) these claims qualify as cognition. Now, given the constraints we have just reviewed, it seems that these theses cannot be true simultaneously unless we have available to us *non*-sensible intuition of our own mind as a thing in itself. However, to maintain that we have such non-sensible self-intuition would contradict Kant's claim that our understanding is discursive.

[24] For further discussion of the givenness condition, see Chapter 2.4.

Note that Kant not only rules out the possibility of non-sensible intuition for us in general, but he also takes pains to emphasize that we do not have intuition of *ourselves* that is non-sensible and, thus, can supply basis for cognition of ourselves as things in themselves.

> [For cognition of my self] self-intuition is required, which is grounded in an *a priori* given form, i.e., time, which is sensible and belongs to the receptivity of the determinable. Now I do not have yet another self-intuition, which would give the determining in me, of the spontaneity of which alone I am conscious […]. (B157-58n)

As Kant is pointing out, besides sensible self-intuition, which is conditioned by its form (i.e., time), we do not have "yet another" self-intuition, that is, non-sensible intuition that would enable us to cognize ourselves as spontaneous noumenal beings. Therefore, "I […] have **no cognition** of myself **as I am**, but only as I **appear** to myself" (B158). In the context of the Paralogisms, Kant adduces the same fact to explain why the rational psychologists' attempt to ground cognition of the thinking self as a thing in itself in pure apperception cannot possibly succeed.

> It is in [empirical intuition of the self] that the thinking self must now seek the conditions of the use of its logical functions for categories of substance, cause, etc., so as not merely to indicate itself as object in itself through the "I," but also to determine its kind of existence, i.e., to cognize it as noumenon; which, however, is impossible, since inner empirical intuition is sensible, and makes available nothing but data of appearance, which affords nothing for knowledge of the separate existence of the object of **pure consciousness**, but can serve merely in behalf of experience. (B430)

For this reason, Kant takes the fundamental mistake of rational psychology (or at least one of its fundamental mistakes) to lie in mistaking pure apperception (which is merely formal consciousness of the logical subject) to be non-sensible intuition of the noumenal self.

> Rational psychology has its origin in a mere misunderstanding. The unity of consciousness, which grounds the categories, is here taken for an intuition of the subject as an object and the category of substance is applied to it. But this unity is only the unity of **thinking**, through which no object is given; and thus the category of

> substance, which always presupposes a given **intuition**, cannot be applied to it, and hence this subject cannot be cognized at all. (B421-22)

Thus, Kant seems to be deprived of any basis upon which presumptive cognition of the mind as a noumenon can be sustained. Moreover, Kant seems to hold that the Critical account of mental faculties rests upon the consciousness of our own mental activity through pure apperception (see Chapter 1). As we have just seen, Kant expressly denies that pure apperception constitutes intuition and, thus, no *object* whatsoever is given through it. Therefore, it seems unlikely that the purported starting point of the critique—the non-empirical consciousness of our own faculties enabled by pure apperception—can underwrite claims about the noumenal character of the mind. If the noumenalist interpretation is correct, we seem to have no other option but to declare Kant guilty of inconsistency.

## 3.6 Circumventing Noumenal Ignorance

More recently, some commentators have attempted to defend the noumenalist interpretation as a charitable reading of Kant. According to these commentators, although Kant makes substantive claims about the mind as a thing in itself, he does not thereby violate his own doctrine of noumenal ignorance. In this section, I will consider two recent approaches, taken by Julian Wuerth and Colin Marshall. Wuerth's 'immediatist' interpretation holds that we have some sort of immediate (thus, intuition-like) consciousness of the noumenal self, which can warrant certain indeterminate epistemic claims about what the self is like in itself.[25] In contrast, on Marshall's 'effect-relative' reading, we do not have immediate acquaintance with the noumenal self but only with its effects (i.e, representations); nonetheless, this suffices for us to individuate and refer to the noumenal self, given that the effects

[25] Julian Wuerth, *Kant on Mind, Action, and Ethics* (Oxford: Oxford University Press, 2014), especially Chapter 5.

are metaphysically prior to the self.[26] I will argue that neither of these approaches provides a satisfactory defense of the noumenalist interpretation against the objection from noumenal ignorance.

*The Immediatist Reading.* Although Wuerth does not think that self-cognition presents us with an exception to the noumenal ignorance thesis, he does take it to be the only one case in which we have cognitive access to the noumenal realm.

> In only one case, namely, that of our relation to ourselves, do we stand in the relations to an existing thing in itself of being the thing in itself. In this one case, we are *immediately* aware of being *a something in general that has powers*—including those cognitive powers that Kant ascribes to us in the first *Critique*, of sensibility, imagination, judgment, understanding, and reason—*by mean of which powers we relate to our accidents*, i.e. *our mental states, which inhere in us.*[27]

The immediate awareness that enables us to cognize ourselves in this way is none other than pure apperception. Pure apperception substitutes for the missing intuition of the noumenal self and thereby provides an epistemic ground for applying the categories to the thinking subject. According to Wuerth, through apperception, we are aware of ourselves or our noumenal soul as a simple, identical substance—"something that *grounds* our mental *accidents* through its exercise of its own *powers*", hence, "a *substance* in the *most basic ontological sense*".[28] This immediate awareness of the soul licenses us to apply pure, unschematized, yet ontologically significant categories to it.

Wuerth argues that his proposal is compatible with the noumenal ignorance thesis as well as the rejection of rational psychology in the Paralogisms. His argument turns on a distinction between *indeterminate* and *determinate* ontological significance. Indeterminate ontological significance is what the categories possess in abstraction from the specific form of our sensible intuition and, hence, what they

[26] Colin Marshall, "Kant's Metaphysics of the Self," *Philosophers' Imprint* 10 (2010): 1–21.

[27] Wuerth, *Kant on Mind, Action, and Ethics*, 127.

[28] Wuerth, 183.

possess qua *unschematized* categories. According to Wuerth, the categories retain indeterminate ontological significance when they are applied to a given object (such as the noumenal self) without being related to empirical intuition. For example, the unschematized category of substance is the concept of "that which exists without being the determination of another".[29] In contrast, determinate ontological significance is what the categories possess *qua* schematized. For example, the category of substance when schematized implies permanence. Wuerth holds that the categories when applied to the soul on the basis of apperception have indeterminate ontological significance.[30] Since, in his view, the noumenal ignorance thesis and the Paralogisms only rule out claims with determinate ontological significance about things in themselves in general and the soul in particular, Kant's claims about the noumenal character of the mind, being indeterminate, violate neither of these constraints.

For Wuerth, indeterminate ontological significance is not equivalent to merely logical significance. This difference is illustrated, with respect to the category of substance, by the contrast between a logical subject of predication and a real subject of inherence. A logical subject is related to its predicate by being the logical ground of its assertion, whereas a real or ontological subject (i.e., a substance) is related to its predicate (or accident) by being the real ground of its inherence. In the former case, the relation between subject and predicate is characterized by a *logical form* (of categorical judgment); in the latter case, the relation between subject and predicate is characterized by a *category* (of substance).

On Wuerth's reading of the Paralogisms, Kant's critique of rational psychology is not directed simply at its attempt to apply the categories to the soul as a thing in itself. Rather, what is at issue is that "the rationalists are not willing to restrict themselves to the indeterminate offerings of our

[29] *MM* 29:771, quoted in Wuerth, 136.

[30] Cf. Wuerth, 134–45.

immediate apperception of ourselves as things in themselves".[31] As a result, they illegitimately infer the soul's substantiality, simplicity, and identity in the determinate, empirical sense from the claims about the soul's substantiality, simplicity, and identity in the indeterminate sense, which can be justified by pure apperception. Thus, from the "useless" propositions about the soul's ontological features, the rational psychologists mistakenly infer their "useful" counterparts, which affirm the soul's permanence, incorruptibility, and personality. According to Wuerth, the confusion of these two senses of the categories just is what Kant refers to by the "equivocation of the middle term" common to all the paralogisms. Hence, Kant's rejection of rational psychology does not indiscriminately rule out all ontologically significant claims about the soul but only those that are determinate and fallaciously inferred from the indeterminate claims underwritten by pure apperception.

Before we turn to the objection to this reading, let me mention one serious weakness of Wuerth's account, which is not philosophical but methodological. To support his contentious thesis that pure apperception delivers immediate awareness of the noumenal soul, Wuerth draws upon an impressive range of textual evidence (there is no need for us to examine this in detail). The sheer amount of textual reference, however, could easily obscure the fact that almost all of it comes from the unpublished corpus, i.e., Kant's personal notes and transcripts of his lectures. As far as I can see, Wuerth has presented no unambiguous evidence from Kant's published works that the immediatist thesis represents his considered view in the Critical period. Any thoughts recorded in Kant's personal notes and lectures from the Critical period should be taken as evidence of his official endorsement only if the same thoughts are asserted or at least indirectly suggested in Kant's published works or else illuminate what he says therein. However, this is not the case with the unpublished passages that

[31] Wuerth, 160.

Wuerth cites in support of his thesis. Thus, despite appearance to the contrary, the basic thesis of the immediatist reading suffers from a lack of textual support.

Moreover, as I will now argue, the thesis that pure apperception delivers immediate awareness of the noumenal soul is inconsistent with some of Kant's views which he explicitly expresses in published works. This thesis effectively assimilates pure apperception to non-sensible intuition, which raises a question about how it can be squared with the discursivity of our understanding. We have seen that the reason we cannot have cognition of things in themselves is that the givenness condition cannot be fulfilled in the case of things in themselves, since we do not have non-sensible intuition, which is not limited by our spatiotemporal forms of sensibility. Wuerth's strategy in circumventing this constraint on cognition is to argue that pure apperception can fill in the role of non-sensible intuition that enables us to cognize our noumenal soul. This implies that pure apperception, on his reading, is an awareness through which the noumenal soul is *given* as an object (this is also suggested by his characterization of apperception as *immediate* awareness of the noumenal soul).

However, to claim that pure apperception is a kind of awareness through which an object is given amounts to holding that through apperception a *manifold* of representation is given. The categories are nothing but forms of transcendental synthesis, and they can be related to a given object only by way of being the rules according to which the understanding combines the manifold given in intuition.

> [The categories are] nothing other than **forms of thought**, which contain merely the logical capacity for unifying the manifold given in intuition in a consciousness *a priori*; thus if one takes away from them the only sensible intuition possible for us, they have even less significance than those pure sensible forms, through which at least an object is given, whereas a kind of combination of the manifold that is proper to our understanding signifies nothing at all if that intuition in which alone the manifold can be given is not added to it. (B305-6)

Thus, pure apperception qua awareness of the noumenal soul cannot lend objective "significance" to the categories and underwrite their application to the noumenal soul unless it makes available a manifold of some sort which can be taken up by the understanding's act of combination.

This raises a difficulty for the immediatist reading. For, according to Kant, the capacity to give a manifold through self-consciousness just is a mark of the intuitive understanding: "An understanding, in which through self-consciousness all of the manifold would at the same time be given, would **intuit**" (B135); thus, the intuitive understanding is "that understanding through whose self-consciousness the manifold of intuition would at the same time be given" (B138). In this sense, Wuerth's interpretation assimilates pure apperception to the kind of self-consciousness characteristic of an intuitive understanding. This contradicts Kant's emphasis that our understanding is discursive and, hence, does not intuit. It also contradicts his corresponding characterization of pure apperception as devoid of anything manifold: a discursive understanding is "one through whose pure apperception in the representation **I am** nothing manifold is given at all" (B138); "through the I, as a simple representation, nothing manifold is given; it can only be given in the intuition, which is distinct from it" (B135).[32] Therefore, the immediatist strategy to defend the noumenalist interpretation against the objection from noumenal ignorance fails.

Note that Wuerth's claim that the purportedly legitimate claims about the noumenal soul have *indeterminate* rather determinate ontological significance makes no difference with regard to the way the givenness condition applies to them. In other words, to apply the categories to a noumenal object

[32] Note that Kant seems to grant that we can cognize our noumenal *existence* through apperception but not the *way* we exist noumenally: "The **I think** expresses the act of determining my existence. The existence is thereby already given, but the way in which I am to determine it, i.e., the manifold that I am to posit in myself as belonging to it, is not yet thereby given" (B157n). I take it that the categorial predicates (including, substantiality, simplicity, identity) represent particular ways of existing and for this reason they require a manifold. See also Kant's discussion of Descarte's Cogito in the Paralogisms.

in such a way that yield claims that have indeterminate ontological significance and yet qualify as cognitions would still require intuition (though *per impossibile* of a non-sensible kind). Consider Kant's following remark.

> **Our** sensible and empirical intuition alone can provide [the categories] with sense and significance. Thus if one assumes an object of a **non-sensible** intuition as given, one can certainly represent it through all of the predicates that already lie in the presupposition that **nothing belonging to sensible intuition pertains to it** thus it is not extended, or in space, that its duration is not a time, that no alteration (sequence of determinations in time) is to be encountered in it, etc. […] What is most important here is that not even a single category could be applied to such a thing, e.g., the concept of a substance, i.e., that of something that could exist as a subject but never as a mere predicate; for I would not even know whether there could be anything that corresponded to this determination of thought if empirical intuition did not give me the case for its application. (B149)

In this remark, Kant is ruling out the possibility of applying the categories to a thing in itself exactly in the manner that Wuerth calls indeterminate. Kant first assumes that non-sensible intuition is really possible and then asks whether the categories could be applied to the object of such intuition (i.e., a noumenon in the positive sense). Since we are not concerned with a spatiotemporal object, we are clearly not considering the possibility of applying a schematized category to that object. Thus, the question being raised can only be understood as whether an *unschematized* category can be applied to a noumenon in such a way that allows us to cognize it. Kant answers with an unequivocal no. The reason is that such an object cannot be given in sensible intuition, the only intuition that is available to us. That this verdict is directed at the unschematized use of the categories is confirmed by the fact that Kant glosses the concept of a substance not as the concept of something permanent but as "that of something that could exist as a subject but never as a mere predicate". It is *this* determination of thought which Kant says we cannot know whether any object could correspond to unless it is given in empirical intuition. It is this determination of thought that requires intuition in order to become

cognition. Therefore, Wuerth's characterization of the purportedly legitimate claims about the soul as indeterminate does not enable them to elude the restriction of the doctrine of noumenal ignorance.

*The Effect-Relative Reading.* Another strategy that promises to avoid the objection from noumenal ignorance also proceeds from a certain understanding of pure apperception: the "intuitive" assumption that the representation 'I' is referring, and the self just is whatever entity 'I' refers to.[33] The question, then, is whether Kant has a positive metaphysical account of the self or the referent of the 'I' and one that is compatible with his claim that we lack cognition of the self as it is in itself. Marshall argues that Kant does and this can be found in his account of the *identity* of self-consciousness in the Transcendental Deduction. In Marshall's view, this account is to be read not only as an account of a formal feature of self-consciousness but also a metaphysical account of the nature of the self as an entity existing outside of representations.

Marshall begins by appealing to an intuitively plausible view that any token of 'I' refers to whatever produced it; if a collection of entities produced the token representation, then 'I' refers to that collection of entities. If this view is correct, then it is plausible to say that the self is constituted by whatever entity or entities produced the thought in which 'I' occurs. Marshall goes on to argue that we can eliminate the requirement that 'I' explicitly occur in the thought that individuates the self—a thought can serve to individuate the self as long as it is sufficiently unified in such a way that makes it possible for the entity or entities that produced it to self-ascribe that thought using an 'I' (in Kant's terms, as long as the thought can be accompanied by the I think). The metaphysical view that emerges from this is one according to which the self is individuated by "an entire course of experience […]

[33] Marshall, "Kant's Metaphysics of the Self," 13. Note that Marshall does not argue for this assumption but only points to its intuitive appeal and claims that "there is no reason to doubt that Kant accepted this".

constituting a single thought (with smaller thoughts and perceptions as components)" provided that that course of experience can be self-ascribed as a single 'I'-thought.[34]

On the effect-relative reading, therefore, facts about the unity of representations are metaphysically prior to facts about the metaphysics of the self—the former metaphysically explain the latter. Thus, Kant's well-known claim that the identity of self-consciousness depends on the unity of representations can be read as a non-trivial metaphysical claim about the identity of the self. An entity or a set of entities constitutes a self in virtue of being causally responsible for the unity of representations: "For any particular unified experience, whatever [noumenal] thing or things are immediately causally responsible for the unity of that experience compose a self"[35], whereby experience is understood in the sense of a single unified course composed of smaller experiences. Thus, a self is individuated by the experience to whose unity it causally contributes. This leaves it open whether the self is a single or multiple substances or accidents. This possibility, according to Marshall, need not contradict the idea that the self is a single individual if we allow that a set of entities can constitute an individual in virtue of their joint causal contribution to a single effect (the production team for a film is a single individual in this sense).

Marshall contends that the effect-relative reading helps to account for Kant's metaphysical commitments that require him to have some positive account of the self without holding him to violate his own doctrine of noumenal ignorance. Although 'I' refers to a noumenal entity or entities on this reading, it remains neutral on whether the noumenal self is one or more substances or accidents. Thus, there is clear sense in which we lack cognitive insight into the nature of the noumenal self apart from its effects.

---

[34] Marshall, 14.

[35] Marshall, "Kant's Metaphysics of the Self," 16. For Marshall's statement that the self must be composed of noumenal entity or entities, see footnote 66 on p. 16.

One of Kant's metaphysical commitments that motivate the effect-relative reading is his commitment to "a conception of a determinate mind with faculties" that underlies his Copernican approach to philosophy. Thus, Marshall shares the view, which I have been arguing for, that Kant's project "makes sense only if we understand the mind as being a genuine entity with a nature of its own that is more basic than the world of appearances, and where at least some aspects of that nature are known *a priori*".[36] However, Marshall does not explain how exactly the effect-relative conception of the self is supposed to solve the problem of Kant's claims about the mind and its faculties, especially if we assume that the mind is a noumenal entity. Suppose that the mind that these claims are about is whatever entity or entities that constitute the self according to the effect-relative reading. While the conception of the self that underlies these claims may abide by the doctrine of noumenal ignorance, it is unclear how the specific claims about faculties (which are not entailed by that conception) are not in violation of it. In particular, it is not obvious what on this reading makes the application of the categories to a noumenal entity (supposedly involved in those claims) possible in the absence of intuition of that noumenal entity.

A deeper but related issue with the effect-relative reading is this: it takes for granted that we have a grasp of our own representation or experience as an *effect* of a cause.[37] But what grounds this grasp? It seems that in order to individuate our self, according to this reading, it must be possible for us to apply the category of causality to our representation or experience (as an object). But what warrants this application? According to Marshall, as long as we are *aware* of a unified course of experience, we have sufficient basis for individuating our self and referring to it (by the 'I') with immunity to reference failure. But what kind of awareness is meant here? Since we are not concerned

[36] Marshall, 7.

[37] I have already shown in the introductory chapter why, in the light of Kant's view, this is not something to be taken for granted.

with empirical causation, this cannot be a matter of cognizing our representation as effect of an empirical cause. It follows that it can only be the non-empirical awareness we have of our representation or experience, i.e., pure apperception, that warrants our application of the category of causality to our representation. As I have previously argued, however, pure apperception can warrant the application of a category only if it is taken to be an awareness through which an object or a manifold is given. Hence, it turns out that, like the immediatist reading, the effect-relative reading also implicitly assimilates pure apperception to non-sensible intuition (perhaps not of the self as a noumenal substance but of our representation as effect of a noumenal cause). To this extent, the effect-relative reading is vulnerable to the objections I have raised against the immediatist reading.

