UNIVERSITY OF CALIFORNIA SAN DIEGO

Spontaneity and Teleology in Kant's Theory of Apperception

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Philosophy

by

Claudi Brink

Committee in charge:

Professor Eric Watkins, Chair
Professor Lucy Allais
Professor Samuel Elgin
Professor Clinton Tolley

2022

The dissertation of Claud Brink is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

For Max and Elena,

I love you

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

All references to Kant's works will cite the *Akademie* edition (1902—) by volume and page number, except references to the first *Critique*, which use the standard A/B pagination. Throughout this dissertation, I have relied upon the translations in the Cambridge edition (1992—) of Kant's works. I use the following abbreviation scheme for ease of reference:

AC: Anthropology Collins

APV: Anthropology from a Pragmatic Point of View

CPJ: Critique of the Power of Judgment

CPr: Critique of Practical Reason

G: Groundwork of the Metaphysics of Morals

JL: Jäsche Logik

MD: Metaphysik Dohna

MFNS: Metaphysical Foundation of Natural Science

ML1: Metaphysik L1

ML2: Metaphysik L2

MM: Metaphysik Mongrovius

MV: Metaphysik Vigilantius

MVM: Metaphysik Volkmann

OP: Opus Postumum

Prol: Prolegomena to Any Future Metaphysics That Will be Able to Present Itself as a Science

R: Reflections

WRP: What Real Progress Has Metaphysics Made in Germany since the Time of Leibniz and Wolff?

ACKNOWLEDGEMENTS

I have been incredibly fortunate to work on this dissertation over many years at UCSD with the support of brilliant Kant scholars, excellent philosophers, and wonderful friends.

I am grateful to my committee for their support, patience, time, and advice over the many years that I have worked on this project.

I would like to thank Sam Elgin for agreeing to join my committee at a very late stage. I am very grateful, and I look forward to continued productive philosophical exchanges in the future.

I must thank Lucy Allais for first instilling a love of Kant's philosophy many years ago when I first started my post-graduate work at the University of the Witwatersrand. She has been a part of my journey ever since then, and her encouragement and support throughout have been a constant source of motivation. Without Lucy, I would not have had the confidence to continue my graduate education. She is an outstanding Kant scholar and an inspiring philosopher, and it has been a great privilege to work closely with her over the course of my entire graduate career.

I have benefitted a great deal from attending seminars and taking part in discussion with Clinton Tolley. His approach to philosophy, his intellectual curiosity, rigor and expansive interests have been a wonderful source of inspiration. His work has been immensely helpful to me, and I am grateful for the many stimulating (if slightly disorienting!) philosophical discussions we've had since I joined the department at UCSD.

I owe a great debt to my chair, Eric Watkins. I have benefitted from having him as my advisor in more ways than I could have anticipated. He is an exceptional mentor, and my engagements with him allowed me to flourish both intellectually and personally. Despite holding me to standards that I often feared were too high, his feedback was always balanced and practical.

His kind, encouraging and forward-looking engagements with my work have helped me both to build confidence as a Kant scholar and philosopher, and to push myself to meet higher standards than I thought I was capable of.

Aside from the many qualities that make him a great mentor, I also benefitted from his deep and nuanced understanding of Kant, as well as his philosophical acumen and rigor. He was always about a hundred steps ahead of me, and could foresee difficulties that my view would face way down the line. Usually, this foresight came well before I myself was even aware of the position I was committed to, let alone the implications it would have downstream. I am extremely grateful for his guidance, his time, and his patience, for all of the feedback, and for the seriousness with which he has engaged with me and with my work. This project would have been vastly inferior in quality without his input and support.

I started the work that forms the basis of Chapters 3 and 5 in Tübingen while on a scholarship from the DAAD. There I had the opportunity to work with Professor Ulrich Schlösser at the University of Tübingen. The philosophy department in Tübingen was a wonderful place to work and study, and I am grateful to professor Schlösser for several very stimulating conversations about my ideas.

I was able to further improve Chapter 3 on the basis of a very rich and helpful discussion of my arguments with Béatrice Longuenesse and Hannah Ginsborg, who both kindly agreed to be commentators on my paper for a presentation that formed part of a Virtual NAKS series in 2021. I would like to thank them both for their generous and insightful discussion of my work. I would also like to thank Patricia Kitcher, who commented on an early version of Chapter 4. Her feedback was immensely helpful in developing the broader framework of that chapter. I am grateful to Michael Oberst, who gave me helpful comments on Chapters 3 and 4. Finally, all of my chapters

benefited in one way or another from discussions with Brian Tracz over the course of many years. His continued feedback on my work has made a lasting impression on my thinking about Kant and philosophy.

Over the many years in I spent in graduate education I have formed deep and meaningful friendships, and all of these had an impact, in one way or another, on my intellectual and philosophical development, but also in many ways on my personal well-being. I am grateful for the many conversations, experiences, and philosophical insights that I got to share with Sebastian Speitel, JP Messina, Brian Tracz, Julie Reid, Chris Fisher, Kathleen Connelly, Lars Lodberg, Irina Schumski, JC Gonzalez, Andy Stephenson, and Adam Parkins. My work, and my life, would be much poorer without all of you.

I would like to thank my family for the many sacrifices, great and small, that they have made to support me in the course of my studies. Among these sacrifices, the most significant one, which I have shared with them, is the physical distance that separates us, and that has, at times, been incredibly painful and regrettable. I would like to thank my sister, Roelien, for always renewing her efforts to bring me into her life-world and for endeavoring to share so many of the beautiful moments that I missed of my nephew's growing up. I would like to thank my mother, Lenie, for encouraging me and for supporting me despite the hardship of being separated from me and more recently from her granddaughter. She has and continues to be an inspiration to me in so many ways, but above all, I admire her ability to face and overcome adversity. I would like thank my father, Ronnie, who instilled a strong sense of responsibility in me, and whose strict work-ethic has taught me a great deal about commitment and determination. I would like to thank my brother, André, for letting me share many of my most outrageous philosophical ideas, and for his powerful scientific counterpoints. I have a particularly fond memory of teaching him Kant's first

Critique over a decade ago in my bedroom when he was in high school; I love that we have continued to discover ways of sharing our intellectual interests.

Finally, and most significantly, I would like to thank my husband, Max Edwards. There is quite simply no person to whom I am more indebted personally and intellectually than I am to him. He gave me feedback on every chapter, every section and every paragraph of this dissertation many times over. He discussed every idea that is present in this work from its inception to its articulation and refinement. He struggled with me to help bring this work to completion and he celebrated every breakthrough along the way with me. He is and continues to be enthusiastically supportive of my work, and his excitement and genuine investment in it are the ground of a self-belief and confidence that I never had before I met him. He is an excellent Kant scholar, a steadfast partner, and an exceptional person, and I cannot conceive what this project, and indeed my life would be without him.

There is a deeper and more fundamental way in which I am grateful to Max. He has been the most wonderful, loving, and attentive parent to our daughter, Elena. Over the course of the last year, especially, as I was working to complete my dissertation, and the stress of finding reliable and good quality childcare became increasingly difficult to navigate, Max stepped in and took care of Elena with more love and devotion than I thought any person capable of. It was wonderful to behold and inspiring to be part of. In the midst of a pandemic, isolated from family and friends, and increasingly concerned about falling behind with his own work, Max never once failed to provide Elena with the love and care that she needed to flourish. There are no words to express how grateful I am to him for that.

And finally, let me thank my wonderful little daughter, Elena, because her presence in my life gave me the perspective that I didn't know I needed to complete this work.

VITA

2010	B.A. (Hons) Philosophy, University of the Witwatersrand
2012	M.A. Philosophy, University of the Witwatersrand
2013 – 2022	Teaching Assistant, University of California San Diego
2022	PhD in Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Spontaneity and Teleology in Kant's Theory of Apperception

by

Claudi Brink

Doctor of Philosophy in Philosophy

University of California San Diego, 2022

Professor Eric Watkins, Chair

In this dissertation, I argue for what I call a *formalist* interpretation of Kant's doctrine of transcendental apperception. On this interpretation, transcendental apperception is not a representation of the subject, or any aspect of a subject, such as her internal activities or states.

Indeed, I maintain that transcendental apperception should not be understood in representational terms at all. Instead, I argue, transcendental apperception must be conceived of as an end-directed *activity* of a subject that is the formal condition of a particular kind of representation, thought. On the view I develop, Kant identifies transcendental apperception with a spontaneous activity of the subject, which functions as an inner organizational principle through which representations are combined in such a way that they have a particular kind of unity, a unity that Kant calls 'formal' or 'logical'. It is this kind of unity that a representation must have in order to inhere in a subject not just as a sensible representation, but as an intellectual representation, which, for finite beings like us, are thoughts.

In developing this account, I draw on Kant's faculty psychology and the underlying metaphysics of causal powers that he inherits from the rationalist tradition, to show that his conception of the understanding as a *spontaneous* faculty signals that he conceives of its operations in *teleological* terms. This teleological dimension of the understanding helps us to articulate an account of transcendental apperception that breaks from the two dominant approaches in the literature, what I will call the epistemic and psychological approaches. On the epistemic reading, transcendental apperception is a kind of self-consciousness that grounds our cognitive agency, and on the psychological reading, transcendental apperception is a mechanism that functions causally to connect representations to produce judgments. The former account, I argue, is flawed insofar as it introduces normative and epistemic considerations into Kant's account of the understanding that are incompatible with key aspects of his faculty psychology; and the latter is flawed because it fails to situate the understanding within the teleological explanatory framework signaled by Kant's claim that the understanding is spontaneous.

INTRODUCTION

Situating Kant's Theory of Apperception

Prior to philosophical reflection, nothing is more obvious and unremarkable than our consciousness of the world around us and of ourselves within it. It is this consciousness of ourselves that constitutes our own unique perspective on the world, a perspective that is pervasive and inescapable: it colors every thought, every perception and every feeling as belonging to *me*. While it might seem like a quotidian detail of our human existence, developing a satisfactory theoretical account of consciousness is at the center of philosophical inquiry into the nature of the mind. Within this branch of inquiry, a key question to be settled concerns the relationship between self-consciousness and our conscious experience of the world. On the one hand, one might think that self-consciousness presupposes consciousness of objects in the world, and is a manifestation of exactly this mode of object-consciousness, but trained on the subject. On the other hand, one might think that self-consciousness is more basic than and presupposed by all of my conscious experience, insofar as I at least implicitly recognize that all of my experiences are *mine*. But if we are persuaded by this second thought, we are confronted with a pressing philosophical task. If self-consciousness is prior to object-consciousness, then we must find a way of giving a positive

characterization of self-consciousness without drawing on the explanatory resources that inform our account of object-consciousness.

In this dissertation, I will argue that Immanuel Kant's theory *of transcendental apperception* gives us the resources to articulate just such an account. Famously, in the second edition of the Transcendental Deduction, Kant claims that 'the 'I think' must be able to accompany all my representations' (B132). The act of the subject that is expressed by the 'I think' has a unity that he calls 'the transcendental unity of apperception'; and this unity, which, Kant says, "precedes a priori all my determinate thinking" (B134), is a condition of the possibility of object-consciousness, for it is only when the "the manifold of given representation" (B135) is subjected to this unity that it can become cognition or objective representation (B139). My aim in this dissertation is to understand these claims, and to show that, properly understood, they express a philosophically satisfying theory of self-consciousness that explains how self-consciousness could lie at the basis of our awareness of the world around us.

In pursuing this aim, I develop a novel interpretation of Kant's theory of apperception that departs from orthodox readings in several important respects. The key to understanding Kant's doctrine of apperception, I will argue, is to take seriously the 'faculty psychology' that forms the starting point for Kant's theorizing about the mind and its various cognitive faculties. Integrating his theory of apperception into this context, I will show, enables us to make sense of Kant's claim that apperception is an 'act' of the 'self-activity' or 'spontaneity' of the subject (B130), and is "the ground of… the possibility" (B131) of the higher faculty of cognition that Kant calls the understanding.

I call the reading that I develop through this interpretive strategy the *formalist* interpretation. On the formalist reading, apperception is not consciousness of any object; indeed,

I argue that transcendental apperception does not have an intentional structure at all. Instead, the formalist view holds that transcendental self-consciousness is a formal principle that confers on representations a certain kind of unity, and in so doing makes thought possible.

To fully understand how it is that transcendental apperception makes thought possible, I will show, we need to recognize a *teleological* dimension to Kant's account of the understanding that has not been fully appreciated in the literature.[1] This teleological element, I will argue, is implicit in Kant's famous characterization of the understanding as a 'spontaneous' faculty of the mind. It is only by appreciating the metaphysics of causal powers that underlies Kant's faculty psychology, I will show, that we can understand the doctrine of spontaneity, and once we do, we will see that the operations of the understanding cannot be explained mechanistically. Recognizing this teleological dimension enables us to flesh out the formalist proposal because it allows us to see how it is possible for apperception to bring the unity that is required for thinking into our representations. Moreover, it is precisely by bringing about this unity that the act of transcendental apperception grounds the understanding as a power of thinking and cognition.

In defending this view, I depart from two dominant rival interpretive approaches to Kant's account of apperception and the accompanying doctrine of spontaneity, which I call the *epistemic* and *psychological* approaches. Both of these approaches, I argue, rest their interpretations on a problematic conception of how the understanding functions. Against the psychological approach, I deny that the understanding can be conceived as functioning in accordance with the same laws that govern natural phenomena; against the epistemic approach, I deny that we can make sense of

[1]A handful of other commentators have also stressed the teleological dimension of Kant's theory of the understanding, including Dörflinger (1996), Fugate (2014), Longuenesse (1998), and, more recently, Schafer (2021). My arguments in this dissertation are sympathetic to the spirit of this modest interpretive tradition; where I go beyond these commentators is in grounding my claims about the teleology of the understanding on a thorough account of the metaphysics of powers that underlies Kant's faculty psychology.

the functioning of the understanding in normative epistemic terms. Neither of these approaches, I argue, is able to locate the unique role that apperception is meant to play in Kant's theory of the understanding. On the view that I develop, it is only if we recognize that the understanding functions in accordance with an end or goal that is essential to its nature as the faculty of thinking that we can understand the role of apperception in making thought possible. And as we will see, spelling out this proposal requires that we depart from both a mechanistic and a normative conception of the understanding.

In order to better clarify the interpretive approach I will take, the positive claims I will make, and the way in which my positive proposals cut against entrenched orthodoxies within the secondary literature, the remainder of this introduction will be structured as follows. First, in Section 1, I present the fundamentals of Kant's faculty psychology and the metaphysics of powers that he inherits from his rationalist predecessors. With an understanding of this intellectual background, I will go on, in Section 2, to review the dominant strands in the secondary literature on Kant's theory of apperception. In this section, I characterize the epistemic and psychological interpretations in detail and locate their fundamental points of disagreement. Understanding the divergences between these two views will create a framework in which I am able to situate my own view and contribution more clearly, and I close the section by explaining how the formalist interpretation is related to both its psychological and epistemic competitors. Finally, in Section 3, I outline the individual chapters of this dissertation with a view both to previewing the content of those chapters and also to helping the reader understand exactly how my interpretive approach, with its appeal to Kant's faculty psychology, underlies the various contentions I will be arguing for.

Section 1: Kant's Faculty Psychology

In the present section, I explain the historical context and conceptual framework that inform Kant's faculty psychology. In drawing on Kant's faculty psychology and the broader metaphysical framework that he inherits from the rationalist tradition, this study takes its cue from several important recent contributions to the study of Kant's critical philosophy. As a number of important scholars have recognized, seeing Kant as continuing in the philosophical tradition of his rationalist predecessors is immensely fruitful for understanding central aspects of his critical enterprise. Despite the fact that Kant's critical work is often portrayed – and lauded – as a wholesale rejection of rationalist metaphysics, we risk not only misconstruing Kant's position, but also concealing the philosophical insights of his work, if we approach it too narrowly from a contemporary perspective. In the spirit of correcting the bias toward approaching Kant's texts intent on determining the "contemporary viability of specific assertions", Karl Ameriks argues, in his groundbreaking monograph, *Kant's Theory of the Mind*, that "the Critique is much more traditional and rationalistic than is generally recognized" (Ameriks 2000, 4). Ameriks focusses primarily on Kant's criticisms of rational psychology in the Paralogisms, with the aim of providing a close reading of these arguments that is informed by Kant's incorporation of various fundamental commitments of the rationalist tradition into his own metaphysical views. On Ameriks' view, we can only fully appreciate the significant heritage of rationalist metaphysics in Kant's thinking by studying the extensive notes on the lectures that Kant gave on the topic throughout his career: "I shall go as far as to argue that only in light of these discussions are his positions [in the Paralogisms] truly comprehensible" (2000, 3).

Since the original publication of Ameriks' monograph, several subsequent studies, building on his fundamental strategy, have extended this historically informed approach to make sense of

other key critical doctrines. Perhaps most notably, in *Kant and the Metaphysics of Causation*, Eric Watkins undertakes to "tell a far different … and more satisfying narrative about Kant's account of causality … by understand[ing] Kant's views and arguments within their proper historical context" (2004, 5). We can tell such a story, Watkins shows, by broadening our perspective beyond our contemporary philosophical preoccupations and approaches and by incorporating textual resources from the entirety of Kant's corpus, including the transcripts of his metaphysics lectures.

Several further studies are worth mentioning in this connection. In his *Kant and Rational Psychology*, Corey Dyck renews Ameriks' efforts to make sense of the Paralogisms by drawing especially on the conception of causal powers in the German Wolffian tradition. A further application of this approach is taken more recently by Nicholas Stang in his attempt to develop a Kantian account of the metaphysics of modality.[2] My hope is to contribute something towards this interpretive approach by extending it to Kant's account of the faculty of understanding, and, more specifically, his theory of apperception. In order to explain the interpretation that results from taking this approach, it will be helpful to present the fundamental commitments of the rationalist metaphysical tradition that informs Kant's faculty psychology.

The historical tradition in which Kant is writing conceives of the mind (or soul) as a *substance* that is characterized by its possession of certain *powers* or *faculties*.[3] These powers are

[2] For articles that make substantive appeal to the rationalist metaphysics of powers, see also Oberst (2019), Smit (2009), and Stratmann (2018). Another important study to mention in relation to this interpretive tradition is Falk Wunderlich's (2011) study of Kant's conception of consciousness, which traces the development of the theory of consciousness through the German rationalist tradition following Wolff and leading up to Kant, with particular emphasis on the development of a shared philosophical terminology for theorizing about consciousness. I do not include Wunderlich's study with the other studies mentioned above because while his work takes very seriously Kant's intellectual relation to the rationalist tradition, he does not focus specifically on the ways in which rationalist metaphysical views about substance and power shape Kant's position.

[3]Note that while these two terms are importantly distinct, I will use them interchangeably for the time being. We will be in a position to clarify the way in which the tradition understands the distinction between faculty and power shortly.

assigned an important explanatory role, for it is in virtue of having these powers that the soul enters into and sustains its characteristic states – representing, feeling, desiring, willing, perceiving, and so on. All substances, on this view, are causally active insofar as their powers serve to bring about the realization of certain accidents or properties. Despite significant divergences among Kant's rationalist predecessors and contemporaries, this basic metaphysical framework is fundamental to theorizing within the Leibniz-Wollfian tradition, both about mind and body or material nature. As we will see, though Kant is more cautious than others about the kinds of substantive metaphysical claims we can make about the ultimate nature of substance, he nevertheless accepts this metaphysical framework as defining the starting point of philosophical enquiry. To enquire into material nature is, ultimately, to enquire into the nature of material substance and its governing powers, and to enquire into the mind is to enquire into the nature of the representational faculties that belong to it. Kant's philosophy of mind is thus articulated within the framework of what is usefully referred to as a 'faculty psychology'.

To understand Kant's faculty psychology and to see how it fits into the broader framework of the Leibniz-Wolffian metaphysical tradition, it would be helpful to provide a brief overview of some of the key concepts and terminology used within this tradition. We can then see more clearly the extent to which Kant incorporates (and modifies) elements of that framework into his own system. To see the outlines of this tradition in a form most germane to understanding Kant's engagement with it, it is helpful to consult Alexander Baumgarten's *Metaphysica*, a textbook account of this metaphysical tradition that Kant himself used throughout his career as the basis for his lecture courses.

Baumgarten begins the *Metaphysica* by laying out a basic ontology, which he characterizes as "a science of the most general predicates of being" (M §4).[4] The idea here is to abstract entirely from the kind of being under consideration and to consider simply being in general by determining the predicates that apply to it as such. These predicates, as the most general predicates of being as such, apply to minds or souls as well as to bodies and their interactions as part of material nature. For our purposes, it is especially important to review just one pair of predicates that is meant to apply disjunctively to all being, namely, *substance* and *accident*. Baumgarten introduces the distinction as follows:

> A being either cannot exist except as a determination of another (in something else), or it can (§ 10). The former is an accident …, and the latter is a substance … because it can exist, although it is neither in something else, nor the determination of something else. (*M* § 191)

Here, Baumgarten claims that everything that exists is either a substance or an accident. It is an accident if its existence depends in some way on another being, which being, if its existence is not again dependent on anything further, is a substance. On the basis of this general distinction of all being into substances and accidents, Baumgarten now introduces several important further concepts, including that of *power*, *faculty*, *receptivity*, *activity*, *suffering*, *nature*, and *principles of alteration*, which clarify the nature of the relationship between substance and accident.

At the center of this cluster of further concepts is the concept of a *power*, which serves to explain the dependence of an accident on a substance. Baumgarten says that an accident *inheres* in a substance (*M* § 192), and given that it inheres in a substance, the substance must contain "some ground of inherence" (*M* § 193). This ground explains why the accident belongs to the substance in question and so constitutes the reason for the inherence of the accident in the substance. Now,

[4] In what follows I will abbreviate references to Baumgarten's Metaphysica with 'M'.

it is this very ground that Baumgarten identifies as a *power* (*M* § 193).[5] The framework I want to sketch emerges from a more careful consideration of how these powers function.

First, we can note that there is a distinction between a *power* and a *faculty*, according to which a faculty is a dormant power whose exercise would suffice to ground a particular accident (*M* § 216). When a power of a substance is in fact exercised, the substance is said to *act*, and it acts insofar as it either sustains or alters the state that it is in. Baumgarten thus defines *action* as "an alteration of state and in general the actualization of an accident in a substance through its own power" (*M* § 210). A faculty is thus identified as "the possibility … of acting" (*M* § 216). On Baumgarten's view, *all* existing substances act, and therefore all substances have faculties. But substances can also *suffer*. When a substance suffers, on Baumgarten's account, it is not acting (and therefore not exercising any of its own powers) in the sense described above, because it is not actualizing an accident through its own power, but rather through the exercise of an external power. The possibility of acting, which we have seen is called the *faculty* of a substance, is therefore contrasted with the possibility of suffering, which in turn is called *receptivity* (*M* § 216). This distinction between faculty and receptivity, and the corresponding distinction between activity and suffering, tracks two ways in which accidents can be actualized in a substance. If an accident is actualized through the substance's exercise of its own internal power, then the substance acts and has a corresponding faculty for acting in this way. If the accident is actualized through the exercise of some external power, in contrast, then the substance suffers and has a corresponding receptivity for being acted on in this way.

We can now see that all of the possible ways that substances can exist, the states that they are and can be in, are determined by the faculties and receptivities that they possess, and

[5] As an aside: Baumgarten goes on to identify power and substance (cf. *M* § 198).

Baumgarten claims that these together constitute the *nature* of a substance. To see this, we must note first that the internal and external powers that bring about the alterations in a substance are called *principles*, insofar as those powers are the "ground[s] of the existence" (*M* § 307) of the accidents that inhere in the substance. These principles together constitute the nature of a being:

> The nature (cf. § 431, § 466) of a being is the collection of those of its internal determinations that are the principles of its alterations, or, in general, of the accidents inherent to it. Hence, to the nature of being pertains its … faculties, receptivities, and all the powers with which it is equipped (*M* § 197).

Since the nature of a substance consists of the inner principles in accordance with which changes of its state takes place, it follows that the complete set of possible accidents that could inhere in a substance is determined by its essence or nature. These inner principles are therefore expressed through its action, in the exercise of its power, and, likewise, through its suffering, in the affection of its receptivity.

One final piece of terminology that is important for our purposes is the term *spontaneity*. On Baumgarten's view, spontaneity is a feature of an action when that action "depends on a sufficient principle that is inside the agent" (*M* § 704).[6] Now, we have seen already that all action, on Baumgarten's view, must be the result of an inner principle of a substance that is sufficient to actualize an accident since otherwise the substance cannot be said to *act* (instead, it would suffer). Accordingly, he says that "every action properly as such depends on a principle that is inside the agent" (*M* § 704). Therefore, he holds that every action (properly as such) is spontaneous.

[6] Leibniz and Wolff share this conception of spontaneity, and as we will see shortly, so does Kant. Here is Wolff (quoted in Baumgarten, 2013): "Spontaneity is the intrinsic principle of determining oneself in acting. And actions are called spontaneous insofar as the agent determines these through a principle intrinsic to oneself without an extrinsic determining principle" (*DM* § 933). And here is Leibniz: "The spontaneity of our actions can therefore no longer be questioned; and Aristotle has defined it well, saying that an action is spontaneous when its source is in him who acts" (1952, § 301).

Within this general framework, we then find that the soul is considered a kind of substance, which is distinctive in that "it is something in a being that can be conscious of something" (*M* § 504). On Baumgarten's view, the soul therefore *is* a power that grounds certain accidents in itself, namely, thoughts:[7]

> I think; my soul is altered (§ 125, §504). Therefore, thoughts are accidents of my soul (§ 210), at least some of which have a sufficient ground in my soul.' (§ 21) Therefore, my soul is a power. (*M* § 505)

Unsurprisingly, given their status as ontological predicates that characterize all being as such, we see Baumgarten applying the concepts of substance, power, and accident in his characterization of the soul. This basic metaphysical framework of power, faculty, action and suffering is therefore meant to apply to the explanation and functioning of our minds as much as it is meant to apply to bodies and their interactions.

Now, the critical Kant accepts this basic ontological framework both as a basis for metaphysical theorizing, and also as a basis for theorizing about the nature of the mind and its constituent faculties. While he makes several important adjustments to the traditional framework, he uses all of the above terminology throughout his critical writings and never indicates a wholesale rejection of the fundamental ontological framework, even when he is otherwise critical of his rationalist predecessors. For the purpose of considering how this framework might inform our interpretation of some of the key doctrines in the Analytic, we can start by determining how Kant incorporates (and modifies) key aspects of it into his philosophy by drawing from the

[7] In general, Baumgarten identifies power with the substance. Now, given that he is also committed to the claim that there are several faculties and receptivities that determine the nature of a substance, it must be that he takes there to be a single *dominant* power to which these faculties and receptivities can all be reduced. Kant rejects this claim.

transcripts of his lectures delivered immediately after the publication of the first edition of the Critique.[8]

Kant both takes over and modifies the traditional conception of a power. Throughout the transcripts, Kant signals that he is in agreement with the traditional conception of a power as a ground of the inherence of an accident in a substance. However, whereas Baumgarten and Wolff take the powers of a substance to be reducible to one fundamental power, and then identify this fundamental power with the substance itself, Kant insists that power is not to be identified either with substance or with accident.[9] Instead, he takes a power to be a *relation* between the substance and the accident, which, therefore, is not reducible to either: "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" (*MM* 29:771).

Despite his slightly heterodox conception of powers, Kant's account of substance and accident is squarely in keeping with the tradition. Kant claims that a substance is an existing thing to which we ascribe certain properties depending on the state that it is in. These properties are what Kant refers to as 'accidents', and Kant accepts the traditional view that accidents depend on substances for their existence. Accordingly, he says that "accidents cannot exist other than in the substance" (*MM* 29:796), and that substances are what remain even when we "leave aside all accidents" (*MM* 29:771). Kant also carries forward a standard conception of the nature or essence of a substance, though he embeds this conception within his slightly idiosyncratic account of power. We have seen that a power is the relation of the accident to the substance insofar as the power is the ground of the inherence of the accident. Kant also takes it to be the case that a substance *acts* insofar as it grounds an accident in a substance through a power: "substance, insofar

[8] For my purposes here I draw primarily from the Mrongovius lecture transcript, which is dated 1782-3.
[9] Cf. Watkins (2005, 261 – 262) and Dyck (2014, 200 – 205)

as its accidents inhere, is in action, and it acts insofar as it is the ground of the actuality of the accidents…" (*MM* 29:773). So, the states of a substance are determined by the various acts through which it exercises its power(s), since it is through the exercise of these powers that accidents inhere in it. The acts that ground accidents in a substance are themselves *internal principles* that constitute the nature of the substance, a view that is again in keeping with the traditional one outlined above (cf. *MM* 29:935, *MFNS* 4: 467).

We see, however, an important departure from Wolff and Baumgarten in Kant's conception of activity and passivity: whereas *activity* and *passivity* (or suffering) are mutually exclusive concepts for Baumgarten, they are not so for Kant. We saw above that Baumgarten contrasts activity with suffering. According to this view, a substance acts if it actualizes an accident through one of its own internal powers, but it suffers when an accident is actualized through an external power acting on it. Since Baumgarten then also identifies the spontaneity of an action as the actualization of an accident in a substance through its internal powers, all actions on his view are spontaneous. Now, as we will see in much more detail later on (Chapter 3), Kant revises this account of activity, spontaneity, and suffering. On Kant's view, activity and passivity are not mutually exclusive concepts. Kant agrees with Baumgarten that a substance "acts … insofar as its accidents inhere, and … insofar as it is the ground of the actuality of the accidents" (*MM* 29:773), but he does not contrast suffering or passivity with this conception of activity. Instead, Kant identifies suffering as a *kind of* activity, distinguishing passivity or suffering, on the one hand, and self-activity or spontaneity, on the other, as different species within a common genus. Corresponding to this difference is a different understanding of the extension of the concept of spontaneity. On Baumgarten's conception, we saw that all action is spontaneous, but, as we will see, this is not Kant's view. Although he is in broad agreement with the tradition's conception of

spontaneity as a determination of the state of a substance in accordance with an inner principle (cf. *ML1* 28:285, *MVM* 28:448; *MFNS* 4:544), and of suffering or passivity as a determination of a substance in accordance with an external principle (*MM* 29:823), he does not have to deny that suffering is a kind of activity and nor, therefore, does he need to claim that all activity is spontaneous.

Now, along with his rationalist predecessors – who, we saw, apply this general ontological framework in their treatment of the soul – Kant conceives of the soul or the mind as a substance with a number of different *faculties* or *powers*, and he characterizes these faculties by investigating the distinctive *acts* that actualize certain states in accordance with a priori *principles*. To see how thoroughly this traditional framework informs Kant's critical project, we merely have to note that the three key critical texts, the *Critique of Pure Reason*, the *Critique of Practical Reason* and the *Critique of the Power of Judgment* are each an investigation and analysis of one of the three faculties into which Kant takes the human mind as a whole to be divided:

1. The faculty of cognition
2. Feeling of pleasure and displeasure
3. Faculty of desire (cf. *CPJ* 20:245, *CPJ* 5:198)

Each of these faculties is again composed of sub-faculties, whose characteristic *activities* account for the accidents that inhere in the human soul (sensing, perceiving, cognizing, thinking, desiring, feeling, and so on). Kant moreover takes these acts to be determined by certain *principles*. And he distinguishes these faculties into lower *receptive* capacities and higher *spontaneous* faculties, depending on whether they contribute to the various changes in the soul through the affection of objects (or powers) outside of it, or from within.

Now, my focus in this dissertation is specifically on the first of the faculties listed above, and, even more specifically, on the spontaneous sub-faculty of the faculty of cognition, the faculty of understanding. It is important to note that Kant distinguishes between a broad and a narrow sense of the term 'understanding' [*Verstand*] (A130-31/B169), and it is the understanding in the narrow sense of the term that will be my focus. In the broad sense, the understanding incorporates all of the higher powers of cognition, which includes the understanding (in the narrow sense), the power of judgment, and reason. In the narrow sense, the term 'understanding' refers to one of the faculties that compose the higher faculty of cognition, and which Kant more specifically identifies as a faculty for cognition, a faculty for judgment, and a faculty for thinking (A127). This faculty, too, is spontaneous, insofar as it "bring[s] forth representations" from itself (A51/B75), and its spontaneity is due to a fundamental *act* which Kant calls pure apperception or transcendental self-consciousness (B130).

In this dissertation, my aim is to bring to bear Kant's faculty psychology and the broader ontological framework outlined above on our conception of how the understanding functions and the respects in which its characteristic function is grounded on the act of apperception. In doing so, I will be arguing for three main claims: (1) Pure apperception is not *reflexive*. It is not a consciousness that takes as its object either the subject, the activity of the subject, or the corresponding state of the subject. As we will see, it is not consciousness of any object at all, but only the *formal* condition of the representation of any and all objects, and it is the act that *grounds* representations as thoughts in the subject. (2) We should not conceive of the rules and laws of the understanding as normative. Insofar as they derive from the nature of the subject as a thinking being, these laws are to be regarded as constitutive of the subject's faculty of thinking, and considered only as such, have no normative dimension. (3) The understanding's function cannot

be understood in mechanistic terms. Its functioning is not to be explained in terms of the causal order that governs material nature that Kant identifies as the whole of appearance. Instead, the understanding must be conceived of as functioning teleologically, and, I will argue, must be understood as operating for a purpose or a goal that is constitutive of its nature. In every case, the key to understanding the ground of my contention will be to appreciate the ways in which Kant's faculty psychology shapes his philosophy of mind.

Section 2: Overview of the Literature

The three claims to which I briefly alluded above define a middle-ground between the two major approaches taken in the literature on Kant's theory of apperception. In this section, I want to provide an overview of each approach in order to locate their points of substantive disagreement; this will allow me to chart the ways in which I see my own position as steering a course between them.

The literature on Kant's theory of apperception and its role in his broader account of the understanding is dominated by a dispute between *epistemic* and *psychological* approaches. The former approach is famously taken by Peter Strawson (1966), whose 'austere' interpretation of the Analytic aims to remove from it any claims about the psychological processes that underlie cognition, for Strawson thinks that such claims violate Kant's doctrine that we do not have knowledge of things in themselves. This tradition is carried forward by numerous major commentators, perhaps most notably by Henry Allison, although in less 'austere' form. On Allison's reading (2004), Kant's project in the first *Critique* is to articulate the 'epistemic conditions' on human cognition, where these conditions are distinct from both psychological and ontological conditions. While Allison does not rule out transcendental psychology wholesale, as

Strawson does, he does not read the Analytic as a project that describes the various psychological processes that generate cognition, but instead as an account of the normative constraints that govern cognition. On Allison's view, "it is absolutely essential to keep firmly in mind the normative concern of the Deduction" (Allison 1996, 58), and this normative concern, he holds, is clearly on display when we consider the epistemic significance of the states that are generated through the functions of the understanding.

Opposed to this epistemic approach is one that finds in Kant a precursor to contemporary functionalist accounts of the mind. On this kind of view, associated most famously with Patricia Kitcher (1990), Kant's conception of the mind and of the understanding can be understood primarily in psychological terms, that is, in terms of the psychological processes that generate cognition. This approach seeks to offer an account of Kant's project that is particularly appealing from the point of view of contemporary cognitive science and psychology in that it does not see Kant as essentially committed to a conception of the mind that cannot be studied, as it is in these disciplines, by observing how the mind functions empirically. Accordingly, *psychological* approaches such as Kitcher's tend to conceive of the workings of the understanding in causal terms, and of the subject that Kant discusses throughout the Analytic as an empirical subject. We see both of these commitments very clearly in the following passage from Kitcher's early work:

> [T]he doctrine of apperception must present a phenomenal, if highly abstract, aspect of the self. If Kant is right that anything of which we can have knowledge must be governed by causal laws, then synthetic connection will be lawful ... The I that thinks will be phenomenal and causally determined… The doctrine of apperception can only be phenomenal, and so we might as well admit this fact, whatever the fallout of this doctrine in the later Critiques. (1990: 139-40)

In this passage, Kitcher embraces the view that Kant's theory of apperception is ultimately an account of the causal processes that govern the phenomenal subject's activities of synthesis.

Against this, epistemic readers have suggested that while it is true that Kant could not be making substantive claims about the nature of the noumenal subject, he should also not be read as making claims about the empirical subject, since any such claims would presuppose the very conditions on cognition that he is at pains to articulate by appeal to the transcendental subject.[10] Instead of rendering Kant's claims in the Analytic as claims about either the noumenal or the empirical subject, the epistemic approach reads Kant as making claims about a formal logical subject that cannot be identified with either the noumenal or the phenomenal self. One way of pursuing this suggestion is to resist assimilating the subject into the causal deterministic system of laws that govern nature, and instead to insist, in Davidsonian spirit, that making sense of our thinking requires its own set of laws that are responsive to a certain ideal of rationality. This commits Kant to making claims about the formal conditions and nature of thought that are independent of any empirical facts about the thinker and that do not entail metaphysical claims about the noumenal subject.

My aim in this dissertation is to develop an interpretation of Kant's theory of apperception that mediates between these two interpretive approaches, drawing insights from both. Both psychological and epistemic readers make important contributions to our understanding of Kant's theory of apperception, and I argue that these contributions can be combined through a careful consideration of the faculty psychology that is the fundamental starting point of Kant's theorizing about the mind. In order to bring out the elements of each view that I will carry forward into my own positive account, it will be beneficial to have a deeper sense of the positive commitments of these views as well as the disagreements between them. As we will see, the disagreement between

[10] Here is Strawson articulating this worry: "The theory of synthesis, like any essay in transcendental psychology, is exposed to the *ad hominem* objection that we can claim no empirical knowledge of its truth; for this would be to claim empirical knowledge of the occurrence of that which is held to be the antecedent condition of empirical knowledge" (1966, 32).

these two approaches emerges most clearly in the way they account for Kant's doctrine that the understanding is a *spontaneous* faculty. As a result of this fundamental disagreement, they have different conceptions of (1) the structure of the act of apperception itself, (2) the status of the laws of the understanding in relation to this act, and (3) the causal order, if any, to which the subject Kant analyzes in the Analytic belongs.

We can start by developing our characterization of the psychological approach. In an effort to recover what Strawson (1966) dismissively refers to as the "imaginary subject of transcendental psychology", there has been a sustained interest in developing an innocuous account of the central psychological features of Kant's doctrine of the understanding as it is developed in the Transcendental Analytic of the first *Critique*. The aim of the commentators who motivated this revival in Kant's psychology is twofold: First, to show that Kant's account can be made compatible with his transcendental idealism and, in particular, with the restrictions that such an account place on cognition; and second, to show that Kant's theory is of interest to contemporary philosophers of mind who want to develop an account of the nature of the mental that is broadly naturalistic. At the forefront of this effort is the work of Patricia Kitcher, whose early work especially focusses on developing a functionalist interpretation of Kant's account of the understanding that holds the promise of removing any suspect metaphysical commitments that might violate the restrictions on our ability to cognize things beyond experience as they are 'in themselves', and that, at the same time, might appear suspect from the naturalistic starting point adopted in the contemporary philosophy of mind.

On Kitcher's (1990) view, we must understand Kant's argument in the Deduction as a response to Hume's account of personal identity, or, more accurately, Hume's insistence that there is no such account forthcoming. Hume famously denies that there is any basis for us to claim that

there is a unified thinker at the basis of our mental activity. He claims that we can only identify a 'bundle' of impressions and perceptions that lack any real connections and therefore exist dispersed and independently of one another. On Kitcher's view, Kant's account of the unity of apperception is meant to be a response to this claim. On her reading, this unity obtains when "cognitive states are connected to each other through syntheses required for cognition" (1990, 105), and these syntheses are precisely meant to establish real connections, of the sort that Hume denied exist, between the mental representations that originate in sensibility. They establish real connections insofar as they are guided by a set of a priori rules or laws that Kitcher identifies with the categories of the understanding (cf. Kitcher 1990, 106). Importantly, Kitcher insists that these rules are descriptive: "rules govern syntheses only as the law of gravity governs the movements of the planets: Theorists can appeal to these rules to describe what is happening" (1990, 83). This commitment to the descriptive status of the laws governing synthesis is required by her more general claim that

> acts or processes of synthesis could not be performed by agents. They are unconscious activities within agents that enable them to have cognitive capacities required for agency. In Daniel Dennett's useful terminology, they are "subpersonal" processes, not acts performed by persons. (1990, 122)

On this kind of view, transcendental apperception and the transcendental synthesis of the understanding that is associated with it, is simply a function or mechanism that connects different mental representations with one another so that there are real connections between them. Once these real connections are established, the unity of apperception is achieved and the subject can thereby cognize the object that is represented through the connected representations.