## 3.7 Conclusion

In this chapter, I have argued that neither of the two versions of the realist interpretation—the phenomenalist and the noumenalist interpretation—offers a satisfactory account of the status of Kant's claims about the mind in the critique. As we have seen, Kant thinks of the laws of the mental faculties investigated in the critique as non-empirical laws, and he characterizes the mind *qua* the bearer of higher faculties as an intelligence, a being not of the sensible but of the intelligible world. This seems to indicate that the mind is not an appearance and the claims Kant makes about it are not empirical propositions warranted by experience. These considerations strongly suggest, on the contrary, that the mind is a noumenal entity, and that Kant's claims regarding it are *a priori* claims about the way the mind is in itself. However, it is difficult to see how one can follow this suggestion without imputing incoherence to Kant due his commitment to noumenal ignorance. We have considered two attempts at avoiding this unattractive consequence, neither of which, I have argued, succeeds.

There is, however, an option which have not been considered so far, namely, to deny that Kant's claims about the mind, to the extent that they involve a use of the categories and derivative ontological concepts, qualify as cognitions in the strict sense. We can agree with the noumenalist interpretation that (i) Kant's claims about the mind are claims about a thing in itself and (ii) these claims involve real, ontologically significant application of the categories, while denying that (iii) these claims qualify fully as cognitions. This interpretive possibility is what I will explore in the remainder of this study. I should emphasize from the start that denying this is not to deny that Kant's claims about the mind in critique make no epistemic contribution whatsoever. Indeed, we can plausibly consider these claims as instances of *formal* cognition (thus, in the same sense in which the propositions of logic are cognitions; see Chapter 2). What we want to deny is that the framework of concepts in terms of which these claims are formulated—their metaphysical scaffolding, so to speak—is grounded in cognition of the mind as a thing in itself. In this way, the alternative reading that I will be defending share features of both the anti-metaphysical interpretation and the noumenalist interpretation.

A closer look at the passages that appear to support the noumenalist reading lends initial plausibility to this alternative. In those passages, Kant sometimes seems to straightforwardly assert that the mind or the human being *qua* the bearer of higher faculties is a noumenal entity (an intelligence, an intelligible object, a member of the intelligible world). However, he very often formulates his claim more carefully as the claim about the way the subject must *regard* herself [sich betrachten, sich ansehen]: "A rational being must regard himself as intelligence (hence not from the side of his lower powers) as belonging not to the world of sense but to the world of understanding" (*GMS* 4:452). It is plausible to couple such claims with Kant's remark that "everyone must necessarily regard Himself [Sich selbst ansehen] as a substance, but regard his thinking only as accidents of his existence and determinations of his state" (A349). I want to suggest that these remarks contain clues for solving the problem of Kant's claims about the mind. They imply how we should understand the

conception of the mind (i.e., the broadly Cartesian conception of the mind as a noumenal substance endowed with representational faculties) that underlie these claims: this conception does not represent (for all we know) what we are in ourselves but how we ought to *regard* or *conceive* of ourselves.

To defend this reading further will require us to give a positive account of the rational basis for asserting such a self-conception (of its "warranted assertability", as Allison calls it). This is what I aim to do in the two final chapters. I will argue that this account can be found in Kant's account of the regulative use of the ideas of reason. Therefore, on my reading, the broadly Cartesian conception of the mind, which underlies Kant's claims about the mind, derives its assertability from the regulative use of the idea of the soul. The significance of the idea of the soul is also hinted at in the same passage in which Kant mentions the necessity of the conception of oneself as a substance.

> One can quite well allow the proposition **The soul is substance** to be valid, if only one admits that this concept of ours leads no further, that it cannot teach us any of the usual conclusions of the rationalistic doctrine of the soul, such as, e.g., the everlasting duration of the soul through all alterations, even the human being's death, thus that it signifies a substance *only in the idea but not in reality*. (A350-51, italics added)

Thus, we can legitimately conceive of ourselves as a noumenal substance if this self-conception signifies a substance "only in the idea but not in reality". But what does this mean? We find an answer in Kant's later discussion of the regulative use of the transcendental ideas, in which he distinguishes two ways in which an object can be "given".

> It makes a big difference whether something is given to my reason as **an object absolutely** or is given only as **an object in the idea**. In the first case my concepts go as far as determining the object; but in the second, there is really only a schema for which no object is given, not even hypothetically, but which serves only to represent other objects to us, in accordance with their systematic unity […]. (A670/B698)

If an object is given "absolutely", it is given in such a way that satisfies the givenness condition of cognition (my concepts can relate to and "determine" this object). I take this way in which an object

can be given to correspond to the use of the category of substance that signifies a substance "in reality". In contrast, if an object is given "in the idea", it is not given for cognition but only as a "schema" of systematic unity of cognition (more on this in the next chapter). Hence, to employ the category of substance such that it signifies a substance "in the idea" is to employ it exhibit the object of a transcendental idea for the purpose of its regulative use. If this is correct, then we can explain why the use of the categories involved in Kant's claims about the mind is legitimate: these claims rest on the legitimate, regulative employment of the idea of the soul, which is the idea of the noumenal bearer of rational faculties.

To make the reading just sketched intelligible will require a revision of the standard interpretation of Kant's account of the regulative use of the transcendental ideas in general. This is what I will undertake in Chapter 4. Then, in Chapter 5, I will show how this general interpretation applies to the case of the idea of the soul and elucidates its role in the Critical investigation.

# Chapter 4

# The Ideas of Reason as Forms of Explanatory Understanding

## 4.1 Introduction

This chapter offers a sketch and a defense of an alternative interpretation of the regulative use of the transcendental ideas that will serve as the basis for my thesis that the regulative use of the idea of the soul underlies Kant's account of mental faculties in the critique. According to the standard reading, the role of the transcendental ideas is to direct *empirical* investigation of nature toward a goal or an end point of inquiry. Therefore, it is not obvious why the idea of the soul should have anything to do with Critical investigation, which is supposed to be *a priori*. I will argue that this reading is mistaken. On my reading, the theoretical role of the transcendental ideas is instead that of forms to explanatory understanding (in Kant's terms, 'comprehension' [Begreifen]), which in Kant's system is a higher degree of cognition. In this way, the role of the transcendental ideas is analogous to that of the categories as forms of empirical judgment. It follows that the regulative use of the ideas must be involved in all sciences (empirical or *a priori*) insofar as they strive toward explanatory understanding. These include the critique, which, as I argued in Chapter 1, is conceived as a science that has an explanatory goal.

Kant's view on the positive use of the ideas is intimately connected with his ambivalent attitude toward metaphysics. While Kant rejects the claims of traditional metaphysics as illegitimate claims of knowledge, he insists that the very same metaphysical enterprise has a positive role to play in our pursuit of morality and scientific knowledge. In the Transcendental Dialectic of the First *Critique*, Kant's aim is, on the one hand, to expose errors at the foundation of the special disciplines of

traditional metaphysics (rational psychology, cosmology, and theology). On the other hand, he seeks to habilitate a positive *theoretical* use of the concepts of the principal objects of these metaphysical disciplines (the soul, the world-whole, and God). In particular, Kant argues that while these concepts have no legitimate *constitutive* use so as to attain knowledge of the supersensible realm, they nonetheless have an "excellent and indispensably necessary" *regulative* use in furnishing cognition of nature with the unity of a system (A644/B672).

On the standard reading, the regulative role of the ideas is to provide guidance or direction for empirical investigation of nature in different domains. Thus, the three ideas correspond to three special sciences: the soul to psychology, the world-whole or cosmos to physics, and God to unified natural science.[1] The adoption of this reading, however, has engendered widespread suspicion. Commentators tend to find Kant's claim that these ideas play a necessary role in each of these sciences to be contrived or arbitrary. Although the three objects of the ideas appear to tenuously correlate with the subject

[1] Statements of the standard reading can be found in Peter Strawson, *The Bounds of Sense* (London: Methuen, 1966).; Jonathan Bennett, *Kant's Dialectic* (Cambridge: Cambridge University Press, 1974).; William Walsh, *Kant's Criticism of Metaphysics* (Edinburgh: Edinburgh University Press, 1975).; Heiner Klemme, *Kants Philosophie des Subjekts* (Hamburg: Meiner, 1996), 229f.; Michelle Grier, *Kant's Doctrine of Transcendental Illusion* (Cambridge: Cambridge University Press, 2001).; Henry Allison, *Kant's Transcendental Idealism: An Interpretation and Defense*, 2nd ed. (New Haven: Yale University Press, 2004).; Frederick Rauscher, "The Appendix to the Dialectic and the Canon of Pure Reason: The Positive Role of Reason," in *The Cambridge Companion to Kant's Critique of Pure Reason*, ed. Paul Guyer (Cambridge: Cambridge University Press, 2010), 290–309.; Camilla Serck-Hanssen, "Der Nutzen von Illusionen: Ist Die Idee Der Seele Unentbehrlich?," in *Über Den Nutzen von Illusionen: Die Regulativen Ideen in Kants Theoretische Philosophie*, ed. Bernd Dörflinger (Hildesheim: Olms, 2011), 59–70.; Corey Dyck, *Kant and Rational Psychology* (Oxford: Oxford University Press, 2014), 208f.; Rachel Zuckert, "Empirical Scientific Investigation and the Ideas of Reason," in *Kant and the Laws of Nature*, ed. Angela Breitenbach and Michela Massimi (Cambridge: Cambridge University Press, 2017), 89–107.; Katharina Kraus, "The Soul as the 'Guiding Idea' of Psychology: Kant on Scientific Psychology, Systematicity, and the Idea of the Soul," *Studies in History and Philosophy of Science Part A* 71 (2018): 77–88.; Katharina Kraus, "The Parity and Disparity Between Inner and Outer Experience in Kant," *Kantian Review* 24, no. 2 (2019): 171–95.; Katharina Kraus, *Kant on Self-Knowledge and Self-Formation* (Cambridge: Cambridge University Press, 2020)..

matters of psychology, physics, and unified natural science, it is not easy to identify the specific guidance that can be obtained by postulating them. William Walsh's terse verdict represents the prevalent opinion: "Kant's appeal to particular ideas is of little or no significance".[2] While some other commentators, such as Henry Allison and Rachel Zuckert, seek to offer a more sympathetic account, their success lies at most in vindicating the spirit of Kant's view rather than offering an account of the exact cognitive role of each particular idea.[3]

In this chapter, I sketch an alternative interpretation which portrays Kant's view in a new and hopefully more favorable light. On the reading I am defending, the transcendental ideas act as forms of cognition of a higher degree than experience and, hence, as conditions of the possibility of this higher degree of cognition. Kant calls this higher degree of cognition 'Begreifen', usually translated as 'comprehension' but perhaps better translated as 'explanatory understanding'. To begin with, I show, in Section 4.2, that relevant text passages do not compel us to accept the basic tenet of the standard reading, which construes the regulative role of the ideas as defined or delimited by distinct domains of empirical inquiry. In Section 4.3, I identify five textual puzzles which the standard reading fails to accommodate. I will argue these puzzles can be solved by adopting the alternative reading, which I sketch and defend in the rest of the chapter.

## 4.2 The Domain Specificity Thesis

As already pointed out, the standard interpretation subscribes to the basic thesis that the regulative role of each transcendental idea is delimited by a domain of inquiry, corresponding to a special empirical science. Call this the domain specificity thesis. The main intuitive appeal of this thesis stems from the fact that it offers a straightforward way to explain why systematic unity of cognition

[2] Walsh, *Kant's Criticism of Metaphysics*, 246.. Cf. Strawson, *The Bounds of Sense*, 160.; Bennett, *Kant's Dialectic*..

[3] Allison, *Kant's Transcendental Idealism, 2nd Ed.*, 423ff.; Zuckert, "Empirical Scientific Investigation.".

requires three different ideas and how their roles are differentiated from one another. As I will show in this section, however, the superficial plausibility of this thesis is undermined by close examination of textual evidence.

The relation between the ideas and their domains has been construed in two different, though not incompatible, ways. In the first way, each idea is said to direct empirical investigation in its domain toward a goal—the unification of all scientific concepts and laws in that domain into a unified theory. This passage from Michelle Grier's seminal study offers a representative statement of this reading.

> Kant apparently seeks to establish that the three transcendental ideas serve as maxims that somehow guide our empirical inquiries. Although he is not at all clear on this issue, it does seem that Kant views the three transcendental ideas as presuppositions that direct our unification of knowledge into scientific theory. More specifically, he suggests that the idea of the soul grounds empirical investigations in psychology, the idea of the world grounds physics, and the idea of God grounds the unification of these two branches of natural science into one unified science.[4]

According to the second way in which the role of the ideas with regard to their domains, these ideas do not (or not only) guide investigation within antecedently determined domains but serve to demarcate the domains and, hence, dictate the subject matters of special sciences in the first place. This passage from Frederick Rauscher exemplifies this reading.

> One would assume that, since [the three transcendental ideas] are derived *a priori*, they play a role in the chart of concepts that is determined *a priori*. That is, the pure ideas would be posited *a priori* as the highest possible concepts to be reached via the methodological principle of homogeneity. They would presumably be *a priori* ideas dictating the domain of sciences at the highest level: soul for psychology, world for physics, and God for some science that ranges over everything. Reason would then be providing the *a priori* structure of the sciences at the broadest level.[5]

[4] Grier, *Transcendental Illusion*, 265.

[5] Rauscher, "The Appendix to the Dialectic," 298.

Both readings share the view that the specific function of each idea is defined by a particular domain of objects, which corresponds to one of the special sciences: psychology, physics, and unified natural science. These domains further accord with the division of the objects of our senses. The domains of psychology and physics correspond, respectively, to the objects of inner sense and the objects of outer sense; the domain of unified natural science corresponds to the objects of sense in general.

To my knowledge, there is no unambiguous textual evidence that Kant himself is committed to the domain specificity thesis. Its apparent plausibility presumably derives from two sources: (i) the association of the ideas with the *metaphysica specialis* and (ii) Kant's conspicuous reference to the distinction between inner and outer sense in his description of the roles of the ideas. Let us examine each of these textual considerations in turn.

Historically speaking, what Kant calls transcendental ideas are the concepts of the objects of the *metaphysica specialis* or special metaphysics. The soul is the object of rational psychology, the world-whole is the object of rational cosmology, and God is the object of rational theology.[6] In the First *Critique*, Kant argues that these spurious sciences result from reason overstepping its boundaries. Therefore, it is tempting to think that when reason is held within its proper limits, there will be three corresponding immanent sciences in which these ideas find their legitimate application. However, I wish to argue that there is no reason to suppose that for every spurious metaphysical science there is a corresponding empirical science. For one thing, empirical science and metaphysics do not investigate the same objects (or the same aspect of objects): whereas empirical science studies appearances, metaphysics strives to attain knowledge of noumenal entities. Now, one might argue that the objects of special metaphysics are the noumenal grounds of the objects of special empirical sciences, and this provides the links between special metaphysics and corresponding empirical sciences. However, the

[6] E.g. A334/B391.

problem with this argument is that, given our cognitive limitation, we are not entitled to assume that a given object of empirical inquiry is related to a particular noumenal object as its noumenal ground, for instance, that the soul is the noumenal ground of inner appearances, which empirical psychology studies. Even if such an assumption were justifiable, it is unclear what advantage it would bring to the empirical investigation, which seeks empirical explanation for empirical phenomena.

What is potentially misleading is that Kant retains the ties between transcendental ideas and traditional metaphysics even when he characterizes their legitimate uses. Consider the following passage from the Appendix to the Transcendental Dialectic (henceforth, Appendix):

> Following the ideas named above as principles, we will **first** (in *psychology*) connect all appearances, actions, and receptivity of our mind to the guiding thread of inner experience **as if** the mind were a simple substance that (at least in this life) persists in existence with personal identity, while its states […] are continuously changing. Then **second** (in *cosmology*) we have to pursue the conditions of the inner as well as the outer appearances of nature through an investigation that will nowhere be completed, **as if** nature were infinite in itself and without a first or supreme member […]. Finally and **thirdly**, (in regard to *theology*) we have to consider everything that might ever belong to the context of possible experience **as if** this experience constituted an absolute unity, but one dependent through and through, and always still conditioned within the world of sense, yet at the same time **as if** the sum total of all appearances (the world of sense itself) had a single supreme and all-sufficient ground outside its range […]. (A672/B700, italics added)

It is relatively obvious that in this passage "psychology", "cosmology", and "theology" refer to the metaphysical sciences. But, as I have just argued, there is no compelling reason to assume further connection between these metaphysical sciences and special empirical sciences. Thus, the references to special metaphysics do not by themselves indicate a commitment to the domain specificity thesis.

In my view, Kant's point in invoking the *metaphysica specialis* in this context is instead to indicate that the doctrines of traditional metaphysics are relevant for the regulative functions of these ideas.

The metaphysical propositions examined in the Paralogisms, the Antinomies, and the Ideal are precisely what gives content to the ideas of the soul, the world-whole, and God. For instance, the content of the idea of the soul is derived from the conclusions of the paralogistic inferences, namely, that the soul is a substance, that it is simple, that it is identical over time, and that it stands in mutual interaction with material substances. As some commentators have argued, the Critical Kant has not completely abandoned synthetic *a priori* principles of pre-Critical metaphysics but only demoted them to subjectively valid, regulative principles.[7] Following this line of interpretation, I maintain that the synthetic *a priori* principles of traditional metaphysics presented in the three main chapters of the Transcendental Dialectic *are* the regulative principles that the ideas embody—these principles make up the content of the *as-if* attitudes that Kant mentions in the passage we have just considered.

As already mentioned, some commentators interpret the regulative role of the transcendental ideas as consisting in dictating or generating the domains of special empirical sciences and, thus, as analogous to their role with respect to the *metaphysica specialis*. On this reading, the ideas fulfil this role by acting as the highest genus concepts of the special sciences.[8] Can this offer a way to establish the connection between special metaphysics and special empirical sciences? The problem with this reading is that it is unclear why highly general, empirical concepts would not better suit the purpose than *a priori* concepts of unconditioned objects. Why would the highest genus concept of empirical psychology, for example, not more properly be that of empirical thinking substance than that of

[7] E.g. Mark Fisher and Eric Watkins, "Kant on the Material Ground of Possibility," *The Review of Metaphysics* 52, no. 2 (1998): 369–95; Eric Watkins, "Kant on Rational Cosmology," in *Kant and the Sciences*, ed. Eric Watkins (Oxford: Oxford University Press, 2001), 70–89; Noam Hoffer, "Kant's Regulative Metaphysics of God and the Systematic Lawfulness of Nature," *The Southern Journal of Philosophy* 57, no. 2 (2019): 217–39; Uygar Abaci, *Kant's Revolutionary Theory of Modality* (Oxford: Oxford University Press, 2019), 208–48.

[8] E.g. Rauscher, "The Appendix to the Dialectic," 296.; Serck-Hanssen, "Der Nutzen von Illusionen."; Kraus, *Kant on Self-Knowledge*, 235ff.

noumenal thinking substance (soul)? In fact, Kant's view seems to be that the highest genus concept of each special science is an empirical concept. In the *Metaphysical Foundations*, he explicitly identifies the highest genus concepts of empirical psychology and physics as the *empirical* concepts of thinking being and matter (*MAN* 4:470).[9] These concepts are clearly not identical to the ideas of the soul and the world-whole, which are pure concepts.