What is important to note is that, on Kitcher's interpretation, the real connections between representations that are generated through synthesis are *causal* connections, and she takes this feature of Kant's account to anticipate contemporary functionalism:

> Kant [adopts] an account of the representational content of judgments that is like that defended by contemporary functionalists. Functionalists take cognitive states to have particular contents in virtue of their causal relations to stimuli, responses, and other cognitive states... Could Kant have held a view that is so close to contemporary wisdom? … [b]y the time Kant was writing, causal connection was widely held to be essential to representation, and he always supported this view. However, he also believed that judgments are not in an immediate [causal] relation to objects but that they are only related through intuitions (A68/B93). Hence, it would be natural for him to believe that the representational character of judgments could only arise through their dependence on intuitions. (1990, 111)

So, on this conception, the understanding functions, through synthesis, to create relations of existential dependence between mental states. These syntheses are causal processes governed by rules inherent in the understanding, and in virtue of being so governed, they generate cognitive content from the representations given in intuition.

Now, this conception of the understanding has two important implications for how Kitcher interprets apperception and spontaneity. First, on Kitcher's view, apperception is not reflexive: it does not constitute the self's awareness of itself, or even of its own state: "Apperception does not indicate any awareness of a separate thing, a 'self,' or even that different cognitive states belong to a separate thing, a 'self'" (1990, 105). Instead, on her view, apperception is simply the transcendental synthesis, that is, the causal process that connects different representational contents so as to create the unity of apperception required for cognition. As a result of this view of apperception, she also downplays the sense in which the act of synthetic connection is meant to be spontaneous. Accordingly, she ascribes to apperception only a *relative* spontaneity, which she takes to be consistent with the idea that the understanding functions as a causal system that establishes real connection between mental states.

We can summarize Kitcher's psychological position by distinguishing three claims to which she is committed:

(1) Transcendental apperception is not reflexive. It does not represent the subject itself; nor does it represent any state of the subject as belonging to it; and nor does it represent the subject as the active principle that synthesizes or connects different states with each other. Transcendental apperception simply is that synthesizing activity itself.

(2) The mind, and, more specifically, the functioning of the understanding, can be explained in the same causal terms as the processes that we observe in the phenomenal world.

(3) Accordingly, the rules and laws of the understanding do not function as norms that an agent is responsive to in generating the connections between representations. There rules are descriptive – they simply tell us how the synthesizing activity proceeds.

Taken together, these tenets of Kitcher's view lay the groundwork for an interpretation of Kant's theory of apperception that is consistent with a broadly naturalistic conception of the mind.[11]

Whatever the appeal of this kind of reading from a contemporary standpoint, however, several prominent commentators complain that views such as Kitcher's fail to do justice to some of the most central and also most valuable aspects of Kant's account of cognition. Allison (1996), for example, charges that "[Kitcher's] version not only leaves a distorted view of the historical Kant, but also obscures some of his most important insights, specifically, those regarding self-consciousness and spontaneity" (1996, 55). Allison's fundamental concern with Kitcher's account, which he shares with others, is that, by characterizing apperception in causal functional terms, she is effectively removing the subject from Kant's account of cognition entirely. As we will see

[11] Of the three commitments distinguished here, I take the second and third to be absolutely essential to any psychological view, whereas the first is specific to Kitcher's particular articulation of the view. What is essential to a psychological account is the claim that Kant's theory of apperception is part of a naturalistic, descriptive theory of the understanding. Within the context of such an account, there is room for disagreement about whether apperception involves any form of reflexivity. Andrew Brook, for example, another influential proponent of the psychological approach, treats transcendental apperception (what he calls 'apperceptive self-awareness') as a form of awareness, albeit sparse, of ourselves as subjects of representation (see Brook 1994).

below, these commentators have objected to all three of the commitments distinguished above, and understanding the grounds of these objections will help us appreciate the distinctive features of the epistemic approach.

In outline, the complaints are as follows. The first complaint, leveled by both Allison and Pippin, among others, is directed against Kitcher's denial (claim 1 above) that self-consciousness is reflexive. According to both commentators, reflexivity is an ineliminable feature of Kant's theory of apperception, without which we simply cannot make sense of his account of cognition. Another complaint is directed at Kitcher's claim, embodied in claims 2 and 3 above, that we can make sense of how the understanding functions in *causal* terms, and accordingly her insistence that the understanding is only relatively spontaneous. According to these commentators, there are two problems with this view. First, they claim that the causal view of the understanding makes it hard to understand the sense in which judgments, the characteristic exercises of the understanding, can really be ascribed to the subject. Second, they claim that the approach is inconsistent with Kant's doctrine of the spontaneity of the understanding. According to these readers, that doctrine rules out the possibility that the understanding can be characterized in terms of the causality that operates in the phenomenal world and that characterizes material objects:

> What the spontaneity [of the understanding] reveal[s] is that Kant also thought that such analysis revealed that there was one feature of any such subject that *could* never be the kind of property a phenomenal subject could have – spontaneity. (All phenomenal subjects are necessarily and completely parts of causal series.) (Pippin 1987, 458)

To see how these objections reflect the positive commitments of the epistemic view, it will be especially helpful to focus on Allison's disagreement with Kitcher.

Allison's first criticism is that Kitcher's view fails to explain how objective representation is possible. On Kitcher's view, as Allison presents it, "[t]he mind is nothing but a causally and

contentually connected system of mental states" (1996, 57), within which any given mental state, M_n, depends on the contents of certain prior states, $M_1 – M_{n-1}$, for its content. According to Kitcher, these contentual connections between states suffice for a state to qualify as objective representation, and, importantly, the contentual connections are generated through "acts or processes of synthesis [that] are not performed by agents" (1996, 122); the subject is not conscious of either these processes themselves or of the rules in accordance with which they take place. Allison thus rightly asserts that Kitcher's view of the mind is one that "allows no room for a consciousness of the I and its activity" (1996, 57). But removing this form of consciousness from Kant's account, according to Allison, makes it impossible to understand the possibility of objective representation.

According to Allison, we cannot understand the possibility of objective representation without recognizing the reflexivity that self-consciousness alone can introduce into cognition. Subpersonal processes of the kind that Kitcher appeals to will not account for this possibility, Allison argues, because they cannot generate the distinctive kind of unity that characterizes cognition. Synthetic processes, as Kitcher understands them, only account for "the *de facto* presence of a rule-governed unity of representation in consciousness" (1996, 59), and, according to Allison, the mere presence of such a unity is insufficient for objective representation or cognition. Cognition, for Allison, has a distinctive epistemological significance because it constitutes a way in which a subject *takes things to be*. This epistemological significance requires, over and above the mere existence of a *de facto* unity of representations, "the thought or conceptual recognition of this unity" (1996, 59). It is only once this recognitional component is in place that the relevant representation can count as a way in which the subject takes things to be. And since

every cognition by its nature requires recognition of the unity of its constituent elements, it follows that there is an ineliminable reflexive dimension built into objective representation:

> In other words, this unity must not only be *in* a single consciousness, it must be *for* that consciousness in the sense that the mind must be able to represent it to itself or, equivalently, to recognize it as such. This reflexive dimension of cognition follows directly from the normative nature of the claim of objective connection. (1996, 59 – 60)

Given that Kitcher's account does not open space for this reflexive dimension, Allison concludes that her interpretation is unable to adequately explain the possibility of objective representation.

Allison goes on to argue that in order to become conscious of the unity of the representations that constitute a cognition, the subject must be capable of becoming conscious of its identity. To unite two representations, R_1 and R_2 in consciousness, and for that connection to be objective, it is not sufficient for R_2 to depend causally on R_1, or vice versa. What is required in addition is that the I that thinks R_1 is conscious of its identity with the I that thinks R_2, since without this awareness of self-identity, it would not be possible to become conscious of R_1 and R_2 as constituents of a common unity.[12] Allison therefore clarifies that, on his reading, the subject's consciousness of her identity is not a consciousness that is separate from the recognition of the

[12]Here is Pippin making a parallel point:

> Kant's clearest expression of the apperception thesis occurs in the same passage where the definitions above occur, section 16 of the second edition Deduction. There Kant states and explains the claim that condition the Deduction. There Kant states and explains the claim that 'it must be possible for the "I think" to accompany all my representations' (B131), and goes on to assert that 'The principle of apperception is the highest principle in the whole sphere of human knowledge' (B135). I shall take this 'absolutely fundamental' highest principle to assert that all human experience is ineliminably reflexive. It is reflexive because, according to Kant, whenever I am conscious of anything, I also 'apperceive' that it is I who am thusly conscious. And all the problems arise from trying to understand what he means by this. He at least means that, in any, say, remembering, thinking or imagining, while the object of my intending is some state of affairs or other, I am also aware as I intend that what I am doing is an act of remembering, thinking, or imagining, and that I bring to these acts a subject identical with the subject of prior acts of intending. (1987, 459)

unity of representations in judgment; rather, to think this unity is "at the same time, to think one's identity" (60).[13]

As we have seen, Allison's first objection revolves around what he calls the 'epistemological significance' of cognition —the fact that cognition constitutes a way in which the subject takes the world to be; by removing reflexivity from Kant's account of apperception, he argues, Kitcher cannot account for this significance. He now goes on to argue that these same considerations speak against her view that the understanding is only relatively spontaneous. On Kitcher's reading, the spontaneity of the understanding can be explained in functionalist terms: in calling the understanding 'spontaneous', as she reads him, Kant is alluding simply to the fact that the understanding has a mechanism for generating new outputs given the inputs that it receives from the senses. Once again, Allison argues that this feature of Kitcher's reading fails to account for the epistemological significance of cognition. On Allison's view, the objection is straightforward: Kitcher's minimalist conception of spontaneity cannot account for the 'normative dimension' of the act of judging.

We can see why the relative spontaneity approach cannot capture this normative dimension by recognizing that on this view there are only causal connections between representations. According to Allison, we typically say that 'one has understood … the proposition *P*" if one has "arrived at it in the correct way" (63). The difference between the accounts of spontaneity can then be seen by considering how they would need to characterize the subject's arriving at a belief in the 'correct way'. On Allison's view, Kitcher would have to say that arriving at a belief in the correct

[13] Cf. Pippin: "My asserting that S is P is not an assertion of mine unless I am aware that I am asserting, not entertaining the possibility of S is P, and no such complex judgment, requiring, as Kant thinks he can show, continuity over time, is possible unless there is one continuous subject of experience over time, aware of its continuity in any conscious act. This apperceptive characteristic does not mean that the fact that I am perceiving rather than imagining is itself directly attended to, but that such an aware- ness is an inseparable component of what it is consciously to perceive, imagine, remember, etc." (1987, 459-60)

way requires only that it was caused appropriately, where presumably that means that the belief-generating system does not malfunction. Allison wants to insist, however, that this could not be Kant's position. The causal story, he claims, is not sufficient for understanding.[14] On his view, the correct way of arriving at a belief that *P* is to take into account one's prior beliefs and to see that *P* is justified in light of them, or warranted by them. One therefore takes one's prior beliefs as *reasons* for holding *P*, and to do so is to treat one's belief-forming process as assessable in light of epistemic norms. Allison claims, moreover, that I can take my beliefs to function as reasons only if I recognize them as such, and "[s]ince this is an act that the subject must perform for itself (self-consciously) rather than a cognitive state in which it finds itself, it follows that we must assume an absolute and not a merely relative spontaneity in order to conceive of its possibility." (64)

The final concern that these commentators raise is with Kitcher's insistence that the mind is part of the phenomenal world and that its operations are governed by the causal structure and laws of that world. According to Kitcher, it follows from

> the impossibility of noumenal knowledge... [that] the doctrine of apperception must present a phenomenal, if highly abstract, aspect of the self. If Kant is right that anything of which we can have knowledge must be governed by causal laws, then synthetic connection will be lawful. (Kitcher 1990, 139)

This feature of Kitcher's position, they claim, is inconsistent with key aspects of Kant's transcendental idealism. The transcendental subject that is analyzed and discussed in the Analytic

[14] Here is Pippin, again: "The dependence of M2's occurrence on other mental states is one thing; but M2's being a representation of, say, m is quite another. For that to be explained, I must be said to take the individual representation to be in a connection with other representations, or possible representations, as a condition of such possible intentionality. Or, for a judgmental representing to be an epistemic claim ('objective' in Kant's language), I must take up the contents of intuition and the mental states that can be said to be produced by such intuitions, and make such a claim. It cannot both be said simply to occur in a relation of existential dependence on other mental states, and to be 'objective' (possibly true or false) representing of mind, unless I so 'take' it" (1987, 468).

is meant to be a condition of the possibility of empirical cognition and experience; as such, this subject is presupposed by any cognition that we have of the phenomenal world, and that includes cognition of ourselves as subject. By situating this subject within the domain of phenomenal causal laws, Kitcher's account thus makes the subject depend upon the very causal order that is supposed to depend upon *it*. Consider, for example, the way in which Richard Aquila (1989) expresses this point:

> functionalism appears to get things backward. This is because, for Kant, causes and effects, at least insofar as they are possible objects of knowledge, are essentially governed by systems of causal laws. So according to the functionalist approach, the very idea of conceptual content would have to presuppose the idea of a system of causal laws.... But in Kant's own thinking, the very *idea* that a system of laws naturally obtains seem to be derivative from the idea of a being who is capable of representing determinate sorts of objects. (1989, 31)

In her account of the mind, Kitcher wants to help herself to the explanatory resources of natural science, but Aquila objects that those resources are not available to her given the foundational role played by the transcendental subject in explaining the possibility of the very laws that govern phenomenal nature.

Aside from this worry that functionalism gets things backward, there is also a more general concern that we cannot fully appreciate the significance of the fact that Kant characterizes the understanding as *spontaneous* without recognizing that by doing so, he is signaling that the higher faculties of cognition function independently of the causal constraints that govern nature. Here is Pippin:

> The formal conditions of knowledge require that the content of cognition be 'actively' conceptualized in a way that is finally, at some stage, causally independent of the causally produced reception of that material, and of any initial causal-series processing of that information, [consequently] a 'thinker' cannot 'really' be a causal system… (1987, 451)

Pippin's view brings together many of the strands we have been discussing. With Allison, Pippin holds, first, that Kant all but identifies the spontaneity of thinking with apperception, and, second, that apperception is 'ineliminably reflexive'. It is precisely this ineliminable reflexivity, according to Pippin, that points to "a non-relative or absolute spontaneity in apperception, one in principle not susceptible to th[e] kind of indirectly causal analysis [proposed by Kitcher]" (1987, 466). For this reason, Pippin concludes that Kant is committed to denying that the subject is part of phenomenal nature (but also not necessarily, for that reason, noumenal).

We can helpfully articulate the main divergences between the psychological and epistemic readings along the following three dimensions:

(1) *Reflexivity* The first concerns the nature of the act of apperception or self-consciousness. On the epistemic reading, this act must be reflexive in some sense: it must be a consciousness of the activity or the state or the identity of the thinking subject. For the epistemic reader, as we have seen, it is only because transcendental apperception constitutively involves the subject's awareness of itself that judgment constitutes an epistemically significant mode of representation that is subject to assessment in light of rational norms. The psychological reader, in contrast, is not obliged to take a stand on this issue, though Kitcher, who identifies transcendental apperception with sub-personal acts of combination, holds that apperception lacks a reflexive dimension.

(2) *Normativity* The second concerns whether the rules or laws that govern the activities of the subject in generating cognition are normative. On the epistemic reading, these rules must be rules of which the subject is conscious and that she is, or can be, aware of when she combines representations. On the psychological reading, these rules are

descriptive. The subject is not aware of the rules as governing the act of combination, and the rules simply describe how the act proceeds.

(3) *Causality* The third concerns whether the subject is part of the causal order of nature and whether her activity in thinking is governed by the laws of that order. On the epistemic reading, she is not, instead her thinking is governed by rational norms in light of which she is in some sense responsible for the judgments that she makes and the beliefs that she has. On the psychological reading, she is not independent of that order, and the causal laws that govern natural processes also govern her activity in thinking.

Whereas Kitcher's stance on these three issues yields an interpretation on which Kant is cast as a precursor to both functionalism and naturalism, the epistemic reading settles the issues in such a way that Kant emerges as a thinker who removes the subject from the natural order, who insists on an irreducibly reflexive dimension to all cognition, and who thinks of the laws governing the understanding as *norms* that the subject recognizes as such.

Now that we have a clear sense of the fault lines dividing these two approaches, I can describe the path that my view will tread between them. I side with the psychological reader on two major issues: *Reflexivity* and *Normativity*. With the psychological reader, I hold that the act of apperception is not reflexive, and I agree, too, that the rules of the understanding are descriptive and not normative. I agree, moreover, that the processes that generate judgments are sub-personal and not acts of the subject as such. Finally, I side with the psychological reader in viewing the spontaneity of the understanding as a merely 'relative', not an 'absolute' spontaneity – though we will see that my understanding of this Kantian distinction departs sharply from that presupposed by both the epistemic and the psychological reader. As we will see, in characterizing the understanding as relatively spontaneous, it is not incumbent on us to accept that its functions are

explicable in the causal terms that are appropriate to phenomenal nature, and indeed, on the third issue, *Causality*, I side with the epistemic reader. I accept that, in characterizing the faculty of understanding and the act of apperception as spontaneous, Kant signals that the subject must be understood as independent of the causal order of material nature, and that, if we are to make sense of her activity of thinking it cannot be simply in terms of mechanical causal processes. But, I will show, we can articulate a non-mechanistic conception of the understanding without reverting back to either a reflexive conception of apperception or a normative conception of the laws of the understanding.

Section 3: Overview of Chapters

In this final section, I preview each of the five chapters that make up the dissertation. My hope is that, in addition to clarifying the individual contributions of each chapter, this section brings together the preceding two by helping the reader to understand how my interventions in the secondary literature on apperception, as laid out in Section 2, are motivated by appeal to the rationalist metaphysical framework described in Section 1.

In Chapter 1, I take aim against the claim that transcendental apperception is reflexive. We have seen that this claim is a mainstay of the epistemic reading, and indeed it is the common element uniting several otherwise disparate interpretations of Kant's theory of apperception. For all of their differences, these commentators agree that transcendental self-consciousness is a consciousness that is directed *inward*, towards the subject whose consciousness it is (the 'self'), or to some aspect of this subject, such as its activity of representation or the resulting state of consciousness. My argument in this chapter is primarily negative, although I start sketching the shape of the alternative formalist view that I then defend fully in Chapter 4.

At the heart of the reflexive account of transcendental apperception is an assumption about the nature of consciousness in general, according to which, consciousness is an intentional relation, which results from an activity of the subject that is directed at some object; this object is then meant to contribute to the content of the resulting conscious state. Now, if we carry over this general conception of consciousness to the case of self-consciousness, we find that self-consciousness too must involve some intentional relation to an object, but that the object in this case is internal rather than external to the subject. Thus we arrive at the reflexive proposal, on which the object of self-consciousness is the self, or some aspect of the self, such as its activity or state. In Chapter 1, I argue that Kant does not and could not have conceived of self-consciousness in these terms: Kant does not take self-consciousness to be consciousness of or about the subject or any feature of the subject. The basis for this claim is that Kant makes it clear that self-consciousness is itself a condition of the possibility of intentional relations to objects, and it, therefore, cannot itself be understood in these intentionalist terms.

Against this proposal, I introduce my own interpretation, which centers on the claim that transcendental apperception constitutes the *form* of representation of any and all objects in thought. This form is not itself a representation that the subject is conscious of; it is, instead, an *organizing principle*, without which the representations given in sensibility cannot inhere in the subject as thoughts. This formalist proposal, which I develop and motivate throughout the rest of the dissertation, takes Kant's faculty psychology as a starting point. Kant identifies pure apperception as an *act*, and it is therefore to be understood as a ground of inherence of certain accidents – thoughts – in a subject. He also characterizes it as an 'act of the self-activity of the subject' (B130) and as *spontaneous*. Given his faculty psychology, he must accordingly take it to be an act that is determined through an inner principle of the subject; I therefore read the 'self' in Kant's use of the

term ‘self-consciousness’ not as indication of the *object* of this consciousness, but as an indication of the *origin* of this consciousness within the subject.

Given this close connection between spontaneity and self-consciousness, a full defense of my interpretation of the doctrine of spontaneity will both motivate and clarify the formalist account of apperception. For this reason, in Chapters 2 and 3, I will develop an interpretation of the spontaneity of the understanding with an ultimate view toward explaining the sense in which apperception constitutes the form of thinking.

Having discussed the issue of reflexivity in the first chapter, in Chapter 2, I turn to a further commitment of the epistemic view. As we saw above in the discussion of Allison’s view, the claim that self-consciousness is reflexive is in service of a further commitment about the epistemic significance of the cognitive states generated by the understanding, and one way of spelling out what this epistemic significance amounts to is to appeal to Kant’s claim that the understanding is spontaneous. The epistemic reader wants to claim that the spontaneity of the understanding is meant to signify some *agency* that a subject has in determining what propositions to accept or reject, and that Kant’s project in the Analytic should be read as an account of the rules and laws that govern the processes that generate these cognitive states or judgments. These laws function as *norms* in accordance with which the agent’s judgments are subject to evaluation, and it is precisely because of this normative dimension that those judgments constitute epistemically significant forms of representation.

Having laid out this view of spontaneity, which I call the cognitive agency reading, I proceed to argue against it. To do so, I show that Kant is not committed to the view that the cognitive states generated by the understanding involve any commitment on the part of the subject regarding the *truth* of the content of the state. First of all, I point out that Kant’s discussion of the

various propositional attitudes that characterize an agent's commitment to the truth of a proposition does not occur in the Analytic at all; rather, this topic is treated in depth in his theory of holding-for-true (*Fürwahrhalten*), which is laid out much later in the *Critique*, in the Canon. Kant's focus in the Analytic, I argue, is restricted to the merely *semantic* functions that generate the contents of our thinking. In addition to these textual considerations, I aim to show that the normativist conception of the rules and laws of the understanding does not fit well with Kant's faculty psychology. Against the normativist conception, I argue that Kant's faculty psychology motivates a *constitutivist* reading of these laws, on which they simply describe the way that the understanding functions rather than prescribing rules to it. Once we recognize that the laws of the understanding flow from the nature (in the technical sense) of the understanding, we see that a constitutivist reading of how those rules function better fits with Kant's account. But if there is no normative dimension in Kant's account of the understanding, then we cannot make sense of the view that the spontaneity of the understanding is meant to signify the cognitive agency of the subject since, as we have seen, such a position assumes that the subject is responsive to certain norms.

In rejecting the normativist conception of the understanding and, with it, the cognitive agency reading of spontaneity, one might think that I must therefore accept the conception of spontaneity endorsed by proponents of the psychological approach to apperception. Recall, Kitcher insists that the spontaneity of the understanding is merely 'relative', which she takes to signal that the operations of the faculty are consistent with laws of the causal order that govern nature conceived of in phenomenal terms. According to this kind of view, the understanding functions *mechanically* to generate the contents of thought and cognition, and, therefore, the laws of the understanding simply describe those functions. But while Kitcher's reading certainly gives us *one* way of rejecting the normativist presuppositions of the cognitive agency reading, I will argue in

Chapter 3 that this mechanistic conception of the understanding is just as problematic as the normativist conception it replaces.

In Chapter 3, I argue against the mechanistic conception of the understanding by showing that once we set Kant's conception of spontaneity against the theoretical backdrop of his faculty psychology, we will see that Kant conceives of the human cognitive power in teleological terms.

I start by introducing Kant's account of activity. I show that Kant's notion of spontaneity is part of a broader theory of activity. Spontaneity is one specific form of activity; it is contrasted with a second form of activity, namely, *passivity* or *receptivity*. I explain how both can be understood as species of a single genus, and show that for Kant, mechanistic explanation is restricted to passive beings; *spontaneous* beings, by contrast, must be made intelligible through an alternative, teleological mode of explanation. I then turn to the specific form of spontaneity that characterizes the human cognitive capacity, and differentiate human *cognitive* spontaneity from the merely *conative* spontaneity that characterizes living beings.

As Kant makes clear, both in the third *Critique* and in his metaphysics lectures, living beings, unlike material nature, cannot be explained wholly in mechanistic terms and therefore must be understood teleologically. Now, importantly, Kant characterizes the spontaneity of living beings as belonging to their conative or desiderative powers; as such, it underlies the power of the being to act in accordance with its representations and thereby to act in such a way as to preserve, sustain and promote its life. Life is thus the inner principle of the conative faculties of all living beings, insofar as the promotion of life is the end or goal of the faculty of desire in animals. A teleological explanation of the animal will allow us to understand how this inner principle grounds the inherence of accidents in the animal.

Now, human beings can be distinguished from non-human animals to the extent that not only the desiderative power, but also the cognitive power is spontaneous. As a result, the human being can not only act in accordance with representations that allow her to preserve, sustain and promote her life; she can also *represent* the world in accordance with a certain inner principle. While life is the inner principle of the conative faculties of all living beings, insofar as the promotion of life is the end or goal of the faculty of desire in animals, the inner principle of the cognitive faculty of human beings must be understood as an inner principle of intelligibility that allows the subject to represent the world as a whole of lawfully interconnected objects. This inner principle of intelligibility is nothing other than the pure consciousness or apperception through which a subject's representations are connected in accordance with the rules of the understanding. And it is only by explaining the faculty in teleological terms that we can understand the way in which this inner principle grounds the inherence of representations in the subject.

In closing this chapter, I consider an objection to my account of spontaneity. The objection claims that my account overgenerates because it entails that human sensibility, which Kant explicitly characterizes as a receptivity, should also count as spontaneous. This objection, I show, overlooks an important feature of Kant's account of inner principles and also relies on a contentious and by no means obligatory account of pure intuition.

In Chapter 4, I return to a direct discussion of apperception: building on the results of the previous chapter, I develop and defend the formalist interpretation of apperception first announced in Chapter 1, according to which apperception is not consciousness of any particular object, but constitutes the organizational form of consciousness of objects as such.

I begin by arguing that the systematic role of pure apperception in Kant's account of the understanding is to make thought possible. I go on to show that it plays this role by introducing a

certain organizational form, which Kant identifies as *qualitative* unity, into the representations given in sensibility. It is only in virtue of having this unity, which is the organizational form of all thinking, that the representations given in sensibility can be *thought*. This appeal to qualitative unity puts us in a position to understand Kant's claim that pure apperception is the ground of possibility of the faculty of understanding (B131): by providing representation with qualitative unity, pure apperception is the ground of the inherence of representations in the subject as thoughts. We must therefore conceive of pure apperception as an *act*, through which the power of the understanding is exercised, that allows us to think the objects of sensible intuition. Finally, I argue that this act must also be understood teleologically. Its purpose is to make sensible representations given in intuition intelligible, and it does so by bringing a certain unity into representation by combining them in accordance with an *idea* of the form of understanding itself.

Finally, in Chapter 5, I clarify both the sense in which I take the understanding to be purposive, and how it is related to another higher faculty of cognition, namely, reason. I do so by situating the understanding in relation to the distinction between absolute and relative spontaneity, which Kant develops in his metaphysics lectures. I start by arguing for a novel interpretation of this distinction: whereas the standard reading holds that relatively spontaneous faculties are subject to mechanistic explanation, I argue for a reading on which both relatively spontaneous and absolutely spontaneous powers must be understood teleologically. The real difference between such powers is not that one is end-directed whereas the other is not. The difference, rather, is that only absolutely spontaneous powers set ends for themselves; relatively spontaneous powers are no less end-directed, but they must be given an end from an external source. With this account in place, I show that the understanding is relatively spontaneous, whereas reason is absolutely spontaneous in the senses argued for.

I argue that conceiving of the understanding as relatively spontaneous in this way helps us to see both that the end of the understanding is essential to it, and to see that this teleological reading can avoid the normative presuppositions of the cognitive agency view. I then argue that this distinction helps us to make sense of the relationship between reason and understanding. As a relatively spontaneous power, the understanding must be given its end by an external power, and I argue that it is precisely *reason* that provides the understanding with the end for which it acts. This end is an idea of the form of the whole of cognition; in producing its cognitions the understanding functions in accordance with this idea, and it must do so if it is to generate a unified and systematic (even if incomplete) representation of nature.

I close Chapter 5 by considering an objection to my account of the relationship between reason and understanding. In claiming that reason's idea of systematicity is a requirement, not only of the sophisticated cognitive task of scientific theory building, but also of the more elementary task of generating the particular cognitions that make up experience, one might worry that my account cannot make sense of Kant's insistence that the role of reason in experience is 'regulative' rather than 'constitutive'. I show, however, that this objection rests on a misreading of the regulative-constitutive distinction, on which only regulative uses of ideas can be necessary for experience. Against this reading, I propose an alternative reading, in terms of a contrast between the determination of a faculty and the determination of an object; this alternative, I show, is textually well-motivated and gives my view the resources to respect Kant's insistence that the legitimate use of reason in experience is only ever regulative.

With the formalist view explained and defended, I conclude the dissertation with a postscript on practical self-consciousness. The formalist model, I suggest, can be exported beyond the context of Kant's theoretical philosophy into his practical philosophy, where it promises to

make sense of the deep parallels between theoretical and practical cognition, the sense in which the moral law depends upon the absolute spontaneity of reason, and the connection between practical self-consciousness and transcendental freedom.

CHAPTER 1

Transcendental Apperception and Reflexivity

At the outset of the Transcendental Logic in the *Critique of Pure Reason*, Kant observes that our cognitive faculty consists of two 'sources in the mind', one receptive, the other spontaneous. He goes on to insist that for cognition to be possible, the distinctive product of each of these 'two fundamental sources' must be *united* (A50/B74). Kant undertakes the project of explaining how they are united in the Deduction of the Pure Categories of the Understanding, for there he shows that the representations that the spontaneous faculty 'brings forth' — the categories — can legitimately be applied to what is given to us in sensible intuition, that is, that the categories have 'objective validity'. At the start of this famous section in the B-edition of the *Critique*, Kant introduces a principle from which he proposes to prove the objective validity of the categories. That principle is what he calls the original synthetic unity of apperception, or also the transcendental unity of self-consciousness. The absolute centrality of this principle for Kant's project is evident in his famous declaration that "[T]he synthetic unity of apperception is the highest point to which one must affix all use of the understanding, even the whole of logic and,

after it, transcendental philosophy" (B134n), and that it is "the supreme [principle] in all of human cognition" (B135).

Notwithstanding the central role it plays in the Deduction, there is little consensus in the literature about what the synthetic unity of apperception *is*, and what role it plays in the argument of the Deduction. The difficulty of developing a satisfactory account of the original synthetic unity of apperception is not altogether surprising given that almost *every* aspect of Kant's argument in the Deduction — its aim, structure, starting point, and outcome — is a matter of controversy.

Amid the interpretive discord, however, an important trend has nevertheless emerged about the *structure* of the transcendental self-consciousness that plays such a central role in Kant's argument. The trend is to assume that Kant takes self-consciousness to be *reflexive*.[15] This seems to follow simply from the fact that Kant identifies it as *self*-consciousness, for the term by itself suggests that whatever else this consciousness is, it stands in some kind of relation with itself. We might then be tempted to go on to explain this relation as follows: it is a consciousness whose object is either (a) the *self* (from which it emerges) or (b) *itself* (the consciousness that it is). Of course, there are many different ways of spelling out what exactly it would be for consciousness to take either (a) or (b) as its object. If, for example, this consciousness took (a), the self, as its object, then questions immediately arise about the nature of the self that is revealed in this consciousness: is the conscious self aware of itself as a noumenal or a phenomenal being, or as something else altogether? Or, if this consciousness took (b), itself, as object, then we can ask whether it is conscious of itself as an activity, or as a state, or as the content of a state. All of these

[15]As we saw in the Introduction, this assumption, though very widespread, is not completely universal. Some psychological readers, such as the early Kitcher, do not take self-consciousness to be reflexive, but still, this reading is the norm, both within the epistemic and psychological camps. Indeed, even Kitcher goes on, in her later work, to embrace a reflexive account. We will discuss Kitcher's more recent contributions later in the chapter.

questions are taken up and treated in great depth in the literature on Kant's conception of self-consciousness. For the moment I just want to note that these otherwise very different interpretive lines share something in common, namely that the structure of self-consciousness is *reflexive* — that is, in some or other sense it is consciousness that is directed inward, at itself.

We can describe this common structure more clearly by bringing into focus the intentionalist conception of consciousness that is prevalent in most interpretations of Kant's theory of apperception. In the contemporary literature on consciousness, there is a distinction between purely *qualitative* conceptions of consciousness, according to which conscious states are *not* necessarily representational, and *intentionalist* conceptions of consciousness according to which conscious states are necessarily representational or intentional.[16] The latter conception ties consciousness to intentionality essentially, so that we can say that *every* state of consciousness is consciousness *of* something, whereas qualitative conceptions of consciousness insist that, while there might be some conscious states that are intentional, intentionality is not necessary for a state to be conscious, for there are some conscious states that are purely qualitative and do not represent anything. Advocates of both qualitative and intentionalist conceptions of consciousness might also deny that intentionality is sufficient for consciousness; there might be intentional states that are not conscious, [17] but the intentionalist must deny, at least, that there are conscious states that are not intentional.

[16] Cf. Siewert (2017) for a helpful overview of the relationship between consciousness and intentionality in the contemporary debate.

[17] An externalist, like Putnam (1981), might think that the phenomenal character of a mental state is independent of any relation of the state to the environment that caused it. If the meaning or the representational content of the state is determined entirely by the environment and not 'internally', then it would be plausible to claim that intentionality is properly speaking independent of the specific 'qualitative feel' or consciousness that is associated with being in that state.

If we were to import this intentionalist model to the specific case of self-consciousness, we would need to say something about what it is that the consciousness in question is directed at, or what it is that this consciousness represents. Clearly, the intentional object is the self, and so it is not difficult to see how the reflexive structure that is attributed to self-consciousness arises once we presuppose an intentionalist account of consciousness. Self-consciousness is regarded as just a special case of object-consciousness. What makes it 'special', and confers on it its distinctive reflexive structure, is just that the specific object at which it is directed is not external to it. Since the object at which this consciousness is directed is either itself or is something that is somehow contained in it, we might think of it as involving some inward-directedness or some form of introspection.

In this Chapter, I will consider what I will now call intentionalist interpretations of Kant's conception of transcendental self-consciousness, and I will argue that they fail to accurately describe Kant's view. I call these interpretations intentionalist because they assume that consciousness is necessarily intentional and then attempt to explain self-consciousness as a special case of the intentionality inherent in consciousness. These views take consciousness to be more primitive (explanatorily and metaphysically) than self-consciousness since the latter is just consciousness of a particular, and perhaps in some way to be specified 'special', object — the self, its activities, or its states. Note, too, that reflexivity is a necessary feature of self-consciousness on such an intentionalist view.

I will consider the intentionalist view from two perspectives, one historical and one contemporary. The historical perspective provides us with the 'classic' reflexive model that Dieter Henrich aptly calls the "Reflection Theory" and which he also attributes to Kant. Henrich (1982 [1966]) argues that Fichte first appreciates the deep philosophical difficulties that such a model

faces, and criticizes Kant's adoption thereof. Henrich, taking himself to be following Fichte, then proposes an alternative account of self-consciousness according to which it is a primitive *non-reflexive* act. The contemporary perspective changes tack and introduces important semantic and epistemic considerations to be accounted for in a theory of self-consciousness. Kant scholars who have taken these considerations seriously attribute to Kant a different intentionalist model, which I will call the 'self-ascription model' of self-consciousness. This model too endorses the reflexive structure that is common to intentionalist accounts of self-consciousness, insofar as self-consciousness is taken to be consciousness of either the activity of generating the internal states of a subject, or of their content. My aim is to argue that Kantian transcendental self-consciousness does not have a reflexive structure, and so it is not a kind of consciousness that can be adequately explained on the basis of any intentionalist model, either the Reflection Theory or the self-ascription model. Self-consciousness on Kant's view, I argue, is in fact a kind of *pre*-reflexive consciousness that first makes possible reflection and, with it, conceptual object-consciousness. On this interpretation, Kant takes transcendental self-consciousness to be metaphysically and explanatorily prior to our consciousness of objects.

On my view, which I introduce in this chapter and continue to defend and elaborate throughout this dissertation, the term 'self' in 'self-consciousness' does not function to indicate the intentional object of self-awareness; instead, it must be understood as functioning in the same way as the term 'self' in Aristotelean 'self-movement': to indicate a certain distinctive kind of spontaneous activity, namely, one that originates from *within* rather than from *outside* the actor. As I will show throughout the rest of the dissertation, this re-alignment in our thinking about the nature of self-consciousness is motivated by Kant's account of *spontaneity*, and his faculty psychology more broadly, and it points us toward a distinctive, non-intentionalist interpretation of

apperception. The aim of the present chapter is to clear the way for this positive project by pressing the negative case against the intentionalist reading. I will show both that the intentionalist reading is a pervasive feature of contemporary Kant scholarship, and that it faces fundamental textual objections that should lead us to question its viability as an interpretive model.

I proceed in four sections. In Section 1, I introduce the main topic of my project, transcendental apperception, and provide a brief overview of the shape of the interpretive position that I aim to develop in the rest of the dissertation. In doing so, I provide a gloss of some of Kant's key terminology, including his distinction between pure and empirical apperception. Clarifying this distinction at this early stage is important, as doing so will help me clarify the way in which I see Kant's account of pure apperception as departing from intentionalist presuppositions. In Sections 2 and 3, I give evidence of the influence of the intentionalist model in Kant scholarship. I begin, in Section 2.1, with what I take to be the canonical formulation of the intentionalist conception of self-consciousness, what Dieter Henrich call the Reflection Theory. I detail the model, and go on to consider, in Section 2.2, a recent alternative to the model, due to Karl Ameriks, which I argue in fact inherits its principal intentionalist assumptions. I go on, in Section 3, to discuss the most common interpretive trend in the Anglophone literature on Kant, what I will call the self-ascription theory of apperception. For this purpose, I consider three of the most influential and sophisticated views, the accounts of Strawson (Section 3.1), Kitcher (Section 3.2), and Allison (Section 3.3). I show that despite fairly wide-ranging interpretive disagreement, these authors all share the basic commitment that is common to intentionalist views, namely, that self-consciousness is consciousness of either an activity or the product of an activity of the mind. Moreover, in attributing this view to Kant they all share a further commitment, namely that Kant holds self-consciousness to be subject to an 'objectivity condition', on which the application of

the categories is necessary for self-consciousness. This observation forms the basis of Section 4, where I argue that Kant cannot be regarded as subscribing to the objectivity condition. Since, as I also argue, the only textually plausible version of intentionalism requires that we ascribe to Kant the objectivity condition, my objection to the latter serves as a fundamental objection to all intentionalist interpretations. Finally, to conclude, I say something to briefly anticipate the shape of the account to be developed in subsequent chapters and the rationale behind my focus on the topic of *spontaneity* as a pathway to understanding apperception.

Section 1: Kant's Theory of Apperception

In this section, I introduce my topic and clarify some of the central terminology that Kant uses in articulating his account of apperception. As we will see, Kant draws a distinction, within the broader domain of apperception, between empirical and pure apperception. Aside from helping to bring the focus of my project on pure apperception more clearly into view, explaining Kant's distinction between these two forms of apperception will be useful for clarifying certain aspects of the positive position that I want to develop. In particular, my contention will be that while the intentionalist assumption that informs interpretations of Kant's theory of self-consciousness is justified in the case of empirical apperception, it does not carry over to his account of pure apperception. I will suggest that we should not assume that Kant conceives of pure apperception, first and foremost, in representational or semantic terms and that we can gain a better understanding of his account by focusing on the metaphysical terminology that he introduces in developing his view.

It is useful to start by introducing a kind of working conception of Kant's notion of apperception. Since my aim in the rest of the dissertation is to work out Kant's theory of

apperception in much more detail, the way that I introduce it here will almost certainly seem unacceptably cursory and brisk, but it will nevertheless be helpful to introduce some fundamental assumptions (some of which might not be wholly uncontentious) just to clarify the way that I take Kant to be using terminology.