We must conclude, therefore, that the fact that the ideas are associated with special metaphysical sciences gives us no reason to suppose that their legitimate uses are tied to the domains of special empirical sciences. Let us turn to the second presumptive textual support for the domain specificity thesis and see if this fares better than the first. The distinction between inner and outer sense seems to play a crucial role in Kant's characterization of the regulative roles of the psychological and the cosmological ideas. In the passage we considered above, Kant says that through the idea of the soul we "connect all appearances, actions, and receptivity of our mind to the guiding thread of *inner experience*" (A672/B700, italics added). In another passage, he states that the regulative use of this idea is concerned with the "systematic unity of all the appearances of inner sense", which cannot be attained by means of the empirical concept of "that which the soul actually is" (presumably the empirical concept of thinking being). The passage continues:

> With [the idea of the soul], however, reason has nothing before its eyes except principles of the systematic unity in explaining the appearances of the soul namely by considering all determinations as in one subject, all powers, as far as possible, as derived from one unique fundamental power, all change as belonging to the states of one and the same persisting being, and by representing all **appearances** in space as entirely distinct from the actions of **thinking**. (A682-3/B710-1)

[9] Cf. A848/B876. Kant also describes the empirical concept of matter as "the highest empirical principle of the unity of appearances" (A618/B646).

The recurring references to inner sense and inner experience in these passages make it seem natural to suppose that the regulative role of the idea of the soul is specific to the domain of objects of *inner* sense (i.e. the mind and mental states) and, correspondingly, that the cosmological idea operates exclusively within the domain of objects of *outer* sense (i.e. material objects).

However, contrary to what one might expect, Kant's description of the role of the cosmological idea lays equal emphasis on inner and outer senses. In employing this idea, we "pursue the conditions of the *inner* as well as the *outer* appearances of nature through an investigation that will nowhere be completed" (A672/B700; italics added).[10] The cosmological idea concerns "nature in general, and the completeness of conditions in it in accordance with some one principle", whereby nature "is twofold: either thinking nature or corporeal nature" (A684/B712). Thus, if we can infer anything at all about the domains in which the ideas operate from these descriptions, it would be that the domain of idea of the soul (i.e., objects of inner sense) is contained in the domain of the idea of the world-whole (i.e., objects of inner and outer sense), which, in turn, appears to be co-extensive with the domain of the idea of God (i.e. objects of sense in general). Suppose this is the case; it is unclear what is left of the idea that the domains are what differentiate the three ideas functionally from one another and, hence, of the domain specificity thesis.

Further, from Kant's claim that the idea of the soul serves to bring systematic unity to inner experience or inner appearances, it does not follow that this idea is only relevant for cognition of mental items. In Kant's view, the objects of inner sense are not limited to mental items but also include objects of outer sense. Otherwise, it would not be the case that objects of outer sense are also in *time*, which is the form of inner sense. Kant explains this point as follows.

[10] Kant also says that the cosmological idea is derived from the relation of representation to objects as appearances (A333-4/B390-1). Note that here no distinction is made between inner and outer appearances.

> Time is the *a priori* formal condition of all appearances in general. Space, as the pure form of all outer intuitions, is limited as an *a priori* condition merely to outer intuitions. But since, on the contrary, all representations, whether or not they have outer things as their object, nevertheless as determinations of the mind themselves belong to the inner state, while this inner state belongs under the formal condition of inner intuition, and thus of time, so time is an *a priori* condition of all appearance in general, and indeed the immediate condition of the inner intuition (of our souls), and thereby also the mediate condition of outer appearances. (A34/B50)

Since outer objects are also "representations" and, hence, "determinations of the mind", they are at the same time objects of inner sense, though *mediate* ones. For this reason, time, the form of inner sense, is a condition not only of inner appearance but of "all appearance in general". Consequently, the connection between the idea of the soul and inner sense does not entail that the regulative role of this idea does not extend to outer appearances. On the contrary, as I will argue later, the idea of the soul plays a role in the explanation of outer objects insofar as these are considered *representations*. If this is correct, it cannot be maintained that the functional division between the psychological and the cosmological idea is explicated in terms of the distinction between inner and outer sense.

Finally, we might question the assumption that the privileged link between inner experience and the idea of the soul entails that its function is specifically related to *empirical* cognition of the self or the mind. Kant holds that the psychological idea derives its content from the relation of representations in general "to the subject" (A333-4/B390-1). Interestingly, however, he contrasts this relation with the relation of representations "to objects", where "objects" explicitly include objects of both inner and outer senses. This indicates that the "subject", in the former relation, is not to be understood as an object of empirical cognition. If this is correct, then the idea of the soul is grounded in a relation of representations in general to the subject that does not entail empirical cognition of the subject as an object.

This conclusion is confirmed by a passage from the first-edition Paralogisms, where obviously the same relation is paraphrased as the relation of representations to "the conditions of a thought in general" (A397). These conditions, Kant emphasizes, are "not objective at all"; rather, they pertain to "the I, in the universal proposition 'I think'", i.e., the I of pure apperception (A398).[11] As we have discussed in the previous chapters, the self-awareness in which pure apperception consists does not deliver cognition of the self as an object (phenomenal or noumenal) but only denotes the relation of every representation to an identical logical subject. Consequently, the apperceptive subject, whose absolute unity the idea of the soul represents, must be sharply distinguished from the *empirical* subject, an object of inner experience. The apperceptive subject "is only the formal condition, namely the logical unity of every thought, in which I abstract from every object" (A398).[12]

## 4.3 Five Interpretive Puzzles

In the last section, I attempted to undermine the apparent plausibility of the domain specificity thesis through a close examination of potential textual evidence for it. Now, I will point out that the standard reading, which proceeds from the assumption of this thesis, leaves a number of central aspects of Kant's account deeply puzzling. Later, I will argue that my alternative interpretation can accommodate these puzzles better than the standard reading.

First, on the standard reading, it is not obvious how, given their specific content, the ideas are fit for the role attributed to them. More specifically, it is unclear what useful or relevant guidance scientific investigation can draw from the postulation of the unconditioned objects these ideas

[11] In a pre-Critical reflection (ca. 1778-79), Kant simply states that the idea of the soul derives from "apperception" instead of "relation to the subject" (R5553, 18:229). For illuminating discussion of this reflection, see Paul Guyer, "The Unity of Reason: Pure Reason as Practical Reason in Kant's Early Conception of the Transcendental Dialectic," *The Monist* 72, no. 2 (1989): 139–67.

[12] Cf. A334/B391.

represent. The soul, the world-whole, and God are unconditioned objects and, hence, in principle unknowable. In what way are the concepts of these objects relevant for the investigation of the objects of empirical science, which are conditioned and knowable?

Second, it is puzzling how the ideas can derive their regulative or directive force from the kind of attitude (i.e., the as-if attitude) in which are deployed. According to Kant, the ideas fulfil their regulative role by figuring in the attitude of *regarding* or *conceiving* of "all the connection of things in the world of sense **as if** they had their ground in this being of reason [i.e., the soul, the world-whole, or God]" (A681/B709). Thus, we employ the ideas as constituents in an essentially assertoric attitude towards objects of experience, an attitude of taking these objects to be a certain way, which, however, is characterized by mean of analogy (i.e., the as-if). However, it is not obvious how such an as-if attitude can give rise to any directives for ongoing empirical investigation that are non-trivial and sufficiently specific.

Third, the domain specificity thesis stands in tension with Kant's characterization of the transcendental ideas in terms of real conditioning relations. Just as the categories originate from the understanding in its "real use", so the ideas originate from reason in its real use.[13] Reason in logical use is concerned with forms of inferences. Reason in real or pure use is concerned with real conditioning relations, which are, roughly, explanatory relations of metaphysical dependence. The three transcendental ideas are supposed to represent the three species of real conditioning relation (discussed in the three main chapters of the Transcendental Dialectic) by representing the objects that are unconditioned with respect to each of these species of real condition. We will discuss real conditioning relations in greater detail in Section 4.6. The question this raises is this: what is the relationship between the three conditioning relations and the special sciences within which the ideas

[13] Cf. A299/B355.

purportedly operate? The standard interpretation seems to make sense only if each species of real conditions is relevant for one special science only. To mention just one example, causation is associated with the idea of the world and, hence, the standard reading would seem to suggest that causation belong properly to physics. But is it not the case that psychology is also interested in (mental) causation and causal explanations (for mental states)?

Fourth, a related puzzle: Kant also holds that the three transcendental ideas are derived from the "conditions of all representations in general" (A405/B432; cf. A333-4/B390-1) or "the universal conditions of thinking" (A397).[14] The idea of the soul is derived from the *subjective* conditions of representations. The other two ideas are derived from two sorts of *objective* conditions: (i) the idea of the world is derived from objective conditions pertaining to appearance and (ii) the idea of God is derived from objective conditions pertaining to "objects of thinking in general" or "the possibility of objects in general". According to Kant, the "absolute unity" or the "absolute totality of synthesis" of these three conditions constitutes the content of the transcendental ideas. A similar question can be raised: what is the relationship between these conditions of thinking or representation and the special sciences withing the ideas are supposed to operate? The general problem, then, seems to be that there is a mismatch between the domain-specific roles that the standard reading attributes to the ideas and what Kant explicitly says they contain or represent.

Fifth, according to Kant, the ideas employed by reason in its theoretical use and its practical use (with respect to its postulates) are one and the same ideas (with somewhat different designations).[15] As Kant explains, pure practical reason does not generate new ideas but only provides objective reality to the ideas originating from reason in its theoretical use.[16] On the standard reading, the theoretical

[14] Cf. A333-4/B390-1, A397ff., A405-6/B432-33.

[15] Cf. B395n.

[16] Cf. *KpV* 5:136.

role of the ideas is completely different and unrelated to their practical role. However, given Kant's thesis that reason is a unity, it is reasonable to expect that the role of the ideas with regard the theoretical use of reason would in some way be continuous with their role with regard to the practical use of reason. While this consideration does not necessarily show that the standard reading is flawed in failing to accommodate it, it does give us a reason to prefer an alternative interpretation that can help to elucidate such continuity.

## 4.4 Reason, Comprehension, and the Unconditioned

The alternative interpretation that I am about to sketch offers more satisfactory responses to these puzzles that the standard interpretation, or so I will argue. I hope that it will also become clear that the view my reading attributes to Kant is philosophically more attractive, while also remaining faithful to textual evidence.

In rejecting the domain specificity thesis, I propose that we should instead construe the transcendental ideas in their regulative use as forms of explanatory understanding (what Kant calls 'comprehension' [Begreifen]), the understanding *why* something is the case, which, as I will argue, is not limited to empirical explanations or explanations by appeal to natural causal laws. In this way, the positive cognitive role of the ideas is analogous to that of the categories, which serve as forms of empirical cognition (with respect to its conceptual or judgmental component). Thus, there are as many distinct forms of explanatory understanding as there are transcendental ideas (namely, three). We can say that the transcendental ideas are conditions for the possibility of explanatory understanding, just as the categories are conditions for the possibility of empirical cognition. Herein lies the rational basis for our employment of the ideas: to the extent that explanatory understanding can be shown to be possible, we are rationally justified in employing the ideas and asserting the existence of their objects. Note, however, that there is an important difference between the warrant we have for employing the

transcendental ideas and the warrant we have for employing the categories. The categories have objective validity without qualification because we can unproblematically assume experience to be possible without qualification. In contrast, the transcendental ideas have only subjective validity or (at most) "indeterminate" objective validity, because the extent or the degree to which explanatory understanding is possible is undetermined (we do not know, to begin with, whether things are so constituted such that a complete explanation of a given fact is possible).

To see why this reading is plausible, let us begin with an architectonic consideration. The main sections of the first Critique follow the progression from the lowest cognitive faculty to the highest cognitive faculty: "All our cognition starts from the senses, goes from there to the understanding, and ends with reason, beyond which there is nothing higher to be found in us to work on the matter of intuition and bring it under the highest unity of thinking" (A298/B355); "all human cognition begins with intuitions, goes from there to concepts, and ends with ideas" (A702/B730). The Transcendental Aesthetic focuses on the sensibility, the lower faculty of cognition. The Transcendental Analytic proceeds to the higher faculty of cognition, beginning with the first sub-faculty, the understanding, and ending with the second sub-faculty, the power of judgment. Finally, the Transcendental Dialectic is devoted to reason, the third sub-faculty. In the Aesthetic and the Analytic, Kant is occupied with articulating the *a priori* forms of representations (and formal representations) originating in the respective faculties (space and time, the categories, and the schemata) and, hence, their *a priori* contribution to the possibility of cognition of objects. Therefore, it is natural to expect that the Dialectic would as well carry out the task of articulating the *a priori* forms of representations proper to reason, which count as well among the conditions for the possibility of cognition. It is also obvious that such *a priori* forms of representations, if exist, would be none other than the transcendental ideas, which Kant terms, clearly in allusion to the categories, the pure concepts of reason.

This agenda is conspicuously signaled in the opening sections of the Dialectic, in which reason is introduced as the faculty of "cognition from principles" (A302/B357), the highest degree of (theoretical) cognition. As Kant says, reminding us of where we are in the progression of the *Critique*, "all our cognition starts from the senses, goes from there to the understanding, and ends with reason, beyond which there is nothing higher to be found in us to work on the matter of intuition and bring it under the highest unity of thinking" (A298/B355). This seems to indicate that, like the sensibility, the understanding, and the power of judgment, reason, too, is a source of *a priori* forms of representations, which contribute another, higher level of unity to cognition of objects. However, according to a common reading (which more recently has been resisted by many commentators), in these introductory sections Kant merely *entertain* the possibility of positive contribution of reason to cognition in order to show it, in the three main chapters of the Dialectic, to be false.

In the light of Kant's affirmation of a positive theoretical use of the ideas in the Appendix, any reading according to which reason has no positive contribution to cognition whatsoever is to be rejected. Nonetheless, one could be enticed to endorse a weaker reading according to which the regulative use of the ideas mentioned in the Appendix is different from the real or pure use of reason that Kant anticipated in the introductory sections to the Dialectic. I want to argue that this weaker reading, too, must be rejected. Although the term 'regulative use' does not appear until the Antinomies, Kant's characterization of the legitimate positive use of the ideas in the introductory sections of the Dialectic makes it clear that this positive use of the ideas just is their regulative use.

> The manifold of rules and the unity of principles is a demand of reason, in order to bring the understanding into thoroughgoing connection with itself, just as the understanding brings the manifold of intuition under concepts and through them into connection. Yet such a principle does not prescribe any law to objects and does not contain the ground of the possibility of cognizing and determining them as such in general, but rather is merely a subjective law of economy for the provision of our understanding, so that through comparison of its concepts it may bring their universal

> use to the smallest number, without justifying us in demanding of objects themselves any such unanimity as might make things easier for our understanding or help it extend itself, and so give objective validity to its maxims as well (A305-6/B362-63).
>
> Although we have to say of the transcendental concepts of reason: **They are only ideas**, we will by no means regard them as superfluous and nugatory. For even if no object can be determined through them, they can still, in a fundamental and unnoticed way, serve the understanding as a canon for its extended and self-consistent use, through which it cognizes no more objects than it would cognize through its concepts, yet in this cognition it will be guided better and further. (A329/B385; cf. A321/B377ff.)

In stressing that reason, through its ideas, does not determine objects so as to cognize them, Kant is simply paraphrasing his later claim that the positive use of the ideas is merely regulative and not constitutive. Recognizing the usually neglected connection between these early remarks and the Appendix puts the regulative use of the ideas into perspective. It makes it clear that Kant's view on the regulative use is not, as it were, an afterthought (a mere "appendix"), which occurs to Kant after having already rebutted the possibility of all cognitively significant use of the ideas. On the contrary, the Appendix is more plausibly read as a further explication of the positive role of reason, adumbrated from the beginning of the Dialectic, in supplying cognition with higher forms of unity through its pure concepts. In this sense, the Transcendental Dialectic as a whole has a positive task, which is completely in keeping with what has been pursued in the preceding main chapters of the *Critique*.

This proposed reading also helps to elucidate Kant's point in forging an analogy between the transcendental ideas and the categories, which is that both supply *a priori* forms to cognition (though of different degrees, as we will see). As already pointed out, Kant alludes to the categories or "the pure concepts of understanding" in naming the transcendental ideas "the pure concepts of reason". These denominations imply that each set of concepts originates *a priori*, as a complete system, from the nature

of each faculty.[17] Correspondingly, Kant remarks that in order to derive a complete "table" of the pure concepts of reason, the Dialectic follows the example of the Transcendental Analytic of "how the mere logical form of our cognition can contain the origin of pure concepts *a priori*, which represent objects prior to all experience, or rather which indicate the synthetic unity that alone makes possible an empirical cognition of objects" (A321/B377-78). In the case of reason, the "mere logical form of our cognition" corresponds to the forms of syllogism, which represent the acts of reason in its logical use (cf. A303/B359). By applying the forms of syllogism to "the synthetic unity of intuitions under the authority of the categories", we can somehow derive all the pure concepts of reason, which exhaustively represent the acts of reason in its real use.

The analogy between the transcendental ideas and the categories holds not only with regard to their origin but also extends to their cognitive function. At the opening of Book 1 of the Dialectic, Kant suggests that the pure concepts of reason play an analogous cognitive role to that of the categories, though they concern a different (higher) degree of cognition, which Kant calls comprehension.

> "A concept of reason" […] deals with a cognition (perhaps the whole of possible experience or its empirical synthesis) of which the empirical is only one part; no actual experience is fully sufficient for it, but every experience belongs to it. Concepts of reason serve for **comprehension** [Begreifen], just as concepts of the understanding serve for **understanding** [Verstehen] (of perceptions). (A310-11/B367)

I will say more on the difference between comprehension and understanding below. For now, note that Kant also fleshes out the same parallel in terms of the contrast between two distinct kinds of unity of cognition: "the unity of understanding" and "the unity of reason".[18]

[17] A311/B368, A321/B378, A326/B383.

[18] A302/B359, A307/B363, A326/B383, A422/B450, A645/B673.

> If the understanding may be a faculty of unity of appearances by means of rules, then reason is the faculty of the unity of the rules of understanding under principles. Thus it never applies directly to experience or to any object, but instead applies to the understanding, in order to give unity *a priori* through concepts to the understanding's manifold cognitions, which may be called "the unity of reason," and is of an altogether different kind than any unity that can be achieved by the understanding. (A302/B359)
>
> Pure reason leaves to the understanding everything that relates directly to objects of intuition or rather to their synthesis in imagination. It reserves for itself only the absolute totality in the use of concepts, and seeks to carry the synthetic unity, which is thought in the categories, all the way to the absolutely unconditioned. We can therefore call this the **unity of reason** in appearances, just as that which the category expresses can be called the **unity of understanding**. (A326/B382-83)

According to these passages, we can distinguish two kinds of unity that our cognition may exhibit. On the one hand, the unity of understanding is the unity of appearances under rules constitutive of experience of objects. On the other hand, the unity of reason is the higher-order unity of those rules (i.e., concepts) under principles. The unity of understanding is "thought in the categories" and characteristic of the degree of cognition called understanding. The unity of reason is thought in the transcendental ideas and characteristic of the degree of cognition called comprehension. Crucially, unlike the unity of understanding, the unity of reason does not constitute a condition for the possibility of experience (hence, "it never applies directly to experience or to any object"). Because the possibility of experience does not depend on the unity of reason and the transcendental ideas, the theoretical use of the ideas is never constitutive but merely regulative.[19]

Now, in the *Jäsche Logik* and the logic lecture transcripts, the terms 'understanding' and 'comprehension' are explicated in terms of the "degrees" [Grade] of cognition, which differ "in regard to the objective content" (*JL* 9:64). These are (from the lowest to the highest): (i) To represent

[19] Cf. A664/B692.

something [vorstellen], (ii) to represent something with consciousness or to perceive it [wahrnehmen, *percipere*], (iii) to be acquainted with something [kennen, *noscere*], (iv) to be acquainted with something with consciousness or to cognize it [erkennen, *cognoscere*], (v) to understand something [verstehen, *intellegere*], (vi) to have insight into something [einsehen, *perspicere*], (vii) to comprehend something [begreifen, *comprehendere*]. What interest us here are 'understanding' and 'comprehension', which are the fifth and the seventh (highest) degree in the ladder of cognition. To understand something "to cognize something through the understanding by means of concepts, or to conceive [concipiren]". In contrast, to comprehend something is "to cognize something through reason or *a priori* that is sufficient for our purpose" (*JL* 9:65).[20] Mediating these two degrees of cognition is 'insight'—"to have insight [einsehen] into [something] (*perspicere*)" is to cognize it simply through reason. Thus, insight and comprehension appear to be two different degrees of rational cognition, i.e., cognition from grounds or principles (see Section 1.2).