Kant identifies apperception as an *act* or *activity* (e.g. at B137; cf. also A547/B575, *APV* 7:135, *APV* 7:142-2, *MD* 28:670) and sometimes more specifically as the *self-activity* of the subject (e.g. at B130, B68-9, B132, AC 25:10), and he also tells us that the 'I' is the representation of this activity (4:334 fn).[18] He nevertheless also refers to apperception as a *faculty* (e.g., at A117n, B153), and indeed as one of the sub-faculties that, together with sense and imagination, comprises the understanding (A94; also cf. A119), and one might wonder how to square these formulations. I will take it that there is no deep tension in saying both that apperception is a faculty and that apperception is an activity, since Kant defines a faculty as 'the possibility of acting' (*MM* 29:772). A power, by contrast, is "a faculty insofar as it suffices for the actuality of an accident" (*MM* 29:823), and a faculty 'suffices for the actuality of an accident' just in case it acts. So, the act of apperception is the actualization of a certain kind of power, which, if the power is actualized in combination with certain other powers, constitutes what Kant calls the understanding.[19]

We must next consider what kind of power apperception is (taken in abstraction from the contributions of the other powers that together with apperception comprise the understanding). I will take it that, for Kant, the power of apperception is a power of *consciousness* and that its distinctive contribution to the understanding is that it facilitates our consciousness of objects.[20]

[18] I assume that when Kant says that the 'I' is the representation of the activity that apperception is, he means that it is the way that this activity is represented linguistically.
[19] There are some important and difficult questions about how faculties and powers are individuated that I cannot address here.
[20] The textual basis for this claim will become clear in Chapter 4.

Accordingly, the terms 'apperception' and 'consciousness' can be used interchangeably and are at least extensionally equivalent.[21]

Within this broader conception of apperception or consciousness, Kant draws an important further distinction between *pure*, or *transcendental* apperception and *empirical* apperception. Strikingly, Kant characterizes empirical apperception, or what he also sometimes refers to as inner sense (e.g., at A107), in explicitly intentionalist terms. He says, for example that "the I of inner sense, that is, of *the perception and observation* of oneself, is not the subject of judgment, but an object" (*APV* 7:142) and, continuing the same passage, concludes that

> [o]ne must therefore distinguish pure apperception (of the understanding) from empirical apperception (of sensibility). The latter, when the subject attends to himself, is also at the same time affected and so calls out sensations in him, that is, brings representations to consciousness. (*APV* 7:142)

It is clear from this passage that Kant takes empirical apperception to consist in a kind of introspection: it is the consciousness through which representations are generated (in time) when we perceive and observe ourselves (and our states). He also claims that it is a consciousness that can be understood in terms of *self-affection*, which is what he claims 'brings representations to consciousness'. Given the role of self-affection in empirical apperception, it is aligned in the passage above with the passive cognitive faculty of sensibility.

We get a better understanding of the kind of consciousness that empirical apperception is by noting that Kant equates empirical apperception with *apprehension*, and he notes that it makes perception – the empirical consciousness of objects — possible:

> If we consciously represent two acts: inner activity (spontaneity), by means of which a concept (a thought) becomes possible, or reflection; and receptiveness (receptivity), by means of which a perception (*perceptio*), i.e., empirical intuition, becomes possible, or apprehension; then consciousness of oneself (*apperceptio*) can be divided into that of reflection and that of apprehension. The first is a

[21] Tolley (2016) makes the historical case for this terminological alignment.

> consciousness of understanding, pure apperception; the second a consciousness of inner sense, empirical apperception. (*APV* 7:135 fn)

Kant distinguishes between two acts in this passage: one act, characterized by spontaneity, makes *concepts* possible, and a second act, which is characterized as receptivity, makes *perception* possible. [22] The second act is called apprehension, and Kant says that it is "a consciousness of inner sense, empirical apperception" (ibid.).[23] Now, the passage above might initially seem to equate empirical intuitions and perceptions, but in numerous passages in the *Critique* and elsewhere, Kant makes clear that perception, rather than being identical to empirical intuition, consists in the consciousness of the appearance that is the object of empirical intuition (see, for example, B160).[24] Since Kant claims that empirical apperception is what makes perception possible, we can thus read him as claiming that empirical apperception or apprehension is the consciousness of certain internal states, *intuitions* [*Anschauungen*], that produces *perceptions* [*Wahrnehmungen*], that is, empirical consciousness of the objects of intuitions, namely, appearances [*Erscheinungen*]. What is important to note, then, is that empirical apperception plays its role in producing perceptions through a distinctive kind of introspection, in which a subject brings her own empirical intuitions to consciousness by becoming aware of the objects of those intuitions.

Now, we might be tempted to carry over this introspectionist model from empirical apperception to transcendental apperception, understanding the latter as just a more abstract or a

[22] Note that Kant also calls receptivity an act; as we briefly explained in the Introduction and will see more clearly in Chapter 3, Kant in fact distinguishes between two kinds of activity, spontaneity and receptivity, and while it might seem oxymoronic to say that an act can be passive, there is for Kant no deep contradiction here, and it is, in fact, exactly how he conceives of receptivity.

[23] I think it is important to note that the 'of' here should be read attributively, so that Kant is claiming that it is a consciousness that *belongs to* inner sense, not a consciousness *about* inner sense.

[24] The case for recognizing this distinction between perception and empirical intuition in Kant has been made most thoroughly by Tolley (2020).

priori version of empirical apperception. Our task then would be to try to spell out more clearly what the non-empirical or pure observation of ourselves would involve. However we spell that out, if we conceive of transcendental apperception just as a 'pure' version of its empirical counterpart, then it too would be a kind of *introspection* or self-observation, though not, of course, observation of our temporally determined internal states. We will explore different variations of this introspectionist approach later in the chapter; as we will see, numerous variants of the approach have been developed, which conceive of transcendental apperception as consciousness of the activity (sometimes conceived as atemporal) that produces our internal states, or the a priori products of that activity, or again as consciousness of the totality of our empirical consciousness or perceptions. For all their differences, what these approaches have in common is that they view transcendental and empirical apperception as isomorphic: the former, like the latter, is understood to involve the introspection of some mental item, the only difference being that the item introspected in the one case is 'pure' whereas in the other it is empirical.

The focus of this dissertation will be Kant's account of pure or transcendental apperception. I will argue that it is in fact a fundamental mistake to view pure apperception on this intentionalist model. In its place, I will develop an alternative model, which I will call the *formalist* model. In the remainder of this section, I want to briefly outline what this alternative approach looks like and indicate upfront some of the textual resources I will draw on to develop this interpretation.[25]

Against the intentionalist model, I will argue that transcendental apperception should not be understood as a kind of self-observation: it is *not* introspective, and, in fact, should not be conceived of in the first place as consciousness *of* anything at all. Thus, it is not simply a pure counterpart of empirical apperception, another species of self-observation that differs only insofar

[25] The fully developed systematic argument for this view is in Chapter 4.

as it has a different intentional object; for I will deny precisely that transcendental apperception has *any* distinctive intentional object at all. Now, this can sound strange if we assume an intentionalist conception of what consciousness is. On an intentionalist view, recall, consciousness is by its very nature consciousness *of something*; that is to say, consciousness is by its nature relational, and in particular, it constitutes an intentional relation of a mental content to an object. If we presuppose this sort of conception of consciousness, saying that transcendental consciousness has no intentional object will sound paradoxical.

We can alleviate some of the concern by noting the following two features of transcendental consciousness. The first is that Kant conceives of self-consciousness as an *activity*. Rather than equating transcendental apperception with the *consciousness of* an activity, as the introspectionist model would predict, Kant equates apperception directly with a kind of activity, what he calls an 'act of understanding' (*APV* 7:142).[26] This characterization of transcendental apperception as an *act*, and not just as an act, but as a *spontaneous* act,[27] is very helpful, for it provides us with the resources to characterize the notion of self-consciousness in metaphysical terms that make it clear that Kant did not conceive of it, first and foremost, in representational or semantic terms. In his metaphysics lectures, Kant characterizes an act as that which grounds the inherence of an accident in a substance (*MM*; 29:773). An act or an activity is therefore an *actualization* of a power of a substance, and a *power* is the relation of an accident to a substance. On the view I will develop, we can helpfully apply this metaphysical framework to the theory of transcendental apperception. I will argue that we should conceive of transcendental apperception

[26] In the first *Critique*, too, Kant calls it an "act of the self-activity of the subject" (B130).
[27] Note that Kant also calls inner sense an act, but he nevertheless thinks that it is an act of receptivity, and therefore passive. I am interested here in the status of apperception as a spontaneous act. I provide a much more detailed analysis of the notions of activity and spontaneity in Chapter 3; for now, I just want to provide some brief orientation for conceptualizing an alternative to the 'introspection' model that I want to reject.

as an act in Kant's technical sense: it is the actualization of a certain power, which, together with certain other powers, comprises the understanding. We should conceive of the understanding as a power that relates certain accidents — *thoughts* — to a substance, namely, the thinking subject. Transcendental apperception is the act through which the power of understanding is actualized, and it is in that sense a condition of the possibility of the understanding (B131; B134 fn).

Second, Kant insists that transcendental consciousness is *formal* (cf. B411, A685/B713). It constitutes the *form* of thinking and is not to be identified with thinking itself. Now, a full discussion of Kant's theory of form is well beyond the scope of the present overview of the position that I wish to defend, but it is nevertheless helpful to say something very briefly about what the form of thinking might be. As I understand him, Kant conceives of form in general as a principle of organization, and therefore as a principle that makes possible a particular kind of coherence and unity among the parts of something. The form of thinking is thus a principle that makes possible a certain kind of unity among our representations, and, in particular, the unity that is required for thinking. This helps us to see that while transcendental apperception, as the form of thought, *structures* thinking, it is not itself a thought. While thinking proper might always be about something, the *form* of thinking, as a condition thereof, is not already related to objects in the way that thought would be. We might say then that transcendental consciousness is a necessary but not sufficient condition for intentionality, for while it makes thinking about objects possible, it is not itself already related to any object. Thus, although it is not consciousness *of* anything, it *is* the form of the consciousness of everything.

This formal conception of pure apperception lies in the background of Kant's claim that pure apperception is 'empty'. In the following passage, Kant denies that pure apperception

contains any 'matter or content' to think about, and calls the study of apperception in abstraction from this external content or matter the science of *logic*:

> This difficulty rests entirely on a confusing of inner sense (and of empirical self-consciousness) with apperception (intellectual self-consciousness), which are usually taken to be one and the same. The I in every judgment is neither an intuition nor a concept, and not at all a determination of an object, but an act of understanding by the determining subject as such, and the consciousness of oneself; pure apperception itself therefore belongs merely to logic (without any matter and content). (*APV* 7:142)

In this passage, Kant denies that pure apperception is a 'determination of an object', and we can now see more clearly why: transcendental apperception is not the determination of some particular object but instead constitutes the form common to *all* object-related consciousness (cf. B137). I take Kant to be making this same point in his repeated denial that any intuitive manifold is given through transcendental apperception (see, for example, B139); transcendental apperception is formal and 'empty' because it does not by itself provide the understanding with any content to think about; rather, this content must always be supplied by an external source (B135). It is simply the condition of the status of the understanding as a power to think anything at all.

As we will see, the interpretive approach that I will adopt sheds light on the close tie, to be described in much more detail in following chapters, between transcendental apperception and spontaneity. Kant draws a distinction between two kinds of activities, or two ways in which a power can be actualized: it can be actualized by a power external to it, and if it is actualized in this way, then the actualized power is passive;[28] but it can also actualize itself, or, as Kant says, be actualized "from an inner principle" (cf. *ML1* 28:285-6), and if it is actualized in this way, then the actualized power is *spontaneous*. Now, since the understanding is a spontaneous power, it must

[28] "All passivity <*passio*> is nothing more than the determination of the power of the suffering substance by an outer power." (*MM* 29:823)

be actualized from an inner principle; that inner principle, I will argue, is the spontaneous activity that Kant calls pure apperception.[29] I will therefore contend that we should not read the 'self' in 'self-consciousness' as indicating the object of self-awareness, but instead as indicating the *origin* of this consciousness within the thinking subject itself rather than from an external source. It is best understood as functioning in the same way as the term 'self' in Aristotelean 'self-movement': to indicate a certain distinctive kind of spontaneous activity, namely, one that originates from *within* rather than from *outside* the actor.

To summarize: I will develop an approach that sharply differentiates pure and empirical apperception. Whereas empirical apperception is helpfully viewed in intentionalist terms,[30] I will develop an account of pure apperception that dispenses with the intentionalist model by denying that pure apperception has any intentional object. Rather than viewing pure apperception in representational or semantic terms, I will argue for a metaphysical reading on which pure apperception is an *act* of the subject that realizes the faculty of understanding. This act, I will show, constitutes the form of all thinking and accounts for the spontaneity of the understanding.

Section 2: Varieties of Reflection Theory

In this section and the next, I review the ways in which intentionalist assumptions have pervaded scholarship on Kant's theory of apperception. In the present section, I focus on an interpretive tradition initiated by Henrich's (1982 [1966]) seminal study of Fichte's theory of self-consciousness, in the course of which he argues that Kant is committed to a philosophically

[29] It is notable that Kant talks, at least of transcendental apperception as a unifying *principle* of the understanding – more on this in Chapter 4.

[30] Within this intentionalist framework, interesting questions are left over about whether the intentional object of empirical apperception shuld ultimately be thought of as internal to the subject. See Newton (2019) for a non-reflexive, 'transparent' account of empirical apperception.

problematic view of self-consciousness he calls the 'Reflection Theory'. In section 2.1, I explain the Reflection Theory and bring out its intentionalist assumptions. In section 2.2, I turn to a recent critique of Henrich's reading of Kant, due to Karl Ameriks (1995). Employing a distinction introduced by Manfred Frank (2004) between 'egological' and 'non-egological' Reflection Theories, I show that Ameriks' rejection of the Reflection Theory in fact carries forward its key intentionalist assumptions.

2.1. Henrich and the Reflection Theory

In Dieter Henrich's (1982 [1966]) seminal paper, "Fichte's Original Insight", he argues that Fichte's philosophy marks the advent of a new stage in the development of the theory of self-consciousness. The development consists in the rejection of the prevailing 'Reflection Theory' of self-consciousness, common both to Kant and his predecessors, according to which "the essence of the Self is reflection" (19). Before providing a brief overview of the challenges that, according to Henrich, Fichte brings to bear on the Reflection Theory, it is worth briefly discussing the nature of that theory and to make explicit the sense in which it is an intentionalist account of self-consciousness.

According to Henrich, the Reflection Theory holds that the Self (or its essence) is identical to the very act by which it is conscious of itself.[31] It is essential to the Reflection Theory that self-consciousness is achieved through an act whereby a subject turns back on itself to make itself an object of awareness. Through this act, the identity of the Subject-self with the Object-self is established, and it is in the subject's grasp of this identity that self-consciousness consists. Henrich

[31] Instead of 'Self,' we can also talk here simply of a *subject*, or of subjectivity, since the issue concerns the nature of the subject, and the aim is to provide an account of what subjectivity consists in.

claims that what is common to Kant and his predecessors is their conception of self-consciousness as having this reflexive structure:

> This theory begins by assuming a subject of thinking and emphasizes that this subject stands in a constant relationship with itself. It then goes on to assert that this relationship is a result of the subject's making itself into its own object; in other words, the activity of representing, which is originally related to objects, is turned back upon itself and in this way produces a unique case of an identity between the activity and the result of the activity. (19)

Bracketing for now the question of whether this view is in fact Kant's, I want to note the structural feature of this type of account of self-consciousness to which Henrich is drawing our attention. The *reflexive* structure of the act of self-consciousness is the result of the intentional relation between the self as representing subject and the same self as object of representation. As Henrich notes, the intentional relation that is established through the act of representing is 'originally' between the representing subject and the represented object, where the representing subject is not identical to the object represented. What makes self-consciousness unique is exactly the identity of these two: the representing subject is also the represented object and that is, on this view, the very essence of the representing subject. Note that on this account there is a more primitive activity than self-consciousness: the activity of representing 'is originally related to objects' and we understand the structure of self-consciousness in terms of this more primitive activity – indeed, as a special case of it. The Reflection Theory is thus a paradigm case of an intentionalist account of self-consciousness.

According to Henrich, Fichte was the first philosopher to recognize insuperable difficulties in this kind of account. The basic thrust of Fichte's criticism can be expressed as a dilemma: either the reflection theory is circular, or it collapses into a reductive position, that is, it fails to capture the subjectivity that is genuinely distinctive of self-consciousness by 'reducing' it to object-

consciousness.[32] The dilemma arises fairly straightforwardly once we appreciate the fact that on this view the self *just is* the very act of becoming aware of itself. Thus, it seems that we must have presupposed some self that is to become the object of awareness. What is the status of this presupposed self? If the self *qua* object of awareness is to be *the same self* that becomes aware of itself, then the theory is circular: it presupposes exactly what it was designed to explain. But if the self *qua* object of awareness is to be some self *other* than the subject-self, then we have abandoned the reflection theory, for it is not *itself* that it becomes aware of, and there is nothing at all *distinctive* about *that* self as object of our awareness; instead of explaining subjectivity, we have explained it away. Henrich's Fichte also thinks that there is an insidious tendency to start with the reflection theory and then, given the problem of circularity, to slip into this reductive position in which the subjectivity that makes self-consciousness distinctive is then explained in objective terms that fail to capture the essence of the self:[33] "In Fichte's opinion everyone who falls victim to [this circle] makes the mistake of representing the Self merely as one object among others" (23). According to Henrich, Fichte reasons as follows:

> The reflection theory does indeed begin with a Subject-Self, but it then proceeds to think of it only as a force capable of acting upon itself. With this the theory gives up the distinctive sense of subjectivity that belongs to self-consciousness. The latter is interpreted instead in terms of a matter of fact activity that really belongs to the sphere of objects. (22)

According to Fichte, once we are in the grips of the Reflection Theory, we rid the subject self of any of the distinctive features that make it a subject rather than an object. Representing the subject-self as a 'power or force' of representation, we apply to it the same conceptual framework that we

[32] By subjectivity here, I will understand something like the distinctive subjective character that Nagel (1974), for example, identifies with the 'what it is likeness' to be a subject in a certain state.

[33] I am not interested here in the broader question of theory reduction that is an important concern in a wide range of philosophical discussions, including the nature of the mind. Here, I will call a theory of self-consciousness *reductive* if it takes self-consciousness to be explicable in intentionalist terms – that is, in terms of the object to which the consciousness is directed.

would apply to any object, and thus the account lacks the resources to distinguish the activity of the subject self from the 'matter of fact' activities that characterize 'the sphere of objects'. Consequently, instead of explaining the subjectivity that makes self-consciousness distinctive, the Reflection Theory explains it away. It is therefore, in Fichte's view, wholly inadequate.

According to Fichte, the very possibility of an act of reflection – the act by which the self turns back on itself and makes itself its own object of thought or determination – presupposes a kind of "primordial selfhood." This "primordial but obscure essence of the Self" (22) then also constitutes the starting point of and the impetus to philosophy. Thus, Fichte aims to escape from the dilemma by introducing a primordial, pre-reflexive, irreducible subjectivity as the starting point of philosophy.

Henrich's work sparked a research program in German philosophy, namely, the development of an historically well-grounded, philosophically plausible non-reflexive account of self-consciousness.[34] Rather than exploring this tradition any further, what I want to note for now is that an assumption built in at the foundation of this tradition is that Kant himself signed up to the reflection theory and embraced its intentionalist implications. Thus, Henrich's telling of the history of the philosophy of self-consciousness has played a significant role in shaping the intentionalist assumptions of contemporary Kant scholarship. In the next section, I will show how, through the influence of Strawson, the Anglophone tradition has also developed an intentionalist bias in its approach to the theory of apperception. Before I do so, however, I want to pause to consider an important attempt, due to Ameriks, to break with Henrich's interpretation of Kant and

[34] Frank (2004; 1991), Pothast (1971), and Cramer (1974) are among the most influential contributors to this program. They are sometimes referred to collectively as the Heidelberg School after the seminars that Henrich taught in Heidelberg. Cf. also Tugendhat (1979) for criticism of their approach. Zahavi (2007) provides a helpful overview of the contribution of the Heidelberg School to the philosophical landscape in Germany.

show that Kant in fact embraced a far more Fichtean account than Henrich's narrative allows. What we will see is that, for all its importance, this account stops short of parting with the intentionalist presuppositions of the Reflection Theory.