Comprehension, as opposed to mere understanding, involves not only forming a judgement which represents an object as falling under a concept but also grasping the *ground* for which the object falls under that concept. In other words, to comprehend *a*, one must not only judge *that a* is *F* but also understand *why a* is *F*. This is implied by Kant's characterization of comprehension as cognition "*a priori*". In this context, Kant is using 'cognition *a priori*' not in the usual sense of 'cognition independent of experience' but in the traditional Scholastic (and more basic) sense of 'cognition from ground'.[21] To cognize something *a priori* in the traditional sense is to cognize it as a consequence or effect of an

[20] Cf. *LB* 24:134-6, *LDW* 24:730-31, *WL* 24:845-6.

[21] Cf. *MM* 29:748. This notion of cognition *a priori* is also couched in the term 'cognition from principles' (e.g., A399/B357). For a discussion of the role of the traditional notion of apriority in Kant, see Houston Smit, "Kant on Apriority and the Spontaneity of Cognition," in *Metaphysics and the Good: Themes From the Philosophy of Robert Merrihew Adams*, ed. Samuel Newlands and Larry Jorgensen (Oxford: Oxford University Press, 2009), 188–251.

explanatory ground—either a ground of being [*ratio essendi*] or of becoming [*ratio fiendi*].[22] Both degrees of rational cognition, insight and comprehension, count as cognition from ground. As Kant clarifies in some of the logic lecture transcripts, the difference between them is one in degree. To have insight into a thing, which we have previously cognized (or understood) through its marks, is to grasp an *insufficient* ground for one of its marks. For example, "to have insight into what gold is I must investigate one of its marks in particular and abstract from it its ground. E.g., why it does not rust, why is it ductile, heavier than others" (*LB* 24:135). By contrast, to comprehend a thing is to have insight into it *sufficiently* and, hence, to grasp a sufficient ground for one of its marks.[23]

This account of rational cognition in general and comprehension in particular reveals the connection between Kant's claim that reason is the faculty of cognition from principles (i.e., explanatory grounds) and, hence, of comprehension and his claim that reason by its nature seeks the unconditioned.

> For a given conditioned reason demands an absolute totality on the side of the conditions (under which the understanding subjects all appearances to synthetic unity), thereby making the category into a transcendental idea, in order to give absolute completeness to the empirical synthesis through its progress toward the unconditioned (which is never met with in experience, but only in the idea). Reason demands this in accordance with the principle: **If the conditioned is given, then the whole sum of conditions, and hence the absolutely unconditioned, is also given**, through which alone the conditioned was possible. (B436)

In short, reason demands not only the conditions for whatever is conditioned but also the totality of its conditions and, hence, the unconditioned. The conditioned is something that stands in need of

[22] Cf. 1:391-8. On Kant's notion of *ratio* (reason, ground), see Béatrice Longuenesse, "Kant's Deconstruction of the Principle of Sufficient Reason," in *Kant on the Human Standpoint* (Cambridge: Cambridge University Press, 2005), 117–42..

[23] Cf. *LB* 24:135, *LDW* 24:730-31, *WL* 24:847. As Kant repeatedly emphasizes, we can never have insight into an *absolutely* sufficient ground of any given thing but only a ground that is "sufficient for our purpose".

explanation. What explains it is called its condition. The unconditioned is what admits of no further explanation and, therefore, what is explanatory basic. Thus, Kant characterizes reason as the cognitive faculty that seeks not only an explanation for whatever calls for explanation but also a *complete* explanation for it, an explanation that specifies all the explanatory grounds (i.e., the totality of conditions) for the explanandum and connects them to that which is explanatorily basic (i.e., the unconditioned). It should be obvious that to possess the complete explanation of a certain feature of an object in virtue of which it is conditioned (that is, to cognize the totality of conditions and the unconditioned with respect to that feature) would amount to cognizing that feature of the object absolutely *a priori* (in the traditional sense) or cognizing it from its absolutely sufficient ground and, hence, to *comprehend* it absolutely. Therefore, in seeking a complete explanation of anything that stands in need of explanation, reason seeks absolute comprehension of any aspect of reality that is not explanatory basic.

As Kant points out in the passage just quoted, reason's demand for the totality of conditions and the unconditioned is what underlies the genesis of the transcendental ideas as concepts of unconditioned objects. In making this demand, reason "assumes" [annehmen] that "when the conditioned is given, then so is the whole series of conditions subordinated one to the other, which is itself unconditioned, also given (i.e., contained in the object and its connection)" (A307-8/B364-65; cf. A409/B436). Kant calls this assumption "the supreme principle of pure reason". Note that this principle has the character of a "transcendental" principle, that is, an assertation about the way things are. It asserts that the unconditioned is "given" in the sense of being "contained in the object and its connection". In this way, reason, like all the other higher faculties of the mind, becomes a source of apparent synthetic *a priori* judgment about the object, which awaits verification by the critique. However, Kant seems to think that reason not only asserts the existence of the unconditioned in the abstract but also proffers a rather specific conception of what unconditioned objects there are (the

soul, the world-whole, and God). Otherwise, reason's commitment to the supreme principle would not give rise to three different transcendental ideas (and specifically these three). As I will argue in the next section, we can plausibly attribute to Kant the view that objects are conditioned in three broad respects. This view together with the supreme principle of reason are sufficient to explain why reason assumes not only that some unconditioned object or other exists but that the soul, the world-whole, and God exist.

Now, the possibility of absolute comprehension of an aspect of reality that is conditioned depends on things being a certain way, namely, that they "contain" the unconditioned condition corresponding to that aspect of reality. The cognition of this unconditioned condition would serve as the ultimate ground or principle from which to derive the cognition of that aspect of reality as its consequence. Thus, the applicability of the concepts of the unconditioned to objects is a condition for the possibility of absolute comprehension for beings that possess reason. This parallels the way in which the applicability of the categories to objects is a condition for the possibility of experience for beings with a discursive understanding. It is clear that for Kant neither cognition of unconditioned objects nor absolute comprehension is possible *for us*.[24] The reason for this is that our cognition is limited to objects that can be given in empirical intuition, which is limited by the conditions of our forms of sensibility.[25] This precludes the possibility of producing a transcendental deduction of the transcendental ideas of the same sort as the transcendental deduction of the categories.

Nonetheless, the impossibility of absolute comprehension for rational beings with sensible intuition does not exclude its possibility for an intellect with a non-sensible form of intuition. Further, what needs to be demonstrated for the deduction of the ideas to go through is not necessarily the

[24] For a helpful discussion of this claim, see Eric Watkins, "Kant and the Grounding of Transcendental Idealism," *Studi Kantiani* 32 (2019): 103–18.

[25] See Section 3.5 for further discussion.

possibility of absolute comprehension *for us*. If it were possible to establish with certainty that absolute comprehension of things is possible for *some* rational beings (perhaps with non-sensible kind of intuition), this would be sufficient for demonstrating the *objective* validity of the synthetic *a priori* principle of reason and the transcendental ideas. However, it is relatively obvious that Kant thinks we cannot *know* whether absolute comprehension is possible for a rational being with a different kind of intuition—according to him, we are not entitled to even assume that another kind of intuition than our own is really possible.[26] Yet, Kant seems to think that the fact that *our* reason demands absolute comprehension gives us at least a *subjectively* sufficient ground for assuming that absolute comprehension is possible for another rational being whose intellect is not limited by the conditions of sensible intuition.[27]

For this reason, although we cannot give a transcendental deduction of the ideas of the same kind as that of the categories, a different, weaker kind of transcendental deduction is nonetheless possible (A669/B697, A671/B699, cf. A336/B393). According to Kant, this deduction demonstrates not only the subjective validity[28] but also the *indeterminate* objective validity[29] of the pure principles and concepts of reason. On my reading, this deduction proceeds from the subjectively grounded claim that absolute comprehension is possible for a rational being whose intellect is not limited by the conditions of sensible intuition to the subjective or indeterminately objective validity of the transcendental ideas as its condition. In this way, the transcendental deduction of the ideas has the same structure of a "regressive" argument as the transcendental deduction of the categories, which proceeds from the possibility of experience to the "determinate" objective validity of the categories

[26] Cf. A254/B309ff.

[27] On subjectively sufficient grounds of assertion, see A820/B848ff.

[28] E.g., A297/B353, A306/B362-63, A666/B694, A671/B699, A680/B708.

[29] A663/B691, A669/B697, A680/B708, 18:224.

as its condition.[30] The difference between the two deductions, as I have pointed out, lies in the strength of the warrant with which we can assert the possibility of the respective kinds of cognition.

I want to suggest that we are subjectively warranted to assert the possibility of absolute comprehension for rational beings whose intellect is not limited by the conditions of sensibility because, unlike in the cases of our sensibility and understanding, we are entitled to take *our* faculty of reason to be representative of *any* faculty of reason in general. In other words, we are entitled to assume that the principles of our reason express the principles of reason considered *as such*. Therefore, the fact that our reason demands absolute comprehension is a reason for asserting that *were* our intellect not limited to the objects of sensible intuition, absolute comprehension *would* be possible for us and, hence, that for any rational being whose intellect is not conditioned by sensibility, absolute comprehension is possible. Although I cannot offer any direct textual support here, this suggestion is not implausible given that Kant seems to rely upon an analogous claim in his practical philosophy. With regard to the practical use of reason, Kant seems to think that we can take our reason to be representative of any reason in general such that we can take the principle governing our own reason (i.e., the moral law) to hold universally for any being with reason.[31] In the practical case, our reason makes laws not for itself alone but also for every rational being. One reason to expect that some version of this claim applies (with weaker warrant) to theoretical reason is that, according to Kant, reason is a unity: "there can, in the end, be only one and the same reason, which must be distinguished merely in its application" (*GMS* 4:391).[32]

---

[30] For a reconstruction of the transcendental deduction of the categories as a regressive argument, see Karl Ameriks, "Kant's Transcendental Deduction as a Regressive Argument," *Kant Studien* 69, no. 1–4 (1978): 273–87.

[31] Cf. *GMS* 4:413-14, 4:449, *KpV* 5:82.

[32] For textual evident that suggests that theoretical reason must as well be seen as legislative, see A653/B681 and A694-95/B722-23.

The positive theoretical use of the transcendental ideas, then, derives its legitimacy from the subjectively warranted assumption that absolute comprehension is possible for an intellect that is not limited by the conditions of sensible intuition. This assumption serves as the basis for a transcendental deduction that follows the same outline as the transcendental deduction of the categories. The difference is that, unlike the possibility of experience, the possibility of absolute comprehension is assumed only on a subjectively sufficient ground.

## 4.5 The Ideas of Reason as Concepts of Synthesis

By taking Kant's analogy between the transcendental ideas and the categories seriously, we have arrived at an alternative understanding of the positive theoretical use of the transcendental ideas as forms of comprehension and, hence, conditions of its possibility. This parallels the way in which the categories are forms of empirical judgment and, hence, conditions of the possibility of experience. This reading draws textual support mainly from the early sections of the Transcendental Dialectic. As we have seen, these passages give a very different picture of the role of the ideas than the standard account, which often focuses narrowly on the Appendix. However, I have argued that there is no reason to suppose that the "regulative" use of the ideas discussed in the Appendix is distinct from the positive use of the ideas that Kant adumbrates throughout the early sections of the Dialectic. But to show that it is plausible to think of the transcendental ideas as forms of a certain degree of cognition is not yet show *how* they can fulfil this role.

In the case of the transcendental ideas, what seems puzzling is that they are concepts of individual things. This makes it difficult to see how these concepts can play the role of forms of a certain kind of representation. Although it is not always completely clear either how the categories are supposed to be forms of empirical thinking or judgment (at least with respect to individual categories), the fact that they are concepts of generic ontological features of objects at least gives intuitive

plausibility to the idea that they provide some sort of structuring unity to our thinking about empirical objects. For example, it is intuitively plausible to think of the concept of the substance-inherence relation as an *a priori* form or template of thought that can be filled in with a variety of empirical content, i.e., empirical concepts of things and their predicates. But can we think of the concepts of the soul, the world, and God in the same vein?

The first thing to note is that Kant characterizes the transcendental ideas not only as concepts of unconditioned objects but also as concepts of the totality in the *synthesis* of conditions for a given conditioned thing.[33] Kant explains that the concept of the unconditioned condition for a given conditioned thing and the concept of the totality in the synthesis of its conditions are two sides of the same coin: "Since the **unconditioned** alone makes possible the totality of conditions, and conversely the totality of conditions is always itself unconditioned, a pure concept of reason in general can be explained a through the concept of the unconditioned, insofar as it contains a ground of synthesis for what is conditioned" (A322/B379). This highlights another important point of similarity between the categories and the ideas: they are both concepts of synthesis, i.e., rules according to which the understanding synthesizes a manifold of representation. Recall that by synthesis in the "most general sense" Kant understands "the action of putting different representations together with each other and comprehending their manifoldness in one cognition" (A77/B103). Conceiving of the ideas as concepts or rules of synthesis seems consistent with Kant's characterization of reason as "the faculty of the unity of the rules of understanding under principles", which gives "unity *a priori* through concepts to the understanding's manifold cognitions" (A302/B359). This suggests a way to flesh out the claim that the transcendental ideas can be construed, in analogy with the categories, as forms of representation.

[33] Cf. A322/B379f., A336/B393, A396-97.

The connection between the ideas and the categories goes even further. Kant explicitly indicates that the synthesis represented in the ideas is genetically related to the synthesis represented in the categories. As we have seen, he holds that reason generates the ideas by carrying "the synthetic unity, which is thought in the categories, all the way to the absolutely unconditioned" (A326/B383). Note that, for some reason, Kant thinks that not all the categories are relevant for the synthesis of conditions but only the relational categories.[34] Therefore,

> There will be as many concepts of reason as there are species of relation represented by the understanding by means of the categories; and so we must seek an **unconditioned**, **first**, for the **categorical** synthesis in a **subject**, **second** for the **hypothetical** synthesis of the members of a **series**, and **third** for the **disjunctive** synthesis of the parts in a **system**. (A323/B379)

Thus, the same syntheses represented by the relational categories represent are represented in their *totality* by the transcendental ideas. But what could it mean for the categorial synthesis to be represented in totality in the ideas?

Note, first, that for Kant's repeated claim that reason makes distinctive contribution to cognition and that the unity of reason is a unity "of which the understanding has no concept" (A326/B383) to make sense, the following must be the case. First, the fact that the categorial syntheses are thought in the ideas in their totality must completely change the *qualitative* character of those syntheses. Otherwise, we could not make sense of Kant's claim that the unity of reason (characteristic of cognition synthesized according to the ideas) differs in *kind* from the unity of understanding (cf. A302/B359). Thus, the notions of synthesis of conditions and its totality must signify more than the simply quantitative difference that can be exemplified by, say, the difference between a slice of pie and

[34] The fact that the other three classes of categories (quantitative, qualitative, modal) come into play later in the explication of each of the transcendental ideas in the Paralogisms, the Antinomies, and the Ideal contributes to further confusion.

a full pie. The fact that a pie is given in its totality does not (normally) change the fact about what kind of pie it is (e.g., a pecan pie). Second, although the relational categories may provide *necessary* basis for deriving the transcendental ideas, this basis cannot be *sufficient*. In other words, it cannot be the case that the kind of synthesis represented in the ideas is already contained in the synthesis of the categories or can be obtained from the latter simply by extending it. This, again, suggests that the relevant notion of totality, which designates reason's distinctive contribution, cannot be understood as a merely quantitative notion.

These considerations seem to entail that the notion of totality that is here relevant is *not* quantitative totality, i.e., the totality represented by the quantitative category of allness or totality, which "is nothing other than plurality considered as a unity" (B111). This unity is the unity in the generation of a homogeneous magnitude in intuition (cf. B115, B162). In contrast, the totality in the synthesis which the ideas represent must be what Kant calls 'qualitative completeness (totality)', which concerns "the quality of a cognition for the connection of **heterogeneous** elements of cognition into one consciousness" (B115). Therefore, the relevant notion of totality is not to be explained in terms of the quantitative category of totality, which is a form of unity that *arises* from the combination of a manifold (cf. B131). Rather, qualitative totality belongs together with what Kant calls qualitative unity, which *grounds* the combination of a manifold, being the unity of the *action* of the understanding which underlies all combination. This is what Kant describes as "the unity of the comprehension [Zusammenfassung] of the manifold of cognition […], as, say, the unity of the theme in a play, a speech, or a fable" (B114), which he identifies with the synthetic unity of apperception. The claim that the ideas represent the categorial synthesis as qualitative totality, if correct, implies the following. Although the transcendental ideas presuppose the forms of synthesis according to the relational categories and, hence, forms of synthesis grounded in the synthetic unity of apperception, they *transform* these forms of synthesis into ones that *cannot* be grounded in or explained by the synthetic

unity of apperception. In this way, reason, through the transcendental ideas, prescribes forms of synthesis to the understanding that would have to be grounded in another kind of rudimentary unity of consciousness than synthetic unity of apperception and, consequently, could be fully realized by our understanding only if it were not discursive.

The difference between the two kinds of unity of consciousness seems to be what Kant tries to capture by his murky distinction between *collective* and *distributive* unity.[35] It is unfortunate that Kant offers very little clarification of this distinction despite its apparent significance. Nevertheless, it is worth briefly considering this distinction because of its obvious relevance for the present discussion. According to Kant, reason generates the transcendental ideas by transforming "the **distributive** unity of the use of the understanding in experience, into the **collective** unity of a whole of experience" (A582/B660).[36] Correspondingly, we employ the ideas to posit "a certain collective unity as the goal of the understanding's actions, which are otherwise concerned only with distributive unity" (A644/B672). What distinguishes one kind of unity from the other appears to be the direction of the priority relation between the whole and the parts. Distributive unity is characteristic of an *aggregate*, a whole in which the parts precede and make possible the whole. Collective unity, by contrast, is what characterizes a proper *whole* [das Ganze], a whole that precedes and makes possible its parts. In the latter case, if a whole is given, all its parts, as limitations of that whole, are also given along with it.[37] The difference between distributive and collective unity characterizes the essential difference between the forms of synthesis grounded in the categories and the forms of synthesis grounded in the ideas and, correspondingly, between the unity of understanding and the unity of reason.

---

[35] A582/B660, A644/B672. For a classic discussion of this distinction, see Michael Friedman, *Kant and the Exact Sciences* (Cambridge: Harvard University Press, 1992), 300–311. Also helpful is Anderson, *The Poverty of Conceptual Truth*, 282–85.

[36] Cf. A323/B380.

[37] The formal intuitions of space and time are paradigmatic examples of collective wholes.

One way to understand what it could mean for the unity of reason, supposedly grounded in the transcendental ideas, to be a collective unity is this. For Kant, as we have seen, the unconditioned represented by an idea is what *grounds* the totality of conditions for a given conditioned object.[38] Hence, the transcendental ideas are concepts not merely of an unconditioned object but specifically of an unconditioned object such that if it is given the totality of conditions is also given along with it. Consequently, the cognition of such an unconditioned object would be able to serves as a principle from which can be derived the cognition of the conditioned object along with the cognition of all of its conditions. As I have argued, absolute comprehension involves just such cognition of an unconditioned object as an ultimate explanatory ground of a conditioned object. Therefore, there is a clear sense in which any body of cognition that qualifies as absolute comprehension forms a whole that precedes and makes possible its parts, since all its parts can be derived from a single principle.

This point illuminates the link between the notion of comprehension (and rational cognition more generally) and the notion of systematic unity of cognition, which is the central idea in the Appendix. The notion of systematic unity of cognition appears to refer to exactly the kind of collective unity characteristic of comprehension, which strengthens my claim that the regulative role of the transcendental ideas just is their role as forms of comprehension. Consider this central passage from the Appendix.

> What reason quite uniquely prescribes and seeks to bring about […] is the **systematic** in cognition, i.e., its interconnection based on one principle. This unity of reason 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. Accordingly, this idea postulates complete unity of the understanding's cognition, through which

[38] Cf. A322/B379.

> this cognition comes to be not merely a contingent aggregate but a system interconnected in accordance with necessary laws. (A645/B673)

This passage basically enumerates all the key features of rational cognition in general (see Section 1.2) and of comprehension in particular. Kant contrasts the unity of a system with the unity of a "contingent aggregate", which implies that systematic unity of cognition is a collective unity. Cognition possesses this unity in virtue of "its interconnection based on one principle", which maps onto the notion that comprehension consists in cognition from a single principle. Further, this unity presupposes an *idea* of "the form of a whole of cognition, which precedes the determinate cognition of the parts". This idea can plausibly be identified with one of the transcendental ideas. As I have argued, the transcendental ideas are concepts of unconditioned objects whose cognitions could serve of as principles from which all the parts of the whole of cognition could be derived. Hence, systematic unity of cognition, considered in its completion, would be the same kind of unity that characterizes comprehension.

The connection between systematic unity and comprehension is also explicitly drawn in the following passage from the *Prolegomena*.

> Nothing can be more desirable to a philosopher than to be able to derive, *a priori* from one principle, the multiplicity of concepts or basic principles that previously had exhibited themselves to him piecemeal in the use he had made of them *in concreto*, and in this way to be able to unite them all in one cognition. Previously, he believed simply that what was left to him after a certain abstraction, and that appeared, through mutual comparison, to form a distinct kind of cognitions, had been completely assembled: but this was only an *aggregate*; now he knows that only precisely so many, not more, not fewer, can constitute this kind of cognition, and he has understood the necessity of his division: this is a **comprehension** [ein Begreifen] and only now does he have a *system*. (*Prol.* 4:322, boldface added)

This passage not only makes clear that systematic unity (as opposed to unity of an aggregate) is a characteristic feature of comprehension but also point out that the pursuit of comprehension

constitutes the goal of philosophy (obviously to be understood as Critical philosophy). More specifically, what philosophy seeks to unite in one system and, hence, to comprehend is concepts and principles which we ordinarily apply *in concreto*.

## 4.6 Forms of Explanatory Understanding

I have tried to show how we might plausibly view the transcendental ideas, in analogy with the categories, as forms of cognition even though they are concepts of individual objects. The basic idea is that the transcendental ideas, like the categories, are concepts of synthesis or rules for combining a manifold of representations. The difference is that whereas the categorial synthesis grounds the unity of experience, what we might call rational synthesis grounds the unity of comprehension, which our cognition can only approximate but never fully realized. The transcendental ideas fulfill this role by representing unconditioned objects that ground the totality in the synthesis of conditions for conditioned objects. Hence, the object of an idea is one whose cognition *per impossibile* could serve as a principle from which alone we could derive the cognition of a given conditioned object together with the cognitions of all its conditions and, thus, absolute comprehension of the conditioned object. In this way, reason prescribes to the understanding a form of synthesis characteristic of comprehension, which Kant also refers to as systematic unity of cognition, by postulating the unconditioned condition for a given conditioned object as a "goal" or "*focus imaginarius*" for the use of the understanding (A644/B672). The understanding shall combine its cognitions *as if* they were to eventually constitute absolute comprehension, which is tantamount to combining them *as if* the totality of conditions and the unconditioned were "contained in the object and its connection". Therefore, reason demand that the understanding combine its representations *as if* the supreme principle of pure reason were true.

Kant thinks that the only way reason can prescribe this form of synthesis or unity of cognition to the understanding is by postulating the unconditioned objects that would ground the totality of synthesis. As he puts it, "reason cannot think […] systematic unity in any other way than by giving its idea an object" (A681/B709). The reason why he thinks this is not obvious presumably has to do with the discursivity of our intellect. Since we do not possess the unity of consciousness necessary for fully realizing the unity of cognition reason prescribes, it is plausible that our grasp of the rules of synthesis that would bring about this unity of cognition would be limited in ways in which our grasp of the rules of synthesis contained in the categories is not.

However, reason does not prescribe its unity to the understanding simply through a single, generic concept of an unconditioned object but rather through the concepts of three specific unconditioned objects. Following the analogy with the categories, this seems to imply that comprehension requires three forms of synthesis just as experience requires twelve forms of synthesis, represented by the twelve categories. But how should we understand the idea that comprehension requires three distinct forms of synthesis? What I aim to do in this section is to sketch a preliminary answer to this question.

According to Kant, conditioning relations, relations between a conditioned object and its condition, come in different kinds. Hence, there are as many respects in which objects can be conditioned as there are conditioning relations they can enter into. Further, an object can be conditioned in one respect (i.e., with regard to one kind of conditioning relations) while being unconditioned in another respect (i.e., with regard to another kind of conditioning relations). We can characterize a conditioning relation in general, roughly, as an asymmetric, transitive, and explanatory relation of metaphysical dependence between objects (in a broad sense that covers a variety of

entities—substances, accidents, events, representations, possibilities, etc.).[39] Thus, an object is conditioned by another object just in case it metaphysically depends (in some respect) on and is explained by the other object but not vice versa. In the Transcendental Dialectic, Kant introduces many kinds of conditioning relations. The Antinomies alone contains discussion of four different kinds of conditioning relations. To mention just two examples: the Second Antinomy focuses on the relation of *composition*, which holds between a composite whole (a conditioned object) and its parts (its conditions). This relation qualifies as a conditioning relation because a whole depends on and is explained by its parts, but not vice versa. The topic of the Third Antinomy is the relation of *causation*, which holds between an event (a conditioned object) and its cause (its condition). A causal relation is a conditioning relation for a similar reason: an event depends on and is explained by its cause, but not vice versa.

The fact that there are three transcendental ideas and, hence, three concepts of the unconditioned condition suggests that all the conditioning relations of the Dialectic can be grouped into three broad kinds or families, each of which is treated in each main chapter of the Dialectic. Correspondingly, there are three broad respects in which an object can be conditioned. As already pointed out, one major flaw of the standard reading, which rests on the domain specificity thesis, is that it is difficult to see what intelligible correlation can exist between the three broad kinds of conditioning relations and the three domains of natural science. However, if we reject the domain specificity thesis, the burden is on us to show that we can plausibly think of any object as, in principle, capable of be conditioned in three different respects and, more specifically, in exactly the same respects in which the objects of the transcendental ideas are unconditioned.

[39] I follow the account of real conditioning relations offered by Eric Watkins, "Kant on Real Conditions," in *Natur Und Freiheit. Akten Des XII. Internationalen Kant-Kongresses*, ed. Violetta Waibel, Margit Ruffing, and David Wagner (Berlin/Boston: De Gruyter, 2018), 1133–40.

I want to argue that this can be done if by an object we understand an object of experience or an appearance. As I pointed earlier, Kant sometimes says that the transcendental ideas are derived from the "conditions of all representations in general" (A405/B432; cf. A333-34/B390-91) or, what I take to be the same, the "universal conditions of thinking" (A397), which come in three classes. These passages vary somewhat in the way they characterize these three classes of conditions:

A. "the relation [of representations] to the subject" (A334/B391); "the conditions of a thought in general" (A397); "the **subjective** conditions of all representations in general (of the subject or the soul)" (A406/B432);
B. "the relation [of representations] to objects as appearances / the manifold of the object in appearance" (A333-34/B390-91); "conditions of appearance" (A334/B391); "the conditions of empirical thinking" (A397); "objective conditions in appearance" (A406/B433);
C. "the relation [of representations] to objects of thinking in general / all things in general" (A333-34/B390-91); "the condition of all objects of thought in general" (A334/B391); "the conditions of pure thinking" (A397); "objective conditions of the possibility of objects in general" (A406/B433).

Despite the slight inconsistency of the terms Kant uses to describe these conditions, the identity of each class of conditions across these passages is confirmed by the fact that in all of them Kant associates what appears to be the same class of conditions with the same transcendental idea: A – the soul; B – the world; C – God. Each of the transcendental ideas represents the "absolute unity" (A334/B391) or the "absolute totality of the synthesis" (A397) or the "unconditioned unity" (A405-6/B432-33) of the respective class of conditions.

Now, if we suppose that these conditions apply to an appearance or an object of experience, we can plausibly think of their division into three classes as representing three ways in which an object can be conditioned. As I have argued in Section 1.2, the critique as an explanatory project presupposes transcendental idealism, according to which objects of experience are mind-dependent representations. Thus, from the standpoint of Critical investigation, objects have a dual aspect: *qua*

transcendentally ideal, they are mind-dependent representations; *qua* empirically real, they are really existing objects in space and time.[40] Note that this dual aspect also allows us to view the objects as subject to two distinct modes of explanation. As representations, their character can be explained by appeal to the faculties of our mind as the source or the ground of representation (what we do in Critical philosophy). As objects, their character can be explained by appeal to their relations to other objects (what we do in natural science).

The dual aspect of an object of experience is, I think, what underlies the distinction between the first class of conditions (the subjective conditions - A) from the other two class (the objective conditions – B and C). Here we follow Kant in dividing the conditions of representation first into two classes, the latter of which can then be divided into two subclasses: "what is universal in every relation that our representations can have is 1) the relation to the subject, 2) the relation to objects, and indeed either as appearances, or as objects of thinking in general" (A333-34/B390-91). If this is correct, then the first class of conditions, from which the idea of the soul is derived, is concerned with the conditioning relation (or the family of conditioning relations) that holds between representations (as conditioned objects) and the representing subject (as their condition)—the relation in which representations depend on and are explained by the nature of the subject, but not vice versa. This can plausibly be identified with the conditioning relation that Kant discusses in the Paralogisms, but also with the conditioning relation that the critique appeals to in explaining the *a priori* features of objects. It should also be noted that if, as I argued, we consider objects to be conditioned in this way only from the "transcendental" standpoint, it follows that the relevant relation between the subject and its representations cannot be an empirical causal relation, the sort of relation presumably investigated in empirical psychology.

[40] Cf. A38/B55.

With respect to the other two classes of conditions, by contrast, we consider objects from the "empirical" standpoint, *qua* real objects in space and time. But how should we understand the suggestion that objects *qua* empirically real can be conditioned in two different respects? The first case (B) is relatively easy. The fact that Kant characterizes this class of conditions as pertain to "appearances" strongly suggests that these conditions correspond to the conditioning relations that hold between two appearances or, as Kant sometimes puts it, the conditioning relations that hold within a "series of appearances" (cf. A414/B441, A446/B474, A456/B484). One can recognize all four conditioning relations discussed in the Antinomies as belonging under this class of conditioning relations. These relations give rise to a set of transcendental ideas, which appear to represent different aspects of the idea of the world-whole. Kant calls these ideas 'world-concepts' [Weltbegriffe]. As Kant points out, what is common among the world-concepts is that they concern the conditioning relations among appearances: "I call all transcendental ideas, insofar as they concern absolute totality in the synthesis of appearances, **world-concepts** […] because they have to do merely with the synthesis of appearances, and hence with the empirical […]" (A407-8/B434, cf. A420/B447). Insofar as objects are considered as appearances and, hence, as conditioned with respect to the "cosmological" conditioning relations, they also admit of a mode of explanation different than one we have previously considered. I take it that the mode of explanation corresponding to the cosmological ideas is one in which we explain a contingent aspect of an appearance (e.g., a particular event) by appealing to another appearance (e.g., a previous state) and the relation between them, which is governed by laws of nature.[41]

What, then, is the respect in which objects of experience are conditioned that corresponds to the third and final class of conditions (C)? Here Kant's own label of these conditions as having to do

[41] Cf. A544/B572ff.

with either "things in general" or "pure thinking" is not particularly illuminating. There are two familiar contexts in which both terms simultaneously find their use. In the first context, "pure thinking" refers to our thinking about the objects according to the rules of transcendental logic and, hence, the categories, which are the concept of an "object in general" (e.g., A21/B36, A55/B60, A96). In the second context, it is noumena that are referred to by objects of "pure thinking" (e.g., A287/B343, A601/B629) as well as "things in general" (e.g., A35/B51, A137/B178, A147/B186). However, it is unlikely that either of these two senses of "things in general" reflects what Kant has in mind with regard to the third class of conditions. On the one hand, if "things in general" were to be understood as objects insofar as they stand under the categories, the third class of conditions would seem to overlap with conditions of the first class, which obviously include conditions of representation grounded in our understanding. On the other hand, nothing precludes interpreting "things in general" as referring to things in themselves insofar as they are conditions of appearances; after all, Kant does claim that appearances depend on things in themselves. However, the problem with this reading is that it is not obvious that we have sufficiently determinate grasp of the conditioning relation between appearances and things in themselves such as to have an idea of something that is unconditioned with respect to that relation, especially one that is more determinate than the concept of a thing in itself (which, by definition, is unconditioned in precisely those respects in which an appearance as such is conditioned).

Fortunately, the discussion of the Ideal chapter makes its clear what kind of conditioning relations is Kant's concern here, namely, the conditioning relation that holds between a thing's real possibility (as a conditioned object) and actuality (as its condition).[42] Hence, "things in general" appears to a shorthand for "*possible* things in general". This is confirmed by the fact that in some

[42] Cf. A571/B599ff.

passages Kant refers to the same conditions as "objective conditions of the possibility of objects in general" (A406/B433) or "the condition of all possible things in general" (A408/B434-35).

Prior to the discussion in the Ideal, real possibility is sometimes mentioned as a condition of objective validity of a concept and, hence, of cognition. According to Kant, we cannot cognize an object whose real possibility we cannot prove, although we might be able to form a non-contradictory concept of it.

> To **cognize** an object, it is required that I be able to prove its possibility (whether by the testimony of experience from its actuality or *a priori* through reason). But I can **think** whatever I like, as long as I do not contradict myself, i.e., as long as my concept is a possible thought, even if I cannot give any assurance whether or not there is a corresponding object somewhere within the sum total of all possibilities. But in order to ascribe objective validity to such a concept (real possibility, for the first sort of possibility was merely logical) something more is required. (Bxxvin)

Thus, when we consider an object as *really* possible (as opposed to merely logically possible), we consider it not only as an object corresponding to a concept whose predicates are logically consistent but also as an object corresponding to a concept whose predicates are drawn from the sum total of all possible predicates. As Kant points out in this passage, the real possibility of the object represented by a concept is a necessary condition for the concept's objective validity and, hence, for its relation to an object. Therefore, it is also a necessary condition of our thought having a relation to an object or object-directed content. Presumably for this reason, Kant calls the conditions of real possibility also "the condition of all objects of thought in general" (A334/B391), the "condition of the possibility of everything that can be thought" (A334/B391), and "the conditions of pure thinking" (A397).

In this way, when we consider an object of experience as a real possibility and, hence, as conditioned with respect to the third class of conditions, we consider it as a *thinkable* object or an object to which we can relate in thought.

> The possibility of objects of sense is a relation of these objects to our thought, in which something (namely, the empirical form) can be thought *a priori*, but what constitutes the material, the reality in appearance (corresponding to sensation) has to be given; without that nothing at all could be thought and hence no possibility could be represented. (A581/B609)

The crucial point is that relation of our thought to an object has both a "formal" condition, i.e., non-contradictoriness, and a "material" condition, i.e., the givenness of its material or content from without. The content of our thought, however, can only be given through experience and, hence, through an actually existing object. More concretely, predicates with which we represent a real possibility must be drawn from an actually existing object that instantiates them and is given to us in experience. In this way, real possibility depends on actuality. Hence, in essence, the conditioning relation that is the topic of the Ideal is the one that holds between a really possible object *qua* an "object of thought" (as a conditioned object) to the ground or the source of the content of the thought or the concept through which it is represented (as its condition). Accordingly, the idea of God as an *ens realissimum* (a being that instantiates all positive predicates to the highest degree) is the idea of "the supreme and complete material condition of its possibility, to which all thinking of objects in general must, as regards the content of that thinking, be traced back" (A576/B604).

I have suggested that objects of experience can be seen as conditioned in three broad respects, corresponding to three perspectives in which they can be considered—as representations, as appearances, and as objects of thought. Accordingly, conditioning relations can be sorted into three broad types or families. Consequently, there are three ways in which objects of experience can be *comprehended* in the Kantian sense and, thus, three kinds of comprehension or explanatory understanding we can, in principle, achieve with regard to these objects. This explains why reason prescribes its form of synthesis to the understanding through three specific ideas. By means of these ideas, reason postulates objects that are unconditioned precisely in those three respects and, hence,

objects whose cognitions *per impossibile* would serve as principles from which alone all three kinds of comprehension could be completely derived. In this way, reason does not prescribe its form of synthesis in a generic manner but *qua* forms of three specific kinds of comprehension. Thus, reason demands that the understanding combine its cognitions not only *as if* they were to eventually constitute absolute comprehension but more specifically *as if* they were to become absolute comprehension of these three respects in which objects are conditioned. Equivalently, the understanding is to combine its cognitions not only *as if* the object and its connection contained the totality of conditions and the unconditioned but more specifically *as if* they contained the soul, the world, and God.

## 4.7 Conclusion

I have argued for a revision of the standard interpretation of the positive theoretical role of the transcendental ideas. On the standard interpretation, the three ideas in their regulative use are supposed to represent the goal or dictate the subject matter of three domains of empirical inquiry, i.e., three special sciences of nature. On my alternative reading, by contrast, the three ideas act as forms of synthesis or unity of cognition characteristic of comprehension. The division among them corresponds to three aspects in which objects of experience are conditioned and, thus, three ways in which they can be comprehended. Kant's view, as we have seen, is the unity characteristic of comprehension (also called systematic unity of cognition) can be prescribed to the understanding in no other way than postulating unconditioned objects whose cognitions *per impossibile* would serve as principles from which comprehension of conditioned objects could be completely derived. The objects of the transcendental ideas are such unconditioned objects that correspond to the three broad ways in which objects are conditioned.

My alternative interpretation diverges from the standard interpretation in three important ways. First, on my reading, the transcendental ideas are not simply methodological or heuristic

principles for scientific investigation; they have transcendental status as forms of synthesis or rules for combining representations, on a par with the categories. Second, insofar as these ideas are relevant for explanatory understanding, they concern not only the empirical mode of explanation (we have seen that this is what the cosmological idea specifically covers) but also non-empirical modes of explanation (those corresponding to the ideas of the Soul and God). Third, on my reading, the role of each idea is not delimited to a specific domain of objects or inquiry. All three ideas pertain to equally to all objects of experience insofar as they are all conditioned in the three respects I have described.

In closing, I want to point out how this alternative reading can provide a solution to the five interpretive puzzles raised earlier.