2.2. Against the Reflection Theory

In a series of important papers, Karl Ameriks (1983, 1995, 1997, 2012) has argued that Kant's position, while more moderate than the one Fichte develops in response to the problems of the Reflection Theory, nevertheless *is* sensitive to those problems and can avoid them. Specifically, Ameriks takes Henrich to place an 'extravagant' and unwarranted demand on Kant's theory, for Henrich reads Kant as attempting to provide an account of the *origin* of self-consciousness. Ameriks argues that it is this demand that motivates Henrich's (and Fichte's) criticism of Kant's theory of apperception as reflectionist; once we withdraw this demand, he argues, a reflectionist reading is no longer forced on us. Ameriks also wants to provide a positive argument for the claim that there is a level of self-consciousness to be discerned in Kant's theory that is pre-reflexive, even if the account falls short of providing an explanation of the *origin* of self-consciousness of the kind that Fichte was seeking.

Kant famously claims that the "**I think** must be able to accompany all my representations" (B132). For all the attention that this passage has received, Ameriks (1983) points out that little work has been done to determine exactly *what* it is that the 'I think' is supposed to accompany. Ameriks contends that Kant distinguishes between several different 'levels' of consciousness, and that it is not immediately clear that Kant's claim at B132 applies to all of them. These levels of consciousness can be distinguished in terms of the increasing complexity of the representational states involved. According to Ameriks, Kant distinguishes the following:

> [1] lower level states ('in' a being) in which a subject may represent something but *without* sensing that he is doing so; [2] higher level states ('for' a being) such that one is simply aware *that* one is having representations; and even [3] higher level states ('cognitively for' a being) which occur when one can claim to know something objective through a representation (of course, even [4] higher level states could be distinguished, such as the reflexive states in which one understands that one is having a representation that provides objective knowledge. (Ameriks 1983, 183-4)

Ameriks claims that the states at the very first level are those that Kant characterizes, at B132, for example, as being 'nothing for me.' The states at the second level, since they *are* something for me, must already have a 'personal quality', a kind of 'mineness' or quality of belonging to me, and thus already constitute a rudimentary kind of self-awareness. This quality is best explained, Ameriks suggests, by analyzing these representations as "*implicitly* of the form 'I think that so-and so…'" (1983, 239; italics in original).

Ameriks wants to use this taxonomy of states to explain what it is that the 'I think' accompanies. He argues that the "*transcendental* 'I think'" applies at the second level:[35]

> Representations that are not at the level of 'nothing to me,' and yet are also not themselves such as to provide objective knowledge would fall in the middle category [the second level] here, and I take it that these are the most likely subjects of Kant's remarks [at B132]. (Ameriks 1983, 184)

Ameriks calls the second level of consciousness *empirical* and the third *transcendental* apperception. Accordingly, empirical apperception will consist of the set of states, individuated by contents $x_{1\ldots n}$, which implicitly have the form 'I think that x,' for any $x_{1\ldots n}$. We can call these states *thoughts* in virtue of their implicit form. Transcendental apperception is thought about that set of thoughts, or at least some subset of those thoughts, such that its structure is best understood as follows: 'I think that (I think that $x_{1\ldots n}$)'. Notice that this makes transcendental apperception

[35]Ameriks uses the phrases 'transcendental unity of apperception', 'transcendental apperception' and 'transcendental "I think"' interchangeably.

explicitly *reflexive*: it is thought *about* thought.[36] It is also important to note that Ameriks takes the doctrine of transcendental apperception to "include the claim that the I which thinks that $x_{[1]}$' is the same I which thinks any of $x_{2...n}$, and is the same "I which thinks *that* I think that x, etc." (Ameriks 2012, 240). That is to say, transcendental apperception "implicitly includes the claim (which may or may not be a correct claim) that all the uses of 'I' within it are co-referential" (Ameriks 2012, 240).

The first step in seeing Kant as committed to a kind of self-awareness that is pre-reflexive involves rejecting what Ameriks (2012) calls the *Strong Apperception Thesis* (SAT). SAT is a thesis specifically about how consciousness at the second, empirical, and third, transcendental levels, are related: It holds that transcendental apperception is *required* for empirical apperception (241). Ameriks claims that SAT is motivated by reading Kant's doctrine of transcendental apperception as providing a compelling answer to the following question: How can the implicit form of the thoughts at the empirical level explain why they all belong to one (unified) consciousness? The answer that SAT allows Kant to provide is that we must be able, in principle, to attach the transcendental I think to the set. If 'I' is co-referential both horizontally across empirical states and vertically between different levels of states, then we can explain the unity of any given set of thoughts by appeal to this *possible* transcendental apperception. As Ameriks points out, however, SAT involves a substantive psychological and philosophical commitment: it rules

[36] Here it is worth noting that there are at least two versions of the Reflection Theory. Frank (2007) draws the distinction between the two versions as follows:

> This [reflection] model implies that self-consciousness comes about either through a consciousness of the I as an object (egological version), or through a kind of higher-order consciousness which is directed towards a first order consciousness (non-egological version). In this way consciousness discovers itself in the position of an object. (2007, 157)

Ameriks' focus tends to be on the egological version of the theory, presumably because that is the view that Henrich most explicitly objects to. He nevertheless acknowledges that it is partly in virtue of an interpretation of transcendental apperception in overtly reflexive, albeit non-egological terms, that Kant is often read as accepting the Reflection Theory. More on this distinction below.

out that there could be beings that can have thoughts, but cannot have thought about thoughts.

As Ameriks points out, SAT is not *sufficient* for the claim that Kant holds the Reflection Theory; for that claim to be true, SAT must also be taken to be an explanation of the *origin* of self-consciousness. The Reflection Theory holds that the act of reflection is necessary and sufficient for self-consciousness to obtain, and that prior to such an act there simply is no self or subject of which to be conscious. Thus, it is only if we see transcendental apperception as that act which *produces* self-consciousness as such that Kant's theory can properly be understood as a Reflection Theory. According to Ameriks, we should not read Kant as aiming to answer the 'extravagant demand' to provide an account of the *origin* of self-consciousness. And if we don't read Kant this way, then there "is no clear reason to believe Kant's doctrine of apperception requires radical modification, in the way Henrich proposes, by something closer to Fichte's philosophy" (Ameriks 2012, 246).

Ameriks maintains, however, that his reading affords the resources to provide a positive account of Kant's position that is Fichtean in spirit. While the reading is not motivated by the Fichtean enterprise of providing an explanation of the *origin* of self-consciousness, it nevertheless identifies a level of self-awareness that is both *sui generis* and pre-reflexive. This kind of reading has the advantage that it remains close to the orthodox Kantian reading of transcendental apperception as a higher-order reflexive act. Nevertheless, Ameriks insists that this advanced level of self-awareness is not necessary for a more primitive self-familiarity that he argues characterizes the second, empirical level of apperception. On his reading it makes sense to regard "a simple 'I think that x' (e.g., "I think that it is warm") state as something that in one sense is *already* a kind of self-consciousness – since it directly expresses a thinking self's activity – even if it is precisely not a reflection upon a distinct 'object-self'" (2012, 249).

For our purposes, what is important to note is that Ameriks locates this pre-reflexive level of self-familiarity *below* transcendental self-consciousness in his taxonomy, at what he takes to be the level of empirical apperception (level 2). At the higher level of *transcendental* self-consciousness, though, it seems that the reflexive structure identified by the Reflection Theory is preserved, albeit with a reconstrual of the intentional object of self-consciousness as a set of thoughts (rather than, say, as an act of the self). To the extent that the general reflexive structure is preserved in the form of a higher-order consciousness, then, Ameriks' conception of transcendental self-consciousness is still intentionalist: it still conceives transcendental self-consciousness as a mode of consciousness that is related to a distinctive object, namely, a set of internal states (thoughts) or their contents.

Manfred Frank has called this kind of intentionalist model the *non-egological* conception of self-consciousness. According to Frank (2007), "'Self-consciousness'…can mean either (A) consciousness of the owner of the consciousness (the *subject* or *I*) or (B) consciousness of *mental states*" (2007, 154). Frank thus draws a distinction between *egological* and *non-egological* iterations of the Reflection Theory. An *egological* Reflection Theory holds self-consciousness to be consciousness of the subject or the self as an object; that is, it holds self-consciousness to consist in an intentional relation that the subject stands in with itself. By contrast, non-egological Reflection Theories take self-consciousness to be consciousness of the mental states or activities of a subject. On such a view, self-consciousness is a higher-order consciousness that takes as its object first-order states or activities of consciousness; thus, unlike on the egological version, the subject itself is not the direct intentional object of this consciousness. The Reflection Theory that Henrich ascribes to Kant, and that Ameriks rightly argues Kant does not accept, is a model of the first, egological kind of Reflection Theory. Yet in rejecting this model, Ameriks opts for the

second, according to which transcendental self-consciousness is a higher-order consciousness of empirical consciousness or apperception, which he understands as a set of thoughts.

As we will now see, it is this latter, *non-egological* version of the intentionalist model that has been most influential in the Anglophone tradition of Kant scholarship. In particular, the interpreters of Kant that I discuss below take him to be committed to a view according to which self-consciousness consists either (1) in consciousness of the activity of the understanding in applying the categories (synthesis), or (2) consciousness of the product of the application of the categories (thoughts or cognitions), or (3) consciousness of both the activity and the product. Despite their important differences, all of these approaches share the assumption that Kant was committed to a non-egological variant of the intentionalist model.

Section 3: Varieties of Self-Ascription Theory

In this section, I explore three of the classic positions that have done most to form the intentionalist starting-point of Anglophone interpretations of transcendental apperception, those of Allison, Kitcher, and Strawson. We saw in the Introduction that these three scholars have widely divergent accounts of what self-consciousness is; whereas Allison and Strawson are proponents of the epistemic approach, Kitcher pioneered the psychological reading of apperception. Nevertheless, for all of their divergences, they all conceive of transcendental self-consciousness as consisting in an awareness either of an activity of the understanding or of the product of this activity.[37] Thus, all three authors accept some version of the intentionalist model that Frank calls the non-egological Reflection Theory.

[37]Now, my claim here that *Kitcher*, like Strawon and Allison, endorses an intentionalist view of transcendental self-consciousness should come as a surprise given my presentation of her view in the Introduction, where I showed that she denies that transcendental self-consciousness is reflexive. But

The aim of this section is to illustrate the way in which all of these views share a common structure. All three views, I will show, are variants of what we can call the *self-ascription theory* of self-consciousness. The main interpretive claim that distinguishes the self-ascription theory is that it is in virtue of the reflexive structure of transcendental self-consciousness, the subject's awareness of her synthetic activity or the products of this activity, that a certain kind of *unity*, *viz*., the synthetic unity of apperception, is achieved. According to this kind of interpretation, the activity of the understanding generates such a unity through the application of the categories to the manifold of sensibility, which Kant calls synthesis or transcendental synthesis. Transcendental synthesis effects a unity insofar as the subject is, in some sense to be specified, aware or conscious of herself as engaging in this synthesis of the manifold, for she is then in a position to think of the contents generated by this act as 'belonging to her', or as *hers*, and hence to self-ascribe those contents.

3.1. Strawson

In 1966, the very same year that Henrich published his article on Fichte's original insight, P. F. Strawson published *The Bounds of Sense*, arguably the most influential work of Anglophone Kant scholarship of the mid-20th century. In that work, Strawson sets out to reconstruct an 'austere' set of transcendental arguments from the first *Critique*, which he thinks are independent of the psychological idiom in which Kant expresses them. According to Strawson, one of the most

whereas in the Introduction I was discussing Kithcer's influential *early* work, in what follows I will be discussing the *second* and more recent of her two monographs on Kant's theory of apperception, and one of the main points on which she changes her mind across these two monographs is precisely on the question whether transcendental self-consciousness is reflexive. Thus, in what follows, the reader should take it as implicit that when I discuss Kitcher's views, I am referring to the more recent account spelled out in her monograph *Kant's Thinker* (2011).

important of these arguments is to be found in the Transcendental Deduction, and concerns Kant's theory of transcendental apperception.

Despite the fact that Strawson takes transcendental psychology to be an 'imaginary subject' (1966, 97), he nevertheless takes seriously the claim that there is a "necessary connexion, by way of synthesis[,] between the unity of consciousness … and the relation of representations to an objective (empirical) world" (1996, 96). Thus, he thinks that the notion of transcendental apperception or self-consciousness, if properly understood, plays an important role in the Deduction, and he aims to explain the connection between self-consciousness and objectivity without relying on (what he takes to be) the dubious aspects of Kant's doctrine of synthesis.

Strawson takes Kant's notion of the transcendental unity of apperception to underlie the subject's capacity to apply a set of concepts (the categories) to the objects that she experiences, and, moreover, he takes the application of these concepts to be necessary for a subject to be able to distinguish herself from a world that exists independently of her experience of it. Since he wants to bypass the entire theory of synthesis, he cannot explain the relationship between the unity of consciousness and objectivity by appealing to facts about the synthetic process. Instead, he wants to "establish a direct analytical connexion between the unity of consciousness and the unified objectivity of the world of our experience" (1996, 96).

On Strawson's view, the role of transcendental self-consciousness is to explain exactly the basis of the possibility of self-ascribing mental states (what he calls *experiences*). He insists that the basic structure of self-consciousness is to be explained in terms of the conditions under which it is possible for a subject to ascribe a set of diverse experiences to herself:

> The notion of a single consciousness to which different experiences belong is linked to the notion of self-consciousness, of the ascription of an experience or state of consciousness to oneself. (98)

His basic argumentative strategy is then to ask under what conditions it is possible for a subject to call a state *hers*, that is, to self-ascribe a mental state. Most fundamentally, he thinks that this would require that the subject is conscious "of the identity of that to which these different experiences, at different times, belong" (96).

Note, however, that Strawson's conception of the transcendental unity of apperception is not that it provides a criterion for the identity of the subject; in fact, he takes it to be a condition for the self-ascription of mental states that *does not require any criterion of identification of the subject at all*:[38]

> When a man (a subject of experience) ascribes a current or directly remembered state of consciousness to himself, no use whatever of any criteria of personal identity is required to justify his use of the pronoun 'I' to refer to the subject of that experience. (165)

It is this use of the 'I' that is operative in the representation that is produced by the transcendental unity of self-consciousness, the 'I think'.

None of this is to say that Strawson treats criteria of self-identity as irrelevant to self-consciousness. Strawson draws a distinction between the transcendental unity of self-consciousness (upon which the self-ascription of mental states depends), and the self-consciousness of an empirical subject; he takes the latter, unlike the former, to provide criteria for personal identity, insofar as the empirical subject's consciousness of herself requires that she can locate herself in a spatio-temporal order (164). Moreover, he takes the 'I' of the 'I think' to refer to the empirical subject of experience despite the fact this 'I' is used without criteria of subject

[38]This feature of Strawson's account has been influential beyond Kant scholarship. In Sydney Shoemaker's landmark paper, in which he introduces the notion of immunity to error through misidenticiation, which has become a mainstay of contemporary discussions of self-consciousness and self-reference, one of his main interlocutors is Strawson (see especially Shoemaker 1968: 564-67).

identity, because, he says, "even in such a use, the links with those criteria are not in practice severed" (165).

Strawson clearly thinks about transcendental self-consciousness and the unity to which it is supposed to give rise in terms of self-ascription, but I haven't yet made it clear in what sense his view should be conceived of as a variety of the intentionalist view that Frank calls the non-egological Reflection Theory. We can see this more clearly in Strawson's commitment that "'transcendental self-consciousness' is associated with the *activity* of the faculty of understanding". In particular, Strawson claims that Kant thinks of self-consciousness, first and foremost, in terms of a consciousness of this activity:

> The key to the unity of consciousness, it seems, is to be sought in the fact that the connectedness of our perceptions is *produced* by the *activity* of the mind. The process of producing such connectedness or unity is called synthesis; and our consciousness of the identity of ourselves is fundamentally nothing but our consciousness of this power or synthesis, or combination, and of its exercise. I can count a given representation as *mine* solely because *I* have combined or synthesized it with others. (94)

However, at least partly because Strawson is skeptical about transcendental psychology, he thinks that we should focus on "the outcome of the synthesizing activities rather than on any special awareness of those activities themselves or of the powers exercised in performing them" (96). Strawson thus argues that we can get a better grasp of the consciousness of this identity by focusing on the necessary connections between experiences that are produced by the synthetic activity of the understanding, rather than the activity itself. It is only by recognizing such necessary connections, he argues, that the subject can distinguish herself from her experiences and therefore recognize an *objective* world of experience.

It is important to note, however, that Strawson's view is not quite so straightforward, since he takes the self-ascription of mental states and experiences to require an *empirical* and not merely

a transcendental criterion of identity. In particular, he takes self-ascription to require the ability of the subject to identify herself as an object in space and time. Transcendental self-consciousness, though insufficient, is nevertheless necessary for self-ascription: "transcendental self-consciousness is not to be *identified* with the possibility of empirical self-ascription of experiences. But it must be recognized as the basic condition of that possibility. And it is shown to *require the fulfillment of the objectivity condition*" (108; my emphasis). In fact, Strawson says that the label of 'self-consciousness' is a misleading one, for, according to him, the objectivity condition is met not by ascribing experiences to a subject, but by the fact that experience has a certain kind of reflexiveness that allows for it to incorporate a distinction between 'the order of the world' and 'the order of experience'. It is this "reflexiveness *which is expressed by Kant in terms of the notion of self-consciousness*" (107); the label is unfortunate because "we are immediately led by it to think in terms of *personal* self-consciousness" (107). Strawson nevertheless takes talk about self-consciousness to be legitimate because the possibility of this distinction within experience between the way things are and the way they are experienced as being is the ground of the possibility of the self-ascription of experiences.

As I read him, his view is that the transcendental unity of apperception, that is, the criterionless self-ascription of mental states (cf. 98), requires the recognition that experience has a kind of 'self-reflexive' structure. In particular, reflection on the *product* of the synthetic activity of the understanding reveals that experience has a self-reflexive structure that consists in the fact that the very possibility of making sense of experience *as* experience requires a distinction between objectivity, how things are independently of our experience, and how they are experienced as being (cf. 107). This latter distinction requires the application of the categories, or 'concepts of the objective' (111) as Strawson calls them. Thus, Strawson's account of transcendental apperception

clearly qualifies as a variant of non-egological intentionalism, on which the intentional objects of self-consciousness are mental states, themselves conceived of as necessarily connected.

Before we proceed, I want to flag one final feature of Strawson's position that will become important later on. It is important to note that the 'doubleness' or duality of experience that Strawson takes transcendental self-consciousness to express, namely, between how things are in the world that we experience and how things are in experience, is the result of the application of the categories (his 'concepts of the objective'). Thus, for Strawson, the same procedure for becoming conscious of the world (recognizing it as independent of our experience of it) also produces self-consciousness in his minimal sense. The procedure consists in the exercise of our conceptual capacity, and in particular in the application of the categories, for the duality is "implicit in the concepts employed in experience" (107). Thus, on Strawson's interpretation, Kant's conception of self-consciousness can be fully explained in terms of the requirements that must be met to satisfy what he calls 'the objectivity condition' (namely, the condition that experiences have 'an objective reference' (89)).

3.2. Kitcher

In her influential recent study of Kant's theory of the cognitive subject, which builds on and departs from the earlier work discussed in the Introduction, Patricia Kitcher (2011) similarly defends the version of the intentionalist model that I have called the self-ascription view. On her view, Kant's central insight in the Deduction is that the transcendental unity of apperception, which she takes to consist in the recognition by the subject of her identity, both requires and is a requirement of cognition of objects. According to Kitcher, self-consciousness is "the consciousness of a subject's identity" (138) and this is at the same time a "consciousness of the

synthesis of appearances according to concepts or rules" (ibid). These two forms of consciousness are necessarily connected because the subject's consciousness of her identity is made possible by her awareness of 'real bonds' or necessary connections between her mental states. These bonds are engendered by conceptual synthesis, and they introduce a kind of mental unity, which unity is just what constitutes the criterion of the subject's identity on Kitcher's view. Thus, on Kitcher's view, "synthesizing produces the unity of apperception" (ibid.), and this very activity of synthesizing is also required for object cognition. If we take seriously Kant's requirement that representations must be brought to the unity of apperception in order to constitute cognitions, Kitcher argues, then "the synthesizing that produces the unity of apperception can take place only when synthesizing also produces object cognition" (114). Thus, Kitcher endorses a version of non-egological intentionalism: Consciousness of the identity of the subject consists in consciousness of the process by which necessary connections between representations are generated.