First, the alternative reading makes a clear sense of why the regulative role assigned to the transcendental ideas requires concepts of unconditioned objects. As we have seen, this role concerns the unity of comprehension as a higher degree of cognition. The peculiarity of this unity is that can be prescribed to the understanding in no other way than by postulating unconditioned object whose cognition would serve as a principle from which comprehension of a given conditioned object could be completely derived. This is why only concepts of unconditioned object could fit the purpose.

Second, it has also become clear in what sense the postulation of the unconditioned objects of the ideas through an as-if attitude gives rise to certain directives. On my reading, the relevant directives are reason's demand that the *understanding* combine its cognitions according to the unity of reason and, hence, that it combine its cognitions as if objects contained the unconditioned objects of the ideas—in Kant's words, as if "all the connection of things in the world of sense […] had their ground in this being of reason [i.e. the soul, the world-whole, or God]" (A681/B709).

Regarding the third and fourth puzzles, unlike the standard interpretation, my interpretation leaves behind no mystery about how the regulative use of the ideas are related to the three broad types of families of real conditions as well as "conditions of all representations in general". These can be

explained in terms of the three respects in which objects of experience are conditioned and, hence, can be comprehended.

Fifth and finally, my interpretation has the potential to elucidate the continuity between the theoretical and the practical use of the transcendental ideas, since it reveals the sense in which the theoretical use of the ideas properly belongs to reason's capacity as the faculty of cognition from principles, which is common to both its theoretical and practical use. While the ideas in their theoretical use are strictly speaking not themselves principles (since we have no cognition of their objects), they are nonetheless placeholders for principles from which, if they were accessible to us, we could derive theoretical comprehension in its completeness. By contrast, in their practical use, these ideas yield practical cognition of unconditioned objects. In this regard, they *are* principles from which a complete system of practical cognition (of freedom and morality)—hence, practical comprehension—can be derived. Therefore, my interpretation has an advantage over the standard interpretation in that it clarifies how the unity of theoretical and practical reason manifests itself with respect to the theoretical and the practical use of its ideas.

# Chapter 5

# The Idea of the Soul and the Explanatory Project of the Critique

## 5.1 Introduction

The previous chapter develops an alternative interpretation of the positive theoretical (or regulative) use of the transcendental ideas in general which makes room for my thesis that the idea of the soul is employed in the critique in such a way that provides a basis for its metaphysically substantive claims about mental faculties. On the standard reading, the idea of the soul in its regulative use serves to either represent the goal or demarcate the subject matter of *empirical* psychological investigation. By contrast, on the reading I propose, it acts as the form of *Begreifen*, which is usually translated as 'comprehension' but which I prefer to translate as 'explanatory understanding', of objects of experience *qua* representations. My aim in this final chapter is to fill in this reading with more detail and offer further defense for it.

I begin, in Section 5.2, by arguing that my interpretation satisfies all the desiderata of a plausible interpretation, which we have established through critical examination of previous interpretations. In Section 5.3, I show how my reading is supported by the way it accords with different aspects of Kant's view, especially on the unity of theoretical and practical reason and the role of the transcendental ideas with regard to the practical use of reason. In Section 5.4, I return to the problem of the source of the explanatory framework of the critique and show my account of Kant's concept of mind as the idea of the soul in regulative use can provide a satisfactory solution to it.

## 5.2 Five Interpretive Desiderata

Let us review some of the key results from our examination of previous interpretations of Kant's claims about the mind. These will provide us with desiderata for a plausible alternative interpretation. In Chapter 1, I argue that the critique has an explanatory task, that of explaining the possibility of synthetic *a priori* cognition. Kant's strategy in fulfilling this task is to appeal to our mental faculties as the explanatory grounds of the properties of objects that we can cognize *a priori*. However, the appeal to faculties in general can have explanatory purchase only against the background of an understanding of faculties as real grounding relation between a substance and its accidents. As I have pointed out, it is implausible that the reflective awareness of our own rational faculties can offer us such a metaphysical understanding of those faculties. In Chapter 2, I reject the anti-metaphysical interpretation, according to which Kant's account of mental faculties carries no metaphysical significance whatsoever. Besides the fact that on this reading it remains unclear how Kant's appeal to mental faculties is supposed to fulfil the critique's explanatory task, I have also shown that the philosophical import of Kant's central claims about mental faculties cannot be understood apart from the conception of these faculties as causal relations. In Chapter 3, we look at the two branches of the realist interpretation. The phenomenalist interpretation, which construes Kant's claims about the mind as empirical claims about a phenomenal object, is dismissed because it was incompatible with Kant's claims that the laws governing our higher faculties are non-empirical and his suggestion that the bearer of these faculties is a noumenal entity. While these considerations can be accommodated by the noumenalist interpretation, it is difficult to see how on this reading Kant's claims about the mind can be reconciled with his commitment to the doctrine of noumenal ignorance.

To avoid the problems that these previous interpretations encounter, a plausible alternative interpretation must be one that (1) identifies a conception of mind that is metaphysically substantive in such a way that it can serve as an explanatory framework for the Critical investigation and allow us

to see how Kant's claims about mental faculties can fulfil the explanatory role ascribed to them, (2) puts us in a position to explain our possession of this conception of mind, (3) specifies the justificatory basis for asserting this conception of mind with respect to ourselves, i.e., explains why we are warranted in taking ourselves to be or to have a mind as characterized by this conception, (4) accommodates Kant's apparent claim that the mind is a noumenal entity, and (5) allows us to see how Kant's conception of mind as noumenon is compatible his doctrine of noumenal ignorance.

The reading that I propose satisfies all five desiderata. On my reading, Kant's concept of mind is to be construed as the transcendental idea of the soul in regulative use. This idea, as we have seen in Chapter 4, is generated in an entirely *a priori* manner by reason. Further, the idea of the soul represents the subject of thinking as an immaterial noumenal substance and, hence, offers a conception of the mind that qualifies as a possible bearer of faculties that explain the properties objects possess *qua* representations. In this way, my reading satisfies desiderata (1), (2), and (4). My reading claims, in addition, that the idea of the soul is employed in Critical philosophy *regulatively* and, thus, in a way that does not issue in a claim to cognition. However, while we cannot know whether the idea of the soul veridically represents what we are in ourselves, we are entitled to conceive of ourselves according to it insofar as this idea is a condition for the possibility of explanatory understanding (more on this below). Therefore, my reading also satisfies desiderata (3) and (5).

Although my reading fulfills all the desiderata for a plausible interpretation, the main difficulty with this reading is that it lacks direct textual support. In the next section, we will consider some indirect textual and philosophical considerations in favor of this reading.

## 5.3 The Idea of the Soul as Self-Conception of Reason

My aim in this section is to highlight several ways in which my reading can be seen to fit with various aspects of Kant's Critical philosophy, which I hope will compensate for the lack of direct

textual evidence. In assessing the plausibility of my reading, it is important to keep in mind that *none* of the readings that we have previously considered enjoys direct textual support. The reason is simple: Kant never tells us what the mind or the bearer of faculties he investigates is. Therefore, the support for any account of Kant's concept of mind will necessarily be holistic, drawn from different ways in which it can be seen to fit into Kant's view as a whole.

Let us begin with a terminological consideration. In terms of frequency, the most common term Kant uses to refer to the mind as the bearer of the faculties is 'Gemüt'. Commentators often emphasizes Kant's use of this term as if it were an indicator that the mind or Kant's concept of it has a special status such that it cannot be subsumed under any metaphysical or representational categories in the Critical system.[1] The *Gemüt*, it is sometimes suggested, is that mysterious entity (if it is an entity at all) that is neither the soul, the immaterial substance that had preoccupied rationalist metaphysicians, nor the object of empirical psychology. Likewise, the concept of the *Gemüt* has completely nothing to do with the concept of the *Seele*, which Kant identifies as one of the pure concepts of reason. Moreover, the overemphasis on Kant's use of the term 'Gemüt' downplays the fact that he not infrequently uses 'Seele' to refer to the mind in the same context and attributes the same mental faculties to the *Seele*.[2] The fact that Kant shows no qualms about using 'Seele' to characterize the central subject of the critique raises a question about whether in using 'Gemüt' he really intends to signal his distance from the traditional metaphysical concept of the *Seele*.

An examination of Kant's discussion of these terms in the metaphysics lecture transcripts also reveals that far from being *sui generis* in the sense described, the term 'Gemüt' is in fact to be understood

[1] E.g. Béatrice Longuenesse, *Kant and the Capacity to Judge* (Princeton: Princeton University Press, 1998), 5–6; Robert Pippin, *Kant's Theory of Form: An Essay on the Critique of Pure Reason* (New Haven: Yale University Press, 1982), 219; Robert Pippin, "Kant on the Spontaneity of Mind," *Canadian Journal of Philosophy* 17, no. 2 (1987): 449.

[2] E.g. A78/B103, A94/B127, B416n, 5:10, 5:177, 5:196.

as related to 'Seele' rather than as opposed to it.[3] These two terms are also related to a third term that Kant more seldomly uses to refer to the mind, namely, 'Geist'. Here are some representative passages.

> One could call *anima* soul [Seele], the subject of feeling, *animus* mind [Gemüt], the subject of thoughts, and *spiritus* spirit [Geist] – as the subject of spontaneity. (*MD* 28:680)

> *anima* is the sensible. *animus* is the intellectual faculty of the soul [Seele]. *mens, nous* is also this. – Soul [Seele] and spirit [Geist] are to be sure two distinct relations but only two faculties of one and the same subject. (*MK*$_2$ 28:753)

> The self underlies consciousness and is what is peculiar to spirit [Geist]. But we can consider this self in three ways: I think as intelligence, i.e., the subject of thinking is intelligence. I think as subject which has sensibility, and am soul [Seele]. I think as intelligence and soul, and am a human being. (*MM* 29:878)

It is likely that in most contexts 'Gemüt', 'Seele', and 'Geist' are not used in their contrastive senses. But when they are, the contrast between them clearly has to do with the division among faculties of the mind into lower and higher faculties (see Chapter 3). 'Seele' appears to refer to the soul as the bearer of lower faculties, whereas both 'Gemüt' and 'Geist' refer to the soul as the bearer of higher faculties and, hence, an intelligence (some other passages seem to suggest that *Gemüt* is the mind *qua* bearer of *both* lower and higher faculties, e.g., *MvS* 28:483). Despite some ambiguity about how these terms exactly divide up the faculties of the mind, it is clear that they do not refer to distinct entities but rather to different perspectives on the same entity. This speaks against taking 'Gemüt' to indicate conceptual distance between Kant's concept of mind and the traditional concept of the soul, which becomes the idea of the soul in the critique.

---

[3] Anthropology lecture transcripts, which I cannot consider here, also contain similar passages. See Julian Wuerth, *Kant on Mind, Action, and Ethics* (Oxford: Oxford University Press, 2014), 219–21. According to Wuerth, based on the anthropology lectures, 'Seele' refers to the soul *qua* bearer of lower faculties, 'Gemüt' refers to the soul *qua* bearer of higher faculties when it is not purely active but *reactive*, and 'Geist' refers to the soul *qua* bearer of higher faculties when it is purely active.

It is important to note that while Kant regards the idea of the soul as the very same concept of Cartesian thinking substance that figures in rationalist metaphysics, its origin can be explained entirely within the framework of Critical philosophy. As Kant often stresses, the transcendental ideas are "generated in an entirely necessary way by reason according to its original laws" (A338/B396)—they are not "thought up only arbitrarily" but "presupposed necessarily by reason" (A337/B394). Kant's explanation of the transcendental ideas in terms of their origination from the *nature* of reason is crucial because it provides a basis for Kant's argument that although these ideas are generated by a dialectical (hence, erroneous), they must have legitimate and purposive use: "Everything grounded in the nature of our powers must be purposive and consistent with their correct use, if only we can guard against a certain misunderstanding and find out their proper direction" (A642/B670).

As pointed out in the introductory chapter, since the dialectical inferences are the *origin* of the transcendental ideas, the very content of these ideas—what they represent—consist in transcendental illusion. In other words, they do not represent objects that we have cognitive access to but objects made up by reason. Nonetheless, Kant does not think that this flaw prevents the ideas from fulfilling important and "indispensably necessary" roles in our cognitive and practical activity. The ideas can remain illusory in their content, but we need not be *deceived* by it—illusion does not entail deception. If we are mindful of their illusory character, we can refrain from ascribing objective validity to these ideas (at least from the "theoretical standpoint"), holding back our assent to their look or pretension of being veridical representations of objects. As long as we do this, we can put the ideas to purposive use in accordance with the ends of our faculties.

> It is not the idea itself but only its use that can be either **extravagant** (transcendent) or **indigenous** (immanent), according to whether one directs them straightway to a supposed object corresponding to them, or only to the use of the understanding in general regarding the objects with which it has to do; and all errors of subreption are

always to be ascribed to a defect in judgment, never to understanding or to reason. (A643/B671; cf. A669/B697)

As Kant emphasizes, no error is incurred as long as the ideas are directed at the use of the understanding rather than to their presumptive objects (with the intent to cognize them). He has in mind specifically the regulative use of the ideas, in which they serve to prescribe forms of a higher degree of cognition to the understanding.

However, recall that the regulative use is not the only legitimate use Kant accords to the transcendental ideas. They are assigned an even more central role in Kant's practical philosophy as the principal constituents of the postulates of pure practical reason. The practical justification of these postulates at the same time lends objectivity validity to the transcendental ideas, though only from the "practical standpoint". It is worth emphasizing again that the ideas that obtain objective reality or validity through the practical proof of the postulates are the very same ideas employed in the regulative use of reason. According to Kant, reason in practical use does not generate its own concepts but merely give reality to concepts generated by its theoretical use (see Section 1.5). It follows that even in the practical use, the ideas do not cease to present us with illusions of unconditioned objects. These illusions are their content, and practical reason does not generate new content here but merely confirms the content it takes over from theoretical reason. Nonetheless, the ideas in practical use are objectively valid because they are necessary for the realization of our practical vocation. The fact that the practical validation of the ideas does not "de-illusion" them is presumably the reason why Kant painstakingly insists that the ideas are valid *only* from the practical standpoint—their objective validity does not carry over to their use in the theoretical context. Practical cognition of the reality of the ideas does not form a basis for theoretical cognition.

The foregoing discussion is meant to allay the worry that the idea of the soul could not possibly be employed by Kant in his own philosophy because they originate from the dialectical use of reason

and, hence, their very content consists in transcendental illusion. As we have seen, despite their dialectical origin and illusory content, Kant makes use of them in a very fundamental way in his practical philosophy. Hence, it should not surprise us if one of these ideas would play an equally significant role with respect to Kant's theoretical philosophy or the Critical system as a whole.

Now, I would like to point to a more specific parallel between the theoretical and the practical case. Broadly speaking, in its practical use, the idea of the soul serves to articulate our self-conception as practical agents. According to Kant, insofar as we are immediately conscious of our practical reason as subject to the moral law, we are already thereby committed to conceiving ourselves as immortal souls—the postulate of "the *existence* and personality of the same rational being continuing *endlessly*" is implicated in the moral law (*KpV* 5:112). But we must be able to think of ourselves as one and the same subjects with regard to the exercise of both our practical and theoretical reason. The two uses of our reason (or three if we count the "affective" use of reason, with respect to feeling) do not take place in isolation but in tandem—our rational faculties constitute a unity expressed by the I.

> The singul(arity) (singularity not simplicity) against which some assume three souls (*animam vegetativam* […], *animam sensitivam* […], and *animam rationalem* […]); but three faculties do not give three souls, and moreover the I brings everything to unity. The question cannot be settled otherwise. A human being constitutes a unity, and we cannot call the principles of life in various parts souls. (*MD* 28:683)

> The fact that the human being can have the "I" in his representations raises him infinitely above all other living beings on earth. Because of this he is a *person*, and by virtue of the unity of consciousness through all changes that happen to him, one and the same person […]. (*Anth.* 7:127; cf. B134n)

Hence, if it is correct to take the idea of the soul to represent our practical self-conception (i.e., our metaphysical conception of the practical I), this gives us a reason to think that the same idea would represent the theoretical conception of ourselves as well. Or, more strongly, given Kant's claim of the unity of reason, it is plausible to think that the idea of soul just is the metaphysical conception of the

rational subject that applies across all uses of reason insofar each can be shown to be in need of such a conception. As I have been arguing, theoretical reason, no less that practical reason, needs a metaphysical conception of the thinking subject.

Let us return to the origin of the idea of the soul in the dialectical inferences of theoretical reason, which is the subject of the Paralogisms. We can see that the idea of the soul is generated exactly in accordance with the need of theoretical reason for a metaphysical conception of the subject of thinking and, thus, in a way that is consistent with Kant's dictum that "everything grounded in the nature of our powers must be purposive". As we have seen, what theoretical reason needs is a concept that bridges the gap between the thinking subject conceived non-objectively (the I of pure apperception) and the metaphysical bearer of faculties that ground *a priori* forms of objects. As the presentation of the individual paralogisms makes clear, the dialectical inferences that generate the idea of the soul are in general inferences leading from logical features of the thinking subject to metaphysical features of the substantive soul. Therefore, the illusion that constitutes the content of the idea of the soul just is what is needed to fill in the gap.

> It is indeed very illuminating that I cannot cognize as an object itself that which I must presuppose in order to cognize an object at all; and that the determining Self (the thinking) is different from the determinable Self (the thinking subject) as cognition is different from its object. Nevertheless, nothing is more natural and seductive than the illusion of taking the unity in the synthesis of thoughts for a perceived unity in the subject of these thoughts. One could call it the subreption of hypostatized consciousness (*apperceptionis substantiate*). (A402)

In an earlier draft of the Paralogisms, Kant explains this "subreption" as a transcendental illusion in which "the unity of apperception, which is subjective, is taken for the unity of the subject as a thing" (18:224, my translation).

Note, however, that the idea of the soul arises not simply from the inference from the ways in which our representations are conditioned with respect to pure apperception to the unconditioned

condition corresponding to these ways of being conditioned. This inference by itself is not erroneous. The error (and, hence, what makes this inference dialectical) lies instead in the misunderstanding of the "conditioning relation" between our representations and apperception. Obviously, this relation is merely formal and, hence, not a real conditioning relation, which is a metaphysical dependence relation. As Kant points out, as long as the nature of this conditioning relation is understood correctly, nothing is wrong in holding the I to be unconditioned with respect to the ways in which our representations are conditioned by it.

> Apperception is itself the ground of the possibility of the categories, which for their part represent nothing other than the synthesis of the manifold of intuition, insofar as that manifold has unity in apperception. Self-consciousness in general is therefore the representation of that which is the condition of all unity, and yet is itself unconditioned. (A401)

However, thinking of the I as an unconditioned condition in the formal sense also does not license us to apply the categories to it and thereby represent it as an object. Hence, the basic error of the paralogisms is to mistake the *formal* conditioning relation between our representations and the I to be a *real* conditioning relation and, hence, the I itself to be an object.

> Because in thinking in general we abstract from every relation of the thought to any object […], the synthesis of conditions of a thought in general (No. 1) is not objective at all, but merely a synthesis of thought with the subject, which is, however, falsely taken to be a synthetic representation of an object. […] Because, further, the only condition accompanying all thinking is the I, in the universal proposition "I think," […]. But it is only the formal condition, namely the logical unity of every thought, in which I abstract from every object; and yet it is represented as an object that I think, namely I itself, and its unconditioned unity. (A397-98)

As Kant implies in the paragraphs that follow, this basic mistake is, in effect, the mistake of taking pure apperception to be *intuition* of the self. But since this self is supposed to an unconditioned object, an intuition of it could only be non-sensible. Hence, this mistake also involves (at least implicitly)

taking our understanding to be intuitive, i.e., to be an understanding "through whose self-consciousness the manifold of intuition would at the same time be given" (B138). This, I suggest, is how we should understand the fundamental mistake of the paralogisms as driven by the nature of reason, since reason has a natural propensity to "pretend" that our understanding is non-discursive (see Sections 4.4 and 4.5).