Above, we saw that Strawson sees a deep connection between object-consciousness and transcendental apperception, since he holds that the requirements on transcendental apperception are precisely those that must be met in order to satisfy the 'objectivity condition' on experience. This same commitment is explicit in Kitcher's position, for she claims that the very synthetic process of which we are conscious in transcendental apperception also produces object-cognition: "How can the necessary connections across representations be forged (and recognized)? In Kant's view, the only possibility is through the rule governed activity of thinking that produces object cognition. Through that activity, the unity of consciousness is created and recognized" (137). Kitcher reads Kant as responding directly to Hume's skepticism about a continuing subject underlying our mental states. According to Kitcher, Hume's skepticism is the result of the fact that he denies that any real necessary connections exist between mental states. Kant's innovation lies

in recognizing that the connections between representations that are required for us to have empirical cognition (experience) are exactly those required for self-consciousness. These connections are made possible by the understanding's activity of synthesizing given representations in accordance with necessary rules, that is, the categories. Thus, we must take the synthesis that makes object-cognition possible to be a necessary condition for the unity of apperception. As Kitcher puts it, "In this way the transcendental unity of apperception would be both a necessary and sufficient condition for object cognition" (114).

To summarize, Kitcher holds that the transcendental unity of self-consciousness yields a criterion for the identity of the subject over time. This criterion consists in a set of necessary connections between mental representations. Self-consciousness obtains when the subject grasps her identity by producing and recognizing these very connections. According to Kitcher, Kant's fundamental, anti-Humean innovation in the Transcendental Deduction is to recognize that the very same conditions that allow us to cognize objects, namely the application of a set of fundamental concepts that determine the persistence of objects through time (as, for example, Kant explains in the Analogies of Experience), also function as conditions for the (mental) unity of the subject. This allows us, upon introspection, to become aware of the identity of a persisting subject to which all these various mental states can be ascribed – contra Hume.

3.3. Allison

Allison (2004) reconstructs the argument of the B-Deduction as starting at B131 with the principle of the transcendental unity of apperception. On his view, "this principle states that the components of a complex thought must be connected in such a way as to allow for the possibility of their ascription to a single thinking subject, which entails that they constitute a synthetic unity"

(164). The self-ascription of a thought is possible only if the thinking subject is aware of its identity, for it is only the awareness of this identity that could impart on each of the individual components of the thought a status that secures its place within the whole. Allison describes this requirement as follows: "[i]f representations A, B and C are to be thought together in a single consciousness, then the I that thinks A must be identical to the I that thinks B, and so forth. In addition, if the subject is to be conscious of these representations as constituting a unity, then it must also be possible for it to become conscious of its own identity as subject with respect to the thought of each of these representations" (165). Allison takes transcendental self-consciousness to consist in the subject's ability to be aware of, or to think, its own identity. This is possible only through consciousness of the synthesis of representations (which takes place in accordance with the categories of the understanding). Allison notes that

> [t]he claim that apperception is possible only through *consciousness* of synthesis is … complex, because of the ambiguity that attaches to the term *synthesis*. It can refer to the act itself or to its product; and Kant apparently intended it in both senses. (170)

Thus it is clear that, for Allison, the subject's consciousness of her identity is explained by her consciousness of synthesis. Since this consciousness is *a priori*, it consists in a consciousness of the rules of synthesis, namely the categories. Unfortunately, Allison does not explain in detail how the self-identification is supposed to rely on the consciousness of these concepts. Instead, his main concern is to show that the mind's thought of its identity consists in the consciousness of a synthetic unifying act (cf. 171).

Thus, the conception of self-consciousness that Allison endorses is clearly intentionalist: he takes self-consciousness to consist in the awareness of a mental act, that is, an act of intellectual synthesis. As we saw above, any view on which a subject's consciousness of itself is mediated by a consciousness of its activity or state is a kind of non-egological Reflection Theory.

And once again, we see the same connection as has been emerging between transcendental apperception and object-cognition. It is important to note, on Allison's reading, that the activity of synthesis itself is what brings about the unity of a thought. It is consciousness of the act that brings about the unity of thought that grounds the subject's ability to think its own identity within the manifold of representation. Allison says that "[t]he consciousness of synthesis [specifically of a fundamental intellectual act of the understanding] is necessary if the mind is to be able to think its identity in the manifold of its representations" (171). This consciousness consists in an act of becoming aware of 'this identical I think', which Allison also holds "is the form of the act of reflection by means of which the mind grasps [not just its own identity, but] the identity in difference in the formation of general concepts" (172). On Allison's view, it is this very act of becoming conscious of the fundamental activity of the thinking subject, namely, the act of thought, which provides the basis of the subject's awareness of its identity across distinct thought-acts. Moreover, on Allison's view, the consciousness of the synthesis that constitutes transcendental self-consciousness is a consciousness of intellectual synthesis, which he identifies with discursive thought. Since intellectual synthesis is made possible by the application of the categories, Allison ultimately claims that the transcendental unity of self-consciousness requires the application of the categories of the understanding.

On Allison's (1996) view, "the crucial theme of the Deduction is the correlation between self-consciousness or the capacity to say 'I' (apperception) and the consciousness or experience of an objective, spatio-temporal world" (1996, 27), and he is credited with explicitly calling the relation between the transcendental unity of self-consciousness and the representation of objects one of 'reciprocity' (Schulting 2013). He claims that the fact "that these somehow mutually condition one another is the nerve center of the Transcendental Deduction," (ibid.) and he takes

the reciprocity thesis to be a commitment "shared by virtually every interpretation from the most metaphysical to the most analytic" (ibid.). The main motivation for Allison's proposal is that there needs to be a "necessary connection (actually a reciprocity) between the synthetic unity of apperception and the representation of objects" (Allison 2004, 173) simply in order for there to be a coherent (progressive) argument in the Deduction from the original unity of apperception to the application of the categories. Allison has consistently defended what is now referred to in the literature as the reciprocity thesis and in his recent monograph on the Transcendental Deduction he writes that "the synthetic unity of consciousness [apperception] is not only a necessary condition of the possibility of the representation of an object but also a sufficient one" (Allison 2015, 353).

3.4 Summary

Before providing some textual reasons to be dissatisfied with this kind of account, I will briefly summarize the main interpretive parallels that unite these otherwise disparate accounts of Kant's conception of transcendental self-consciousness. First, all of the authors above interpret Kant as committed to the view that self-consciousness is consciousness of either the activity or of the products of the activity of the understanding. Kant calls the activity in question *synthesis*, and synthesis itself is understood as a result of the application of the categories to the manifold of sensibility. Self-consciousness is then taken to be a requirement for the identity of the subject without which a subject would not be in a position to self-ascribe mental states (in Kant's terminology, to accompany those states with the 'I think'). Because all of these views share the basic commitment that self-consciousness is consciousness either of an activity or of the product

of the activity of the understanding, they fall squarely into the camp of non-egological intentionalist readings.

We also saw that each of the authors discussed above is committed to the claim that the application of the categories to the manifold of sensibility, which is required for consciousness of objects, is necessary for self-consciousness to obtain. I will call this the objectivity condition. Exactly how we should understand 'objectivity' in the context of the Deduction is a matter of controversy in the literature, and this is reflected in the different versions of the objectivity condition that each of the authors is committed to. What is not in dispute among these authors, however, is that the application of the categories is required for objectivity, for they are concepts of objects without which no objective thought and no cognition is possible. Thus, if objectivity is required for self-consciousness, then so is the application of the categories.

Section 4: Against Intentionalist Models

As the previous two sections have shown, the intentionalist model, primarily in its non-egological version, is a pervasive assumption in interpretations of apperception, one that unites readers in both the epistemic and psychological interpretive traditions. Yet there are powerful reasons for rejecting this model. In this section, I raise a fundamental exegetical objection to the intentionalist approach, centering on its commitment to the objectivity condition on apperception. As we will see, there is strong textual support for thinking that Kant in fact denied that self-consciousness is subject to the objectivity condition, instead conceiving of apperception as both prior to and independent of the categories of the understanding (and the representation of synthetic manifolds as *objects*). In order to reach this conclusion, I start by reviewing some of the evidence that has been put forth for supposing that Kant accepted the objectivity condition, and I show that

the passages in question accommodate an alternative reading. I then go on to show that there are strong textual grounds for doubting that Kant accepted the condition. Since my aim at this point is not to provide a systematic interpretation of Kant's positive view, but only to show that there is a strong textual basis for resisting the objectivity condition, and with it the intentionalist model, I will restrict my focus to central passages in the first *Critique*.

In his recent commentary on the Deduction as well as in his earlier work, Allison (2015, 2004) adduces two passages from the B-Deduction in support of the objectivity condition. Both passages occur in § 17; I address each in turn.

The aim of § 17 is to establish that "[A]ll manifold representation of intuition … stands under [conditions of the original unity of apperception] insofar as they must be **combined** in one consciousness" (B137). For the moment, it is useful to note that Kant's concern is not with the conditions of self-consciousness, but with the conditions under which the manifold of intuition must stand in order to be cognized.[39] He proceeds by spelling out the following condition on cognition: the manifold of intuition must be related to an object. Now the question that arises is whether self-consciousness is both necessary *and* sufficient for this relation to obtain. Allison believes that § 17 supports an affirmative answer to the question and hence concludes that self-consciousness is subject to the objectivity condition. He points our attention to two passages in particular and asserts that "the clear implication of both [passages] is that the synthetic unity of consciousness is not only a necessary condition of the possibility of the representation of an object but also a sufficient one" (Allison 2015, 352). In spite of claiming to find this commitment in both passages, however, he concedes that "it is more evident in the first of these passages, where Kant

[39] What exactly those conditions are is not important at this stage. It is a matter of considerable controversy in the literature. Influential accounts of how to understand Kant's conception of cognition and the relative roles of intuition and concepts are proposed by Allais (2015), Watkins & Willaschek (2017), Tolley (2013, 2017), Schaffer (forthcoming), McLear (2016).

claims that the synthetic unity of consciousness constitutes the relation of a representation to an object" (ibid.). I begin with the less direct passage:

> [1]The synthetic unity of consciousness is therefore an objective condition of all cognition, not merely something I myself need in order to cognize an object but rather [2] something under which every intuition must stand in order to become an object for me, since in any other way, and without this synthesis, the manifold would not be united in one consciousness. (B138)

Allison suggests that while Kant's claim in [1] is only that the unity of apperception is necessary for cognition, his claim in [2] suggests that that the categorical synthesis required for intuition to 'become an object for me' is a necessary condition for the unity of apperception to obtain. That, however, seems to be a misreading, for Kant's claim is only that such synthesis is necessary in order for the manifold to be united in the appropriate way to designate or be related to an object, not that it is a requirement for the unity of apperception as such. Interpreting the passage adequately would require a systematic reading of §16, and specifically an interpretation of what it means for a manifold to be united in consciousness. My own view, which I cannot fully defend here, is that a manifold united in consciousness is what Kant refers to in §15 as a 'combination', which is a *product* of pure apperception in relation to the given manifold of sensibility. This synthetic unity requires that the transcendental unity of apperception is *added* to the given synthetic manifold of sensibility and imagination. So, on my view, the unity of apperception itself does not depend on the synthetic unity that it generates through its relation to the sensible manifold. Therefore, when Kant makes claims about the conditions on a unity in consciousness, as in the passage under consideration, he need not be read as making claims about the conditions of transcendental self-consciousness *itself*. The point for now is not to defend or even fully elaborate this reading; I merely wish to point out that if it is viable, there is a straightforward way of reading the passage on which Kant is *not* stating conditions on pure apperception.

Regardless of our estimate of this first passage, Allison believes that a second passage provides clearer support for his reading:

> An **object**, however, is that in the concept of which the manifold of a given intuition is **united**. [a] Now, however, all unification of representations requires unity of consciousness in the synthesis of them. [b] Consequently [*Folglich*], the unity of consciousness is that which alone constitutes the relation of representation to an object, thus their objective validity, and consequently is that which makes them into cognitions and on which even the possibility of the understanding rests. (B137)

In this passage, [a] can again be read as an assertion of the claim that the unity of apperception is necessary for the relation of the intuitive representation to an object, but [b] seems to make the additional claim that the unity of apperception is *sufficient* for objective representation.

Now, while I think that this passage does provide stronger *prima facie* support for Allison's view than the previous one, it is not conclusive. What Kant actually says in this passage is not that the unity of apperception is sufficient for the relation of representations to an object but rather that it 'constitutes' (*ausmachen*) that relation, and it does not obviously follow from the fact that *x* constitutes *y* that *x* is a sufficient condition for *y*.[40] For example, it does not follow from the fact that statues are constituted out of rock that whenever you have a rock you thereby have a statue. Until Allison does more to tell us why in this special case the constitution claim entails the sufficiency claim, therefore, we are not obliged to take this passage as conclusive evidence that Kant subjected self-consciousness to an objectivity condition.

Having defused Allison's textual case for his reading, I now want to turn to consider some compelling evidence to suggest that Kant accepted what I will call the 'priority thesis', on which self-consciousness as well as the unity that it brings about — the unity of apperception — is prior to and independent of the application of the categories.

[40] Derk Perebook (2006) makes a related criticism of Allison's reading of this passage.

The first compelling evidence that Kant accepted the priority thesis can be found in the Paralogisms chapter. In the A and B-edition paralogisms, Kant insist that apperception or self-consciousness is the ground of the categories:

> 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 is therefore the representation of that which is the condition of all unity, and yet *is itself unconditioned* [*unbedingt*]. Hence of the thinking I… one can say **not so much** that it cognizes **itself through the categories**, but that it cognizes the categories, and through them all objects, in the absolute unity of apperception, and hence cognizes them [all objects] **through itself**. (A401)

> The unity of consciousness, *which grounds the categories*, is [in rational psychology] taken for an intuition of the subject as an object…[T]he subject of the categories cannot, by thinking them, obtain a concept of itself as an object of the categories; for *in order to think them, it must take its pure self-consciousness, which is just what is to be explained, as its ground*. (B422)

> For the inner perception is nothing beyond the mere apperception **I think**, *which even makes transcendental concepts possible*, which say "I think substance, cause, etc." (A343/B401)

The suggestion that Kant thought the categories are necessary conditions for the unity of apperception is contradicted by his explicit claim in the first passage that self-consciousness is unconditioned [*unbedingt*]. It is striking, moreover, that while in all three passages he explicitly claims that apperception or self-consciousness grounds the categories and even their possibility, he does not begin to suggest that the grounding relation is reciprocal or symmetrical. Instead, each passage emphasizes the priority and independence of self-consciousness as a fundamental condition of thought and of a relation of the subject to an object (or to all objects).

Further evidence that he accepted the priority thesis can be found in the opening sections of the B-Deduction, in which Kant first introduces pure or original apperception and its distinctive transcendental unity (B132). In section 15, Kant is concerned to describe the conditions for

combination [*Verbindung*] in general. I will return later to what exactly combination is, but for now it is useful to mention that combination is required in order to form a concept. The overall goal of the passage is to spell out what combination is and the nature of the unity that all combination requires:

> But in addition to the concept of the manifold and of its synthesis, the concept of combination also carries with it the concept of the unity of the manifold. [1] The representation of this unity cannot, therefore, arise from combination; rather by being added to the representation of the manifold, it first makes the concept of combination possible. [2] This unity, which precedes all concepts of combination *a priori*, is not the former category of unity (§ 10); for all the categories are grounded on logical functions of judgments, but in these combination, thus the unity of given concepts, is already thought. The category therefore already presupposes combination. [3] We must therefore seek this unity (as qualitative, §12) someplace higher, namely in that which itself contains the ground of the unity of different concepts in judgments, and hence the possibility of the understanding, even in its logical use. (B131)

The first thing to note is that the unity that Kant has in mind is required for any combination and therefore is not dependent on any combination — this is what I take to be his claim in [1]. Second, it is important to note that Kant explicitly rules out that the unity he has in mind is the category of unity that he introduces in the earlier metaphysical deduction of the categories, for he insists that combination is already presupposed by the functions of judgments and that these functions ground the categories. This I take it is his claim in [2], which concludes that combination is prior to the category of unity. [1] and [2] together yield the claim that the unity that makes combination possible is a unity that precedes the categories, [41] and so must be accounted for independently of them. And this is in fact what Kant claims in [3], for he says explicitly that this unity is the *ground* of the unity of concepts and their functions in judgment and that the possibility of the

[41] *Precedence* here should not be understood in temporal terms, but in terms of a logical or metaphysical priority.

understanding depends on this unity.[42] There does not seem to be any room for the claim that the unity that Kant describes here might in any way depend on the application of the categories.

If we now also accept the claim, which is altogether uncontroversial, that the unity that Kant describes in section 15 is the very transcendental unity that he ascribes to pure apperception in section 16, then the claim that self-consciousness (or its transcendental unity) depends on the application of the categories is also ruled out. Kant is, after all, quite explicit in his claim that the unity in question must be sought somewhere 'higher' than the categories, in a source that makes both the understanding and, with it, the categories possible in the first place.

Now, at this point, the intentionalist reader might attempt to modify her proposal by dropping the objectivity condition on self-consciousness. She might concede that self-consciousness is prior to the application of the categories, but still maintain that self-consciousness involves the subject being conscious of herself as doing something or other, or as being in some or other representational (mental) state, even while admitting that this consciousness does not consist in awareness of the process by which the categories are applied or of the product of that process, as the traditional intentionalist readers discussed previously have all maintained. If this alternative account of apperception could be made good, we would have a version of non-egological intentionalism that, by dropping the objectivity condition, could cope with the passages I have been discussing.

This proposal, however, faces serious obstacles. Consciousness *of* oneself *as* performing an activity or *as* in a mental state already consists in an intentional relation, but what is the status of this intentional relation? It cannot be brought about through an application of the categories, since if it were, the proposal would re-instate the objectivity condition, collapsing back into the

[42] Note that Kant does not say that this unity *is* the understanding.

view that self-consciousness necessarily requires an application of the categories. Therefore, the intentional relation in question must be one that is not mediated by an application of the categories, and the proponent of this view thus owes us an exegetically viable model of pre-categorical intentionality. Now, one might want to appeal here to the model of intuition, since on some readings of intuitions (such as Tolley (2013), Hanna (2011)), intuitions stand in an intentional relation to objects that is not mediated by any application of the categories. The problem, of course, is that assimilating apperception to intuition is in serious danger of falling foul of Kant's repeated insistence that the human intellect is not capable of intuiting, and his denial that any intuitive manifold is given through self-consciousness (see, e.g., B139). There is also a danger that appealing to non-cognitive modes of intentionality to explain self-awareness, such as intuition or perhaps feeling [*Gefühl*], would conflate a *consequence* of the act of self-consciousness with the act itself.

Perhaps this version of the intentionalist proposal could be worked out more comprehensively on a more careful textual basis, though I am highly skeptical. But regardless, another path is open to us, one that dispenses with intentionalism entirely and develops an alternative, non-intentionalist model for understanding apperception. This is the path I follow in this dissertation. In the next, final section, I say something briefly to anticipate the overall shape of the view that I will go on to develop.