I take this account of the origin of the idea of the soul to be Kant's explanation of our *possession* of a metaphysical conception of the mind. His explanation for our *justification* in asserting this conception with respect to ourselves, as we have seen, has to do with its role in prescribing a form of unity of cognition to the understanding that is characteristic of explanatory understanding. As I have argued, in the case of the idea of the soul, the relevant explanatory understanding concerns the features of objects that is due to our way of representing them and, hence, can be cognized *a priori*. To put the same point differently, unless we conceive of our representations *as if* they were accidents inhering in a thinking substance in virtue of its representational powers and, accordingly, of ourselves *as if* we were that substance, we would not be able to comprehend or make sense of the fact that certain features of objects are due to our way of representing them such that these features can be cognized by us *a priori*. If this is correct, the underlying assumption is that the only kind of explanatory relation (at least for us) is the relation of *metaphysical* dependence, i.e., the real conditioning relation, which holds between existing *things*. In this way, the relation of dependence between representations and the thinking subject can only be explanatory insofar as this relation is conceived as a relation of metaphysical dependence and, hence, insofar as its relata are conceived as things. Thus, it is reason's demand of explanatory understanding that ultimately justifies the hypostasis of the thinking subject that is the origin of the "objective" concept of mind.

The tight connection between explanatory understanding and metaphysics, therefore, explains why Kant needs to formulate his philosophical claims as metaphysical substantive claims about

faculties of a thinking substance: it is only in this way that these claims can constitute explanatory understanding of the possibility of synthetic *a priori* cognition. As I have argued, in the absence of a metaphysical framework, these claims would remain claims about forms of representations (see Chapter 2), which are devoid of explanatory significance.

However, since this metaphysical conception of the thinking subject is grounded in reason's demand for explanatory understanding and not in our cognitive access to what the thinking subject is in itself, it does not give rise to a claim to cognition of the thinking subject as an object.[4] Kant expresses this point by contrasting two ways in which an object can be "given" (A670/B698). In an intuition, an object is given "absolutely", which means that it is presented to consciousness in such a way that it can be determined by a concept in judgment and thereby cognized. By contrast, when an object is given "in an idea", it is presented to consciousness as a "schema" of the unity of reason—as the object without whose postulation reason "cannot think its unity" (A681/B790).

Kant sometimes also characterizes the status of the objects of the transcendental idea as *entia rationis* or thought-entities [Gedankending] as opposed to an actual [wirklich] thing.[5] Similarly, he also says that these objects are not real things but "analogues" of real things or of objects of experience.[6] These two characterizations appear complementary in the following way. The former characterization highlights the fact that the ideas are concepts of objects for which, however, no object can be given "absolutely".[7] The latter characterization captures the explanation how, notwithstanding the absence of an object being so given, it is possible to employ the categories to represent the objects corresponding to these concepts. As Kant point out in this passage, we employ the categories to

[4] Again, this does not preclude the possibility of viewing Kant's claims about the mind *qua* claims about forms of representation as *formal* cognition. See Chapter 2.

[5] Cf. A669/B697, A681/B709.

[6] Cf. A647/B702, A674-75/B702-3, A678/B706, A696/B724.

[7] Cf. Kant's definition of *ens rationis* at A291/B348.

represent the objects of the ideas on the basis of an *analogy* with objects of experience. This explains why he calls them analogues of objects of experience.

> Transcendent ideas have a merely intelligible object, which one is of course allowed to admit as a transcendental object, but about which one knows nothing; but for the assumption of such an object, in thinking it as a thing determinable by its distinguishing and inner predicates, we have on our side neither grounds of its possibility (since it is independent of all concepts of experience) nor the least justification, and so it is a mere thought-entity. […] But if we once take the liberty of assuming a reality subsisting by itself outside the entire field of sensibility, then appearances are regarded only as contingent ways intelligible objects are represented by beings who are themselves intelligences; and because of this, nothing is left for us but **the analogy by which we utilize concepts of experience** in making some sort of concept of intelligible things, with which we have not the least acquaintance as they are in themselves. (A565-66/B593-94, boldface added)

Kant's point seems to be that because we have no grasp of *how* the categories are instantiated by things in themselves or perhaps even *whether* they could be instantiated by things in themselves at all (recall that the categories, for all we know, are forms of discursive understanding), the only significance they can have when employed (in an attempt) to represent noumenal objects, like the objects of the idea, must be derivative of the significance they bear when applied to objects of experience. If my interpretation of Kant's concept of mind is correct, it follows that the use of the categories (and derivative concepts thereof) that is involved in Kant's account of mental faculties must likewise be understood as having an analogical significance.[8]

[8] It is striking that, for Kant, the analogical employment of the categories is not limited to the theoretical use of the ideas only. He explicitly states that even with respect to the practical use of the ideas, the categories are to be understood as having a merely analogical significance: "Meanwhile, I would still be warranted in applying [the categories] in regard to their practical use, which is always directed to objects of experience, according to their analogical significance in their theoretical use, to freedom and the free subject, since by them I understand merely the logical functions of subject and predicate, ground and consequence, in accordance with which actions or

The notion that the objects of the transcendental ideas, including the soul, are analogues of objects of experience has a special significance for my reading because it provides a nexus to what comes close to a piece of direct textual evidence for it. Consider the following remark from the *Opus Postumum*:

> Transcendental philosophy [concerns] the relation between theoretical/speculative and moral/practical reason in connection with one another in a system of self-cognition according to the *analogy of an object of possible experience*" (21:108, italics added, my translation).

What Kant seems to be saying, then, is that the critique (transcendental philosophy) is to be considered self-cognition whereby the self is to be construed as an object of a transcendental idea, an analogue of an object of experience. However, Kant wrote this note late in his life and, so, there is no guarantee that it reflects his view in the Critical period. Nonetheless, it does show that Kant himself has considered something like the view I am ascribing to him and that my reading is not too wild to at least be on the table.

## 5.4 Explaining Synthetic *a priori* Cognition

We began our inquiry with the question concerning the source of the metaphysical conception of the mind that underlies the critique's explanatory framework. But how can the regulative use of the idea of the soul, as I interpret it, be seen to accord with the explanatory goal of the critique?

The starting point of the critique is the fact that we can cognize certain properties of objects of experience *a priori*. These are properties that apply to them with necessity and strict universality. We know, for instance, that every object we can experience must be in space and time and that every

effects are determined in conformity to those laws in such a way that they can at the same time always be explained conformably to the laws of nature and the categories of substance and cause, although they arise from a wholly different principle" (B431-32).

change has a cause. Since experience, as Kant points out, never suffices to establish necessity and strict universality, it can only be the case that we arrive at cognition of these properties independently of experience. As we have seen in Chapter 1, this possibility seems puzzling and calls for explanation because there seems to be only two ways in which a necessary (i.e., non-accidental) agreement can obtain between our representations and their objects: either our representations bring their objects into existence or they are caused by their objects. The first is ruled out because our understanding is finite and discursive; the second is also ruled out because the representations in question are *a priori*.

Kant's explanatory strategy with respect to the possibility of synthetic *a priori* cognition relies on three basic assumptions. The first assumption is that objects of experience are representations and not things in themselves.[9] The second assumption is that representations have, apart from their empirical matter, *a priori* forms. The third assumption is that *a priori* forms of representations are metaphysically grounded in the subject's representational faculties. Therefore, we can explain properties of objects that can be cognized *a priori* in terms of the *a priori* forms of representations. The *a priori* forms of representations are in turn explained in terms of the character of the subject's representational faculties. For example, in the Transcendental Aesthetic, Kant explains the spatiotemporality of objects by appeal to the *a priori* forms of intuition, which are grounded in the faculty of sensibility. In the Transcendental Analytic, the categorial properties of objects are explained in terms of the *a priori* forms of synthesis undertaken by the faculty of understanding.

Indeed, all three assumptions require justification since none of them is self-evident. Here, however, we are interested only in the justificatory basis for the third assumption. As I have argued,

[9] "If the objects with which our cognition has to do were things in themselves, then we would not be able to have any a priori concepts of them at all. […] That which is merely in us cannot determine the constitution of an object distinct from our representations, i.e., be a ground why there should be a thing that corresponds to something we have in our thoughts, and why all this representation should not instead be empty" (A129, cf. Cf. A92/B125-26, A370-73, A491-94/B519-22).

by Kant's own light, representations, considered from the standpoint of Critical investigation or the "transcendental" standpoint, are not the sort of thing that can stand in a metaphysical grounding relation ('faculty' is Kant's technical term for just such a relation) to the thinking subject or, for that matter, any metaphysical grounding relation whatsoever. Likewise, the thinking subject, considered from the same standpoint, is also not the sort of thing that can stand in such a relation. This follows from the fact that we are not considering representations and the thinking subject, from the "empirical" standpoint, as objects of experience. Hence, for the purpose of Critical investigation, representations and their subject can only be understood as apprehended through pure apperception. However, apperception presents neither representations nor their subject as objects (something that falls under the categories). But, presumably, only objects are possible relata of a metaphysical grounding relation. Therefore, the third assumption seems to commit a category mistake.

My proposal is that we can understand the justificatory basis for the third assumption as lying in the regulative use of the idea of the soul. As we have seen, this idea supplies a metaphysical conception of the thinking subject as a substance in which representations inhere in virtue of its representational powers. Hence, if we are entitled to employ the idea of the soul with regard to ourselves, we are also entitled to assert the third assumption. And, in my view, we are justified in conceiving ourselves according to the idea of the soul *precisely* because it makes the explanatory understanding that the critique seeks possible. This claim needs some clarification.

Very briefly, according to the interpretation put forward in the last chapter, the transcendental ideas are forms of comprehension and, hence, conditions of its possibility. Reason demands the realization of comprehension while assuming its completion to be possible. We are justified in postulating the objects of the ideas because comprehension (of conditioned objects) is possible only if the objects of the ideas exist and, moreover, the understanding can realize the unity of comprehension in no other way than combining cognitions *as if* these objects existed. Now, there are

three broad respects in which objects of experience are conditioned and, hence, three ways in which they can be comprehended. The idea of the soul, as I have argued, is the concept of the unconditioned condition corresponding to the respect in which objects *qua* representations are conditioned. In other words, the idea of the soul is the idea of a complete explanation corresponding to the explanation which objects *qua* representations stand in need of. As we have seen, in seeking the explanation of the possibility of synthetic *a priori* cognition, the critique also seeks the explanation of properties that objects instantiate in virtue of being representations. In this way, the idea of the soul can be seen as playing the role of the form of precisely the explanatory understanding that the critique seeks.

To put the view under proposal more simply, we are warranted in taking ourselves to be thinking substances in which our representations inhere as accidents because otherwise we would not be able to explain or make sense of the fact that we can cognize certain properties of objects *a priori*. This fact is not inconsequential because, in Kant's view, our cognition in general is so thoroughly structured by synthetic *a priori* judgments that no cognition would be possible, were these judgments not objectively valid. To see why it is imperative (at least for Kant) that this fact be explained we need only to recognize that Humean skepticism fundamentally rests on the assumption that this fact cannot possibly be explained or rendered intelligible.

However, it is important to note that this account need not imply that we *arrive* at the idea of the soul by way of something like an inference to the best explanation. If this were the case, we could not make sense why Kant claims that the transcendental ideas are "given as problems by the nature of reason itself" (A327/B384; cf. A323/B380). This raises the question which of the following is primitive: (i) our apprehension of objects as conditioned (i.e., to-be-explained) in a certain respect, (ii) our grasp of the conditioning relation which corresponds to a particular respect in which an object is conditioned and in terms of which explanation for it is given, (iii) the transcendental ideas as the representation of the unconditioned conditions corresponding to particular conditioning relations.

Given Kant's emphasis on the primitivity of the ideas as pure concepts of reason, the most plausible option is to take the ideas to be primitive. If this is correct, it follows that the ideas dictate, at the most general level, the modes of explanation that are available to us—they specify in a highly general way what can count as explanation for a given explanandum (this must be highly general because there are only three modes which the variety of explanation we actually give can fall under). The idea that there must be some *a priori* constraints or criteria on what counts as explanation for a given explanandum is not without philosophical motivation, since the fact that a given object (e.g., the fact that bodies fall to the ground) stands in need of explanation does not entail anything about what may count as an explanation for it. In other words, the feature of an object in virtue of which it stands in need of explanation alone (in the absence of a worldview, a scientific paradigm, a system of beliefs etc.) does not in any way constrain what could count as an explanation for the object. In Kantian terms, simply from our apprehension of a given object as conditioned, we cannot obtain the concept of the *specific* conditioning relation with respect to which the object is conditioned.

My suggestion is that our concepts of particular conditioning relations are partially grounded in the transcendental ideas. More specifically, the ideas specify three broadest ways in which objects are conditioned and, hence, three broadest types (or families) of conditioning relations we may appeal to in our explanation by representing the unconditioned conditions corresponding to these types (or families) of conditioning relations. In this sense, the ideas set the most general constraints on what counts as explanation for a given explanandum.

To articulate these constraints is beyond the scope of the present work. Nonetheless, we can readily see what implication this suggestion has on the relationship between the idea of the soul and the explanation of the possibility of synthetic *a priori* cognition. The fact that our cognition is permeated by synthetic *a priori* claims about objects may strike us as standing in need of explanation and, thus, embark us on the path of Critical investigation. However, without the aid of the idea of the

soul, we would have no inkling of what shape the explanation for this should take. Fortunately, since we are by nature under the spell of transcendental illusion, we *already* regard ourselves as substance and our representations as accidents inhering in us. And this self-conception turns out to be just what we need in order to make sense of why objects have properties we can cognize *a priori*. The role of the idea of the soul in enabling this explanatory understanding in turns justifies us to assert the self-conception originally imposed upon us by reason. While we derive the warrant for the use of the idea of the soul from its role in enabling explanatory understanding, the idea itself has quite a different source, namely, transcendental illusion arising, as we have seen, from reason's propensity to pretend that our understanding were non-discursive.[10] It is in this sense that the idea of the soul is not derived from inference to the best explanation. To borrow terms Kant uses in a different context, in the case of the idea of the soul, the *quid facti* and the *quid juris* come apart. Its *quid juris* (justificatory basis) lies in its status as an enabling condition of explanatory understanding with regard to the way in which objects are conditioned *qua* representations. Its *quid facti* (explanation of possession), however, lies in the dialectical nature of our reason.

## 5.5 Conclusion

In this chapter, I have shown that my proposal that Kant's concept of mind is to be construed in terms of the idea of the soul in regulative use satisfy all the desiderata of a plausible interpretation, which we have established through our critical examination of previous interpretations. I have also argued that despite the lack of direct textual evidence (from which other interpretations also suffer), my reading is consistent with many aspects of Kant's thinking that are well-attested in the text,

[10] Note that this reading of the source of the transcendental illusion underlying the idea of the soul does not seem to fit Grier's influential account that transcendental illusion in general arises from the move from "P1" to "P2". See Michelle Grier, *Kant's Doctrine of Transcendental Illusion* (Cambridge: Cambridge University Press, 2001).

especially his views on the unity of reason and the role of the ideas in the practical domain. Finally, I elaborate on how the alternative interpretation of the regulative use of the idea of the soul developed in the last chapter provides an answer to the problem of the origin of Kant's concept of mind that underlies the explanatory framework of the critique. As we have seen, Kant's concept of mind stands in two-fold relation to reason. On the hand, the concept has its origin in an illusion imposed on us by reason's dialectical nature. On the hand, the concept derives its warrant from its role in enabling us to fulfil (though never completely) reason's demand that moves the critique and the other inquiries.

# Bibliography


Abaci, Uygar. "Kant, The Actualist Principle, and The Fate of the Only Possible Proof." *Journal of the History of Philosophy* 55, no. 2 (2017): 261–91.

———. *Kant's Revolutionary Theory of Modality*. Oxford: Oxford University Press, 2019.

Allais, Lucy. *Manifest Reality: Kant's Idealism and His Realism*. Oxford: Oxford University Press, 2015.

Allison, Henry. "Kant's Refutation of Materialism." *The Monist* 72, no. 2 (1989): 190–208.

———. *Kant's Transcendental Idealism: An Interpretation and Defense*. 1st ed. New Haven: Yale University Press, 1983.

———. *Kant's Transcendental Idealism: An Interpretation and Defense*. 2nd ed. New Haven: Yale University Press, 2004.

———. "On Naturalizing Kant's Transcendental Psychology." *Dialectica* 49, no. 2/4 (1995): 335–51.

Ameriks, Karl. "From Kant to Frank: The Ineliminable Subject." In *The Modern Subject*, edited by Karl Ameriks and Dieter Sturma, 217–30. Albany: SUNY Press, 1995.

———. "Kantian Apperception and the Non-Cartesian Subject." In *Kant and the Historical Turn: Philosophy as Critical Interpretation*, 51–66. Oxford: Oxford University Press, 2006.

———. *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*. 2nd ed. Oxford: Clarendon Press, 2000.

———. "Kant's Transcendental Deduction as a Regressive Argument." *Kant Studien* 69, no. 1–4 (1978): 273–87.

Anderson, R. Lanier. "Synthesis, Cognitive Normativity, and the Meaning of Kant's Question, How Are Synthetic Cognitions a Priori Possible?" *European Journal of Philosophy* 9, no. 3 (2001): 275–305.

———. *The Poverty of Conceptual Truth*. Oxford: Oxford University Press, 2015.

Baumgarten, Alexander. *Metaphysics: A Critical Translation with Kant's Elucidations, Selected Notes, and Related Materials*. Translated by Courtney Fugate and John Hymers. London: Bloomsbury, 2014.

Beck, Lewis. "Toward a Meta-Critique of Reason." In *Essays on Kant and Hume*, 20–37. New Haven: Yale University Press, 1978.

Beiser, Frederick. *The Fate of Reason: German Philosophy from Kant to Fichte*. Cambridge: Harvard University Press, 1993.

Bengson, John. "The Unity of Understanding." In *Making Sense of the World: New Essays on the Philosophy of Understanding*, edited by Stephen Grimm, 14–53. Oxford: Oxford University Press, 2017.

Bennett, Jonathan. *Kant's Dialectic*. Cambridge: Cambridge University Press, 1974.

Boyle, Matthew. "Die Spontaneität des Verstandes bei Kant und einigen Neokantianern." *Deutsche Zeitschrift für Philosophie* 63, no. 4 (2015): 705–26.

———. "Kant on Categories and the Activity of Reflection." In *Palgrave Handbook of German Idealism and Analytic Philosophy*, edited by James Conant and Jonas Held. London: Palgrave Macmillan, Forthcoming.

———. "Kant on Logic and the Laws of the Understanding." In *The Logical Alien*, edited by Sofia Miguens, 117–44. Cambridge: Harvard University Press, 2020.

———. "Transparency and Reflection." *Canadian Journal of Philosophy* 49, no. 7 (2019): 1012–39.

———. *Transparency and Reflection: A Study of Self-Knowledge and the First Person Perspective*. Oxford: Oxford University Press, Forthcoming.

Brink, Claudi. "Spontaneity and Teleology in Kant's Theory of Apperception." Ph.D. Dissertation, University of California, San Diego, 2022.

Brook, Andrew. *Kant and the Mind*. Cambridge: Cambridge University Press, 1994.

Cassam, Quassim. *Self and World. Self and World*. Oxford: Oxford University Press, 1999.

———. *The Possibility of Knowledge*. Oxford: Oxford University Press, 2007.

Chignell, Andrew. "Kant, Modality, and the Most Real Being" 91, no. 2 (2009): 157–92.

———. "Kant on Cognition, Givenness, and Ignorance." *Journal of the History of Philosophy* 55, no. 1 (2017): 131–42.

Dyck, Corey. *Kant and Rational Psychology*. Oxford: Oxford University Press, 2014.

Eilan, Naomi. "The First Person Perspective." *Proceedings of the Aristotelian Society* 95, no. 1 (1995): 51–66.

Engstrom, Stephen. "Understanding and Sensibility." *Inquiry* 49, no. 1 (2006): 2–25.

———. "Unity of Apperception." *Studi Kantiani* 26 (2013): 37–54.

Fisher, Mark, and Eric Watkins. "Kant on the Material Ground of Possibility." *The Review of Metaphysics* 52, no. 2 (1998): 369–95.

Förster, Eckart, ed. *Kant's Transcendental Deductions: The Three Critiques and the Opus Postumum*. Palo Alto: Stanford University Press, 1989.

Friedman, Michael. "Causal Laws and the Foundations of Natural Science." In *The Cambridge Companion to Kant*, edited by Paul Guyer, 161–99. Cambridge: Cambridge University Press, 1992.

———. *Kant and the Exact Sciences*. Cambridge: Harvard University Press, 1992.

———. "Kant on Science and Experience." In *Early Modern Philosophy: Mind, Matter, and Metaphysics*, edited by Christia Mercer and Eileen O'Neill, 262–73. Oxford: Oxford University Press, 2005.

———. *Kant's Construction of Nature: A Reading of the Metaphysical Foundations of Natural Science*. Cambridge: Cambridge University Press, 2013.

———. "Regulative and Constitutive." *The Southern Journal of Philosophy* 30, no. Supplement (1992): 73–102.

Gaukroger, Stephen. *Explanatory Structures: A Study of Concepts of Explanation in Early Physics and Philosophy*. Atlantic Highlands: Humanities Press, 1978.

Ginsborg, Hannah. "The Appearance of Spontaneity." In *Self, World, and Art: Metaphysical Topics in Kant and Hegel*, edited by Dina Emundts, 119–44. Berlin/Boston: De Gruyter, 2013.

Glezer, Tal. *Kant on Reality, Cause, and Force*. Cambridge: Cambridge University Press, 2017.

Gomes, Anil, and Andrew Stephenson, eds. *Kant and the Philosophy of Mind: Perception, Reason, and the Self*. Oxford: Oxford University Press, 2017.

Grier, Michelle. *Kant's Doctrine of Transcendental Illusion*. Cambridge: Cambridge University Press, 2001.

Grüne, Stefanie. "Givenness, Objective Reality, and A Priori Intuitions." *Journal of the History of Philosophy* 55, no. 1 (2017): 113–30.

———. "Kant and the Spontaneity of the Understanding." In *Self, World, and Art: Metaphysical Topics in Kant and Hegel*, edited by Dina Emundts, 145–76. Berlin/Boston: De Gruyter, 2013.

Guyer, Paul. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press, 1987.

———. "Psychology and the Transcendental Deduction." In *Kant's Transcendental Deductions: The Three Critiques and the Opus Postumum*, edited by Eckart Förster, 47–68. Palo Alto: Stanford University Press, 1989.

———. "Reason and Reflective Judgment: Kant on the Significance of Systematicity." *Noûs* 24, no. 1 (1990): 17–43.

———, ed. *The Cambridge Companion to Kant*. Cambridge: Cambridge University Press, 1992.

———, ed. *The Cambridge Companion to Kant's Critique of Pure Reason*. Cambridge: Cambridge University Press, 2010.

———. "The Unity of Reason: Pure Reason as Practical Reason in Kant's Early Conception of the Transcendental Dialectic." *The Monist* 72, no. 2 (1989): 139–67.

Haag, Johannes. "Faculties in Kant and German Idealism." In *The Faculties: A History*, edited by Dominik Perler, 298–246. Oxford: Oxford University Press, 2015.

Hamann, Johann Georg. "Metakritik über den Purismum der reinen Vernunft." In *Hamann's Schriften*, edited by Friedrich von Roth, 7:1–16. G. Reimer, 1825.

Henrich, Dieter. *Aesthetic Judgment and the Moral Image of the World: Studies in Kant*. Palo Alto: Stanford University Press, 1992.

———. "Kant's Notion of a Deduction and the Methodological Background of the First Critique." In *Kant's Transcendental Deductions: The Three Critiques and the Opus Postumum*, edited by Eckart Förster, 29–46. Palo Alto: Stanford University Press, 1989.

———. "On the Unity of Subjectivity." In *The Unity of Reason: Essays on Kant's Philosophy*, edited by Richard Velkley, 17–54. Cambridge: Harvard University Press, 1994.

———. "The Moral Image of the World." In *Aesthetic Judgment and the Moral Image of the World: Studies in Kant*, 3–28. Palo Alto: Stanford University Press, 1992.

———. *The Unity of Reason: Essays on Kant's Philosophy*. Edited by Richard Velkley. Cambridge: Harvard University Press, 1994.

Heßbrüggen-Walter, Stefan. *Die Seele Und Ihre Vermögen: Kants Metaphysik Des Mentalen in Der Kritik Der Reinen Vernunft*. Paderborn: Mentis, 2004.

Hoffer, Noam. "Kant's Regulative Metaphysics of God and the Systematic Lawfulness of Nature." *The Southern Journal of Philosophy* 57, no. 2 (2019): 217–39.

Hogan, Desmond. "How to Know Unknowable Things in Themselves." *Noûs* 43, no. 1 (2009): 49–63.

———. "Noumenal Affection." *Philosophical Review* 118, no. 4 (2009): 501–32.

———. "Three Kinds of Rationalism and the Non-Spatiality of Things in Themselves." *Journal of the History of Philosophy* 47, no. 3 (2009): 355–82.

Hume, David. *A Treatise of Human Nature*. Edited by David Norton and Mary Norton. Oxford: Oxford University Press, 2000.

Johnson, Robert. "The Moral Law as Causal Law." In *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*, edited by Jens Timmermann, 82–101. Cambridge: Cambridge University Press, 2009.

Kitcher, Patricia. "Kant's Cognitive Self." In *Kant's Critique of Pure Reason: Critical Essays*, edited by Patricia Kitcher, 59–84. Lanham: Rowman & Littlefield, n.d.

———. *Kant's Thinker*. Oxford: Oxford University Press, 2011.

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

Kleingeld, Pauline. "Kant on the Unity of Theoretical and Practical Reason." *Review of Metaphysics* 52, no. 2 (1998): 500–528.

———. "The Conative Character of Reason in Kant's Philosophy." *Journal of the History of Philosophy* 36, no. 1 (1998): 77–97.

Klemme, Heiner. *Kants Philosophie des Subjekts*. Hamburg: Meiner, 1996.

Kohl, Markus. "Kant on the Inapplicability of the Categories to Things in Themselves." *British Journal for the History of Philosophy* 23, no. 1 (2015): 90–114.

———. "Kant's Standpoint Distinction." *Kantian Review* 23, no. 2 (2018): 229–55.

Korsgaard, Christine M. "Realism and Constructivism in Twentieth-Century Moral Philosophy." *Journal of Philosophical Research* 28, no. Supplement (2003): 99–122.

Kraus, Katharina. *Kant on Self-Knowledge and Self-Formation*. Cambridge: Cambridge University Press, 2020.

———. "The Parity and Disparity Between Inner and Outer Experience in Kant." *Kantian Review* 24, no. 2 (2019): 171–95.

———. "The Soul as the 'Guiding Idea' of Psychology: Kant on Scientific Psychology, Systematicity, and the Idea of the Soul." *Studies in History and Philosophy of Science Part A* 71 (2018): 77–88.

Kreines, James. "Kant on the Laws of Nature: Laws, Necessitation, and the Limitation of Our Knowledge." *European Journal of Philosophy* 17, no. 4 (2009): 527–58.

Land, Thomas. "Epistemic Agency and the Self-Knowledge of Reason: On the Contemporary Relevance of Kant's Method of Faculty Analysis." *Synthese* 198, no. 13 (Supplement) (2021): 3137–54.

Levin, Janet. "Functionalism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University, 2018. https://plato.stanford.edu/archives/fall2018/entries/functionalism.

Longuenesse, Béatrice. *Kant and the Capacity to Judge*. Princeton: Princeton University Press, 1998.

———. *Kant on the Human Standpoint*. Cambridge: Cambridge University Press, 2005.

———. "Kant's Categories and the Capacity to Judge." In *Kant on the Human Standpoint*, 17–38. Cambridge: Cambridge University Press, 2005.

———. "Kant's Deconstruction of the Principle of Sufficient Reason." In *Kant on the Human Standpoint*, 117–42. Cambridge: Cambridge University Press, 2005.

Lu-Adler, Huaping. *Kant and the Science of Logic: A Historical and Philosophical Reconstruction*. Oxford: Oxford University Press, 2018.

———. "Ontology as Transcendental Philosophy." In *Kant's Lectures on Metaphysics: A Critical Guide*, edited by Courtney Fugate, 53–73. Cambridge: Cambridge University Press, 2018.

Marshall, Colin. "Kant's Metaphysics of the Self." *Philosophers' Imprint* 10 (2010): 1–21.

Massimi, Michela, and Angela Breitenbach, eds. *Kant and the Laws of Nature*. Cambridge: Cambridge University Press, 2017.

McLaughlin, Peter. "Transcendental Presuppositions and Ideas of Reason." *Kant-Studien* 105, no. 4 (2014): 554–72.

McLear, Colin. "'I Am the Original of All Objects': Apperception and the Substantial Subject." *Philosophers' Imprint* 20, no. 26 (2020): 1–38.

McNulty, Michael Bennett. "Rehabilitating the Regulative Use of Reason: Kant on Empirical and Chemical Laws." *Studies in History and Philosophy of Science Part A* 54 (2015): 1–10.

Meerbote, Ralf. "Kant's Functionalism." In *Historical Foundations of Cognitive Science*, edited by John-Christian Smith, 161–87. Dordrecht: Springer Netherlands, 1991.

Messina, James. "Kant's Necessitation Account of Laws and the Nature of Natures." In *Kant and the Laws of Nature*, edited by Michela Massimi and Angela Breitenbach, 131–49. Cambridge: Cambridge University Press, 2017.

Mohr, Georg. *Das Sinnliche Ich*. Würzburg: Könighausen & Neumann, 1991.

Moore, Adrian. "Sense-Making From a Human Point of View." In *The Cambridge Companion to Philosophical Methodology*, edited by Giuseppina D'Oro and Søren Overgaard, 44–55. Cambridge: Cambridge University Press, 2017.

———. "The Transcendental Doctrine of Method." In *The Cambridge Companion to Kant's Critique of Pure Reason*, edited by Paul Guyer, 310–26. Cambridge: Cambridge University Press, 2010.

Pendlebury, Thomas. "Grounds of Sense: Kant's Image of Theoretical Finitude." Ph.D. Dissertation, Harvard University, 2020.

———. "The Shape of the Kantian Mind." *Philosophy and Phenomenological Research*, Forthcoming.

Pierris, Graciela de. "The Constitutive A Priori." *Canadian Journal of Philosophy* 22 (Supplementary Volume 18) (1992): 179–214.

Pippin, Robert. "Kant on the Spontaneity of Mind." *Canadian Journal of Philosophy* 17, no. 2 (1987): 449–75.

———. *Kant's Theory of Form: An Essay on the Critique of Pure Reason*. New Haven: Yale University Press, 1982.

Pitel, Andrew. "Kant's Critique of Rationalism: Substance, Essence, Noumenon." Ph.D. Dissertation, The University of Chicago, 2021.

Pollok, Konstantin. *Kant's Theory of Normativity*. Cambridge: Cambridge University Press, 2017.

Powell, C. Thomas. *Kant's Theory of Self-Consciousness*. Oxford: Clarendon Press, 1990.

Prauss, Gerold. *Erscheinung Bei Kant: Ein Problem Der Kritik Der Reinen Vernunft*. Berlin: Walter de Gruyter, 1971.

Proops, Ian. "Kant's First Paralogism." *Philosophical Review* 119, no. 4 (2010): 449–95.

Rauscher, Frederick. "The Appendix to the Dialectic and the Canon of Pure Reason: The Positive Role of Reason." In *The Cambridge Companion to Kant's Critique of Pure Reason*, edited by Paul Guyer, 290–309. Cambridge: Cambridge University Press, 2010.

Rawls, John. *Lectures on the History of Moral Philosophy*. Edited by Barbara Herman. Cambridge: Harvard University Press, 2000.

———. "The Categorical Imperative: The First Formulation." In *Lectures on the History of Moral Philosophy*, edited by Barbara Herman, 162–80. Cambridge: Harvard University Press, 2000.

Reath, Andrews. "Formal Principles and the Form of a Law." In *Kant's Critique of Practical Reason: A Critical Guide*, edited by Andrews Reath and Jens Timmermann, 31–54. Cambridge: Cambridge University Press, 2010.

Rohlf, Michael. "The Ideas of Pure Reason." In *The Cambridge Companion to Kant's Critique of Pure Reason*, edited by Paul Guyer, 190–209. Cambridge: Cambridge University Press, 2010.

Rosefeldt, Tobias. "Closing the Gap: A New Answer to an Old Objection against Kant's Argument for Transcendental Idealism." *Proceedings of the Aristotelian Society* 116, no. 2 (2016): 181–203.

———. *Das Logische Ich: Kant Über Den Gehalt Des Begriffes von Sich Selbst*. Berlin: Philo, 2000.

———. "Kants Ich Als Gegenstand." *Deutsche Zeitschrift Für Philosophie* 54, no. 2 (2006): 277–93.

———. "Kant's Logic of Existence." *Journal of the History of Philosophy* 58, no. 3 (2020): 521–48.

———. "Kant's Self: Real Entity and Logical Identity." In *Strawson and Kant*, edited by Hans-Johann Glock. Oxford: Oxford University Press, 2003.

———. "Subjects of Kant's First Paralogism." In *Kant and the Philosophy of Mind*, edited by Anil Gomes and Andrew Stephenson, 221–44. Oxford: Oxford University Press, 2017.

Ryle, Gilbert. *The Concept of Mind*. London: Hutchinson, 1949.

Schafer, Karl. "A System of Rational Faculties: Additive or Transformative?" *European Journal of Philosophy* 29, no. 4 (2021): 918–36.

———. "Kant: Constitutivism as Capacities-First Philosophy." *Philosophical Explorations* 22, no. 2 (2019): 177–93.

———. "Transcendental Philosophy As Capacities-First Philosophy." *Philosophy and Phenomenological Research* 103, no. 3 (2021): 661–86.

Sellars, Wilfrid. ..."...This I or He or It (The Thing) Which Thinks..." *Proceedings and Addresses of the American Philosophical Association* 44 (1970): 5–31.

Serck-Hanssen, Camilla. "Der Nutzen von Illusionen: Ist Die Idee Der Seele Unentbehrlich?" In *Über Den Nutzen von Illusionen: Die Regulativen Ideen in Kants Theoretische Philosophie*, edited by Bernd Dörflinger, 59–70. Hildesheim: Olms, 2011.

Smit, Houston. "Kant on Apriority and the Spontaneity of Cognition." In *Metaphysics and the Good: Themes From the Philosophy of Robert Merrihew Adams*, edited by Samuel Newlands and Larry Jorgensen, 188–251. Oxford: Oxford University Press, 2009.

———. "The Role of Reflection in Kant's Critique of Pure Reason." *Pacific Philosophical Quarterly* 80, no. 2 (1999): 203–23.

Stang, Nicholas. "A Guide to Ground in Kant's Lectures on Metaphysics." In *Kant's Lectures on Metaphysics: A Critical Guide*, edited by Courtney Fugate, 74–101, 2019.

———. *Kant's Modal Metaphysics*. Oxford: Oxford University Press, 2016.

Strawson, Peter. *Individuals: An Essay in Descriptive Metaphysics*. London and New York: Routledge, 1959.

———. *The Bounds of Sense*. London: Methuen, 1966.

Teufel, Thomas. "Kant's Non-Teleological Conception of Purposiveness." *Kant-Studien* 102, no. 2 (2011): 232–52.

Tolley, Clinton. "Kant on the Place of Cognition in the Progression of Our Representations." *Synthese* 197, no. 8 (2020): 3215–44.

———. "The Metaphysics of Powers in Kant and Hegel." In *Powers: A History*, edited by Julia Jorati. Oxford: Oxford University Press, 2021.

Walsh, William. *Kant's Criticism of Metaphysics*. Edinburgh: Edinburgh University Press, 1975.

———. "Philosophy and Psychology in Kant's Critique." *Kant-Studien* 57, no. 1–4 (1966): 186–98.

Warren, Daniel. "Kant on the Substance-Accident Relation and the Thinking Subject." In *Rethinking Kant*, edited by Pablo Muchnik and Oliver Thorndike, 4:35–54. New Castle upon Tyne: Cambridge Scholars Publishing, 2015.

———. *Reality and Impenetrability in Kant's Philosophy of Nature*. New York: Routledge, 2001.

Watkins, Eric. "Kant and the Grounding of Transcendental Idealism." *Studi Kantiani* 32 (2019): 103–18.

———. *Kant and the Metaphysics of Causality*. Cambridge: Cambridge University Press, 2005.

———. "Kant on Infima Species." In *Kant on Laws*, 212–23. Cambridge: Cambridge University Press, 2019.

———. *Kant on Laws*. Cambridge: Cambridge University Press, 2019.

———. "Kant on Materialism." *British Journal for the History of Philosophy* 24, no. 5 (2016): 1035–52.

———. "Kant on Rational Cosmology." In *Kant and the Sciences*, edited by Eric Watkins, 70–89. Oxford: Oxford University Press, 2001.

———. "Kant on Real Conditions." In *Natur Und Freiheit. Akten Des XII. Internationalen Kant-Kongresses*, edited by Violetta Waibel, Margit Ruffing, and David Wagner, 1133–40. Berlin/Boston: De Gruyter, 2018.

Watkins, Eric, and Marcus Willaschek. "Kant's Account of Cognition." *Journal of the History of Philosophy* 55, no. 1 (2017): 83–112.

Waxman, Wayne. *Kant's Model of the Mind: A New Interpretation of Transcendental Idealism*. Oxford: Oxford University Press, 1991.

Willaschek, Marcus. *Kant on the Sources of Metaphysics: The Dialectic of Pure Reason*. Cambridge: Cambridge University Press, 2018.

Williams, Bernard. *Descartes: The Project of Pure Enquiry*. London and New York: Routledge, 1978.

Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. Translated by C.K. Ogden. London and New York: Routledge, 1981.

Wolff, Robert. "Kant's Debt to Hume Via Beattie." *Journal of the History of Ideas* 21, no. 1/4 (1960): 117.

Wood, Allen. *Kant's Rational Theology*. Ithaca: Cornell University Press, 1978.

Wuerth, Julian. *Kant on Mind, Action, and Ethics*. Oxford: Oxford University Press, 2014.

Zöller, Günter. "Main Developments in Recent Scholarship on the Critique of Pure Reason." *Philosophy and Phenomenological Research* 53, no. 2 (1993): 445–66.

Zuckert, Rachel. "Empirical Scientific Investigation and the Ideas of Reason." In *Kant and the Laws of Nature*, edited by Angela Breitenbach and Michela Massimi, 89–107. Cambridge: Cambridge University Press, 2017.

———. *Kant on Beauty and Biology: An Interpretation of the "Critique of Judgment."* Cambridge: Cambridge University Press, 2007.