Conclusion

The interpretation that I develop in the coming chapters breaks sharply from intentionalism in all its forms by denying that transcendental self-consciousness, as Kant understands it, is consciousness *of* anything, be that something the self as object (as per the Reflection Theory); a

collection of empirically apperceived thoughts (as per Ameriks); the product of synthetic activity (as per Strawson); the process of synthesis itself (as per Allison); or some combination thereof (as per Kitcher). We do ourselves a disservice, I think, if we assume from the outset that we even have a firm grasp of the phenomenon that Kant introduces with the term 'apperception' and that it can be understood in terms of our contemporary philosophical terminology and concerns. It is natural and tempting to hear the 'self' in 'self-consciousness' as the placeholder for some object of intentional awareness, and once we hear it that way, we have already signed up to the intentionalist assumption that characterizes scholarship on the topic. But I will argue, against this approach, that the term 'self' in 'self-consciousness', like the term 'self' in the phrase 'self-movement', serves, not to indicate an *object* of consciousness (or, in the self-movement case, an object that is moved), but rather a distinctive *origin* of consciousness (or, in the self-movement case, a distinctive origin of movement). Self-consciousness, on Kant's account, is a form of consciousness that issues in a unique way from within the subject of consciousness itself.

In developing the interpretive resources for spelling out this reading, I will draw heavily on Kant's conception of *spontaneity*. In canonical texts, Kant forges a tight connection between apperception and spontaneity: in section 16 of the B-Deduction, for example, Kant is at pains to emphasize that pure apperception is an 'act of spontaneity [*Actus der Spontaneität*]' (B132), and in section 15, Kant calls the representation of combination, which rests on apperception, an *Actus* of the subject's 'self-activity [*Selbsttätigkeit*] (B130). It is this very spontaneity or self-activity that also serves to distinguish the faculty of understanding from that of sensibility (A50/B74), and there is thus an intimate connection between apperception and the defining feature of the faculty of understanding.[43] Given this close connection between the two phenomena, apperception and

[43]This connection is widely recognized and exploited in the literature on Kant's theory of self-consciousness Cf. Banham (2008), Carl (1997). Engstrom (2006), and Kern (2006). As we will see, many commentators

spontaneity, we can pursue one of two options, depending on which phenomenon we take to be better understood. If we take ourselves to have a stronger handle on apperception than spontaneity, we can use our handle on the doctrine of apperception to gain leverage on Kant's account of spontaneity. But given my aims in this thesis, and my belief that the doctrine of apperception has been consistently misunderstood in the literature, I will pursue the alternative option, treating the doctrine of spontaneity as the key to unlocking Kant's theory of apperception. My strategy will be to develop a detailed interpretation of spontaneity, drawing on the significant volume of published and unpublished texts in which Kant addresses the topic, and then build to an interpretation of how exactly apperception is meant to contribute to or underlie this feature of the understanding.

The ultimate aim of my discussion of spontaneity will be to show, in Chapter 3, that Kant's doctrine of the spontaneity of the understanding requires us to conceive of the functioning of the understanding in *teleological* terms. This teleological conception of understanding will be key to my interpretation of apperception, for I will go on to argue in Chapters 4 and 5 that it is pure apperception that introduces the *telos* or end on which the understanding depends and that is the basis of the unity of the understanding. In the next chapter, I start the work of building toward my own positive account of spontaneity by arguing against a widespread interpretive approach that understands spontaneity in terms of the notion of cognitive agency.

recognize a deep connection between Kant's conception of the understanding as a spontaneous faculty of judgment and self-consciousness. Cf. for example, Ellis (2017), Kohl (2015), and Land (2018).

CHAPTER 2

Spontaneity and Normativity

In the previous chapter, I argued against a very natural and widespread model for understanding Kant's theory of apperception, which thinks of apperception as a particular mode of intentional consciousness. Against this intentionalist reading, I argued that transcendental self-consciousness, as Kant understands it, is not consciousness *of* anything; and I presented a *formalist* alternative to this picture that, drawing on Kant's faculty psychology, thinks of transcendental self-consciousness as a fundamental act of the subject that generates the form of thinking. But we do not yet have a full sense of what this formalist alternative amounts to or what reasons one might have, beyond dissatisfaction with the intentionalist model, for thinking it to be true. For the rest of this dissertation, my aim is to articulate and defend the formalist model, and my strategy, presented at the close of the previous chapter, is to build my reading of apperception on a reading of Kant's doctrine of the spontaneity of the understanding. The interpretation that we will work toward and articulate in Chapter 3 argues that Kant's conception of spontaneity introduces a teleological dimension into his account of understanding that has not been sufficiently understood and that is crucial to a proper interpretation of the doctrine of apperception.

In order to get to that point, though, we will need to start by making the negative case against existing readings of spontaneity, and the aim of the present chapter is to argue against a broad trend in the literature that identifies spontaneity with a kind of *agency* of the cognitive subject.[44] This kind of view, which I will call the *cognitive agency* reading, might seem like a natural view to adopt once we settle on the reflective structure of self-consciousness that, according to the self-ascription view, is at the center of Kant's theory of apperception. But in this chapter, I will press a serious objection to the cognitive agency reading, arguing that such a view relies on a reading of Kant's project in the Transcendental Analytic, which, for all its appeal, we should not accept.

My broader positive aim is to set up a model for thinking of the understanding that is fundamentally continuous with the Leibniz-Wolffian tradition to which Kant belongs. This tradition conceives of the understanding as a *faculty* or *power* of the soul. On this view, the soul is causally efficacious (that is, it represents, desires and feels) in virtue of its having certain faculties or powers. To consider the understanding as a faculty in this sense is to hold that it grounds the inherence of certain distinctive kinds of accidents — namely, *thoughts* — in a subject through certain *activities* that form part of its essence or nature. Having outlined this framework in the

[44] Let me note, right at the outset, that discussions of cognitive spontaneity in the literature on Kant can be divided into two broad categories. One approach takes Kant's conception of spontaneity and self-consciousness in the Analytic as its starting point and seeks to offer an interpretation of that doctrine in terms of the freedom and responsibility that are required for cognitive agency. Allison (1996), Boyle (2015), Kohl (2015), Land (2021), McDowell (1994, 1998), Valaris (2013), and to some extent McLear (2020a), all fall into this category. There is a second category of commentators, however, who discuss Kant's commitment to cognitive agency by reflecting on the extent to which he accepts some form of (indirect) doxastic voluntarism. These discussions do not focus on Kant's conception of spontaneity and self-consciousness but instead on his theory of '*Fürwahrhalten*' or 'assent'. See, for example, Chignell (2007a and 2007b) and Cohen (2013). For reasons that will become apparent below, my argument in what follows has a bearing only on the former of these two categories of discussion. I am interested, in particular, to dispute the claim that Kant's doctrine of the spontaneity of the understanding gives us any purchase on an account of cognitive agency insofar as it entails that we have the type of control over our doxastic, epistemic or even cognitive attitudes that would be required for this kind of agency.

present chapter, I will go on, in the next, to develop the model in further detail and show how it can help us to make sense of Kant's claim that the understanding is spontaneous. Within the scope of the present chapter, the model lends further support to my negative case against the cognitive agency view, for I argue that if Kant indeed incorporates this metaphysics of powers into his account of the understanding, then the normative epistemic framework that the cognitive agency view rests on is not exegetically viable.[45]

I proceed as follows: In Section 1, I provide a brief overview of the notion of cognitive agency. Despite some divergence, contemporary philosophers agree that cognitive agency is manifested through the control that we have over how we judge or what we believe, where this control is conceived of as a kind of freedom or self-determination. Much of the work of contemporary epistemology then lies in working out in what sense our beliefs are or could be under our control, or self-determined, and which norms govern the epistemic activities related to judgment and belief-formation.

In Section 2, I consider how this theory is applied to Kant's project in the Analytic of the first *Critique*. The basic framework for establishing such an interpretation is relatively easy to uncover since it draws on Kant's characterization of the understanding as a faculty of *judgment* and on his claim that this faculty is also *spontaneous*. Since judgment is exactly the activity that is meant to manifest our cognitive agency, and since spontaneity is closely related to Kant's conception of freedom, it is not difficult to find the basis for the cognitive agency interpretation in these fundamental and familiar Kantian commitments. I round out this section by drawing on Thomas Land's recent contribution on the topic to fill in the details of what such an interpretation

[45] Now, this is not to say that this framework cannot be made compatible with a cognitive agency reading, but I will argue that on the most natural and straightforward way of understanding the causal framework that this metaphysical picture gives rests on, it does not obviously leave room for any normative assessment of the way in which powers function. See McLear (2020a) for an alternative interpretation.

of Kant's project might look like. Through my discussion of Land's view, it will become clear that this interpretation of Kant's project in the Analytic relies on two important assumptions: 1) that Kant accepts what I will call the *epistemic* conception of judgment, according to which judgment is or involves the activity of deciding whether or not to assent to the truth of a proposition; and 2) that the laws of the understanding (those studied by general and transcendental logic) are *norms* according to which the subject regulates her judgment or belief-formation.

For the remainder of the chapter, I address each of these assumptions in turn. In Section 3, I argue that there is no clear textual basis for ascribing to Kant the epistemic conception of judgment; in fact, I argue, there are important textual and philosophical reasons for thinking that he accepts an alternative conception of judgment that I call the *semantic* conception. Finally, in Section 4, drawing on the work of Clinton Tolley, I will argue that the laws of the understanding are not norms that govern the activity of judging but that they are constitutive laws that make the understanding the faculty or power that it is. Here I begin to develop Kant's metaphysical conception of faculties or powers, and I argue that the presence of this conception in his theory of understanding strongly suggests that Kant did not accept the normative epistemic framework that the cognitive agency view presupposes.

Section 1: Cognitive Agency

To get a clearer sense of why commentators have found it so appealing to explain Kant's conception of the understanding in terms of cognitive agency, it is helpful to consider in more detail what cognitive agency is meant to be. Of course, it is a matter of dispute in the contemporary literature what exactly constitutes such agency (the question of what agency in general is, is no

less fraught), but there is nevertheless consensus about what we can think of as the minimal necessary but perhaps not sufficient conditions for it.

We can start by considering the following intuitive explanation of the phenomenon offered by Matthew Boyle (2011):

> A creature that can make up its mind is one that does not just perceive and react instinctively to perceptions; it can judge. It is one that does not just desire things and unthinkingly pursue them; it can choose. It is one that does not just habitually associate one thing with another; it can reason. These and other familiar contrasts hang together with the thought that rational creatures are distinguished by their capacity for a special sort of cognitive and practical self-determination, a capacity which makes their relation to their own mental lives fundamentally different from that of a nonrational animal. (2011, 1)

Boyle is suggesting here that we can start thinking about what constitutes our agency (practical or cognitive) by considering the ways in which we, as rational human beings, differ from nonrational animals. What sets us apart is meant to be captured, at least initially, by pointing to a "special sort of cognitive and practical self-determination". For Boyle, this ability is manifested in the practical domain through the choices we make, and it is manifested cognitively when we make judgments. The colloquial expression that is meant to capture our agency in general, the 'making up of one's mind', is understood as pointing to a kind of self-determination that might be even better captured by the expression "making up one's *own* mind" or "making up one's mind *for oneself*". What this expression is supposed to signal is that how one acts or how one judges is, at least to some extent, up to oneself and hence counts as an expression of one's agency — as something that we are responsible for and for which we can be criticized. The difficult philosophical issues are tied up, of course, in spelling out how and under what conditions my actions, and my beliefs and/or judgments are in the relevant sense 'up to me' or 'self-determined'. My focus in what follows will exclusively be on the cognitive side of this agency.

To hold that judgments are 'up to me' or 'self-determined' is to commit to the idea that I am in some sense *free* with respect to what I come to believe, and the dominant way of cashing out the operative notion of freedom is by appeal to the notion of *control*. Thus, I am free with respect to what I believe only if I have some sort of control over what I believe. Since the paradigmatic instance of this kind of control is encountered in our voluntary actions – those for which we are morally responsible – many philosophers have wanted to spell out the notion of control in the cognitive domain by giving an account of the way in which what I believe is subject to my will.[46] The debate is therefore typically framed in terms of the question of doxastic voluntarism: In what sense, if any, is belief *voluntary*?[47] For our purposes, however, it is sufficient simply to rule out that Kant's, or even, more broadly, a Kantian conception of agency would hold that my judgments are 'up to me' only if they are under the direct control of my will, since the will, on Kant's view, is not a cognitive faculty and plays no direct role in generating cognition.

---

[46] The paradigmatic case of this kind of control is, of course, exemplified by our overt voluntary actions. Thus, the contemporary debate focusses on whether belief is in the relevant respect like those actions. Audi (2001) claims, for example that

> Belief is profoundly analogous to action. Both are commonly grounded in reasons; both are a basis for praising or blaming the subject; both are sensitive to changes in one's environment; both can appropriately be described as objects of decision and deliberation, and belief can appear quite action-like when conceived of as formed by assent or by acceptance. (2001, 93)

Audi makes a very strong claim, but much of the literature on this topic is devoted to working out just how strong this analogy needs to be in order to support an 'ethics of belief'. Alston (1988), for example, argues that the idea of our having voluntary control over our beliefs in the right kind of way for us to be responsible for them, in the same way that we are for our actions, is quite problematic, and he specifically criticizes the idea that the model of intentional action is appropriate for modeling the kind of responsibility and control we have over our beliefs: "The inauguration of propositional attitudes simply does not work like intentional action" (1998, 278). Given that the actions that we are responsible for are under our direct voluntary control, is it also the case that our beliefs need to be under our direct voluntary control? Can they be under our voluntary control indirectly, and what would that mean, exactly? I will, for the most part, leave these questions undecided and focus on what I take to be a broadly Kantian conception of cognitive freedom according to which the will is not involved in our cognitive processes and, therefore, on which whatever kind of control we have is not meant to involve voluntarism.

[47] For a helpful overview of this debate, cf. Vitz (2008). See Hieronymi (2008, 370 - 372) for a helpful discussion of control, and Leon (2002) for a discussion of indirect voluntarism with respect to belief formation.

Rather than tying cognitive agency directly to the will, the kind of account that I am interested in is one that finds in Kant's account of cognition a distinctive *normative* dimension that provides us with the resources to make sense of our cognitive agency. According to a view like this, there is a set of norms that govern epistemic activities such as belief-formation, and these are norms that we recognize as binding on us. On this view, we exercise our agency insofar as we recognize these norms and aspire to regulate our epistemic activity in accordance with them.

Commentators who ascribe to Kant a view such as this aim to capture the relevant sense of agency by holding, as Boyle does, that "a rational intellect is characterized by a special sort of freedom", a freedom we possess because our cognitive activities and their outputs are the sorts of thing for which we are and can be held responsible and, consequently, for which we can legitimately be praised or blamed. They are, therefore, the kinds of activities that can be assessed in light of some set of standards or norms. We think of the rational constraints in question not as directly determining (causally) what a subject believes, but rather as normative demands whose force a subject would feel in virtue of her rationality. For our purposes here, it is not necessary to consider what these norms are; instead, I simply want to note that our commitment to such norms provides us with one way of making sense of the claim that our cognitive activities are up to us, or, as we might say, are the result of a certain kind of freedom. And this kind of proposal has a distinctively Kantian ring, since it appeals to exactly the kind of freedom — freedom as norm-responsiveness — that Kant thinks characterizes rational beings. For this reason, numerous commentators have been drawn to the suggestion that Kant's project in the first Critique (in whole or in part) is to articulate the norms that govern our epistemic lives and constitute our cognitive agency.[48]

[48] Several prominent commentators have suggested that the *Critique* should be read as providing a normative framework for our epistemic relation to the world. Of particular relevance here is, of course,

In the passage with which we began, Boyle locates our cognitive agency specifically in our ability to *judge*, and it is worth pausing briefly to consider what this ability consists in according to contemporary views. Boyle mentions three influential accounts of what judgment is, each of which is at pains to distinguish between judgment, on the one hand, and belief, on the other. The first is provided by Peacocke (1998), who defines judgment as "a conscious rational activity, done for reasons" (1998, 88), and goes on to claim that the content of a judgment is stored as a belief. Shah and Velleman (2005) define judgment as "a cognitive mental act of affirming a proposition" and contrast it with a belief, which is a state or an attitude rather than an activity. Cassam (2010) also emphasizes that judgment "is a mental action" whereas belief is not something that the agent *does* or *undergoes*. To summarize, these authors all agree that judgment and belief are very closely bound together, but they suggest that we should distinguish between judgment and belief insofar as the former is the *activity* of making up one's mind and the latter is (in some sense) the *outcome* of that activity. Boyle is concerned about this way of distinguishing judgment and belief since, he

John McDowell (1994), whose influential work interprets Kant as providing an account of intentionality that attributes a normative status to the contents of mental representations. McDowell is also impressed by the "spontaneity" of the understanding as providing a basis for a parallel with freedom in the practical domain (see fn. 6). More recently, Pollok (2017) draws on the concept of spontaneity to interpret Kant's theory in normative terms: "spontaneity, which is required for any judgment, means that the person who judges appropriates the content of, and can be held accountable for, that judgment" (2017, 67). This spontaneity shows up in judgment through the transcendental unity of apperception: "The 'I think' is not a second act, and not even a second-order act of our empirical mind. It rather stands for the imputability of this very same empirical consciousness of a tree" (2017, 64). Longuenesse (2017) shares this normativist conception of self-consciousness: "[u]sing ' I' in ' I think' is premised on nothing but the consciousness of a mental activity one takes to be one's own in virtue of the fact that one takes oneself to be accountable for the correctness of its contents and their connections" (2017, 7). Pollok goes further and claims that "Kant does have a systematic account of what it means for judgments to be normative. Kant's critical answers – in the plural – to the question 'How are synthetic judgments a priori possible?' amount to an original theory of normativity" (2017, 2). Finally, Allison (2004), argues that

> [transcendental] idealism involves a radical reconfiguration of epistemic norms, which goes together with an appeal to epistemic conditions. In other words, it serves as the epistemological counterpart of the shift from heteronomy to autonomy, which is generally recognized as the essence of Kant's "revolution" in ethics (2004: xvi).

Cf. also Brandom (2002); Land (2021); O'Neill (1989); Merritt (2009);Valaris (2013).

argues, it threatens to confine our agency to *judgment*, thereby undermining or diminishing our agency with respect to our beliefs, but this detail of Boyle's position is not relevant for present purposes. The point rather is that these authors express a general consensus insofar as they take judgment to be the process we engage in to arrive at a decision about whether or not to accept some proposition as true, a process that culminates in the acceptance or rejection of the relevant proposition.

Disregarding the philosophical question about whether and how to distinguish between judgments (*qua* process or activity) and belief (*qua* state), it is clear that on this picture judgments are *epistemic* activities (or, if we do not distinguish between judgments and beliefs, epistemic states). I will therefore call this conception of judgment, according to which judgment is the activity of deciding whether or not to assent to a proposition, the *epistemic conception of judgment*. Now, this epistemic conception of judgment is the standard view of what judgment is, not just in the contemporary literature but also in the literature engaging with Kant's theoretical philosophy (although the distinction between judgment and belief is not typically appealed to in Kant scholarship).

To sum up, then: the idea of cognitive agency brings with it the idea that we have some control over our belief-formation. There are different ways of spelling out this conception of control, but the way that seems on the surface to be most germane to a reading of Kant spells out the kind of control we have over our beliefs in terms of some set of norms that are supposed to govern our epistemic and cognitive activities. This approach in turn requires that we accept what I have called the epistemic conception of judgment, which ties judgment directly to belief-formation by identifying judgment with the propositional attitude that result from this process of 'making up one's mind'.

Section 2: Cognitive Agency in Kant

With this general overview of cognitive agency in mind, it is not difficult to see why it has been thought to offer an appealing interpretive approach to Kant's theory of the understanding, and in particular his conception of the understanding as *spontaneous*. Spontaneity is, after all, a certain kind of *freedom*, and the understanding is characterized as a 'faculty for judging' (most famously, at A69/B94). Putting these two points together, the thought is, very naturally, that Kant's characterization of the understanding as spontaneous is meant to capture the sense in which our cognitive judgments are up to us.[49] On this view, in emphasizing the spontaneity of the understanding as a faculty of judgment, Kant's claim is that, in judging, we make up our minds about what to think and believe, and his philosophical project in the Analytic is to provide an account of the normative constraints to which our thinking and judging are subject.[50]

More specifically, there are, I think, three broad textual motivations for applying the cognitive agency model in interpreting Kant's doctrine of the understanding. First, there is undeniably a deep connection between spontaneity and practical freedom, and this has led commentators to model cognitive spontaneity on the kind of freedom that we have, according to Kant, in the practical domain. In the practical domain, it is exactly the recognition of a norm or norms as binding on us as rational agents that expresses our freedom. Many commentators have thus proposed that we transfer this model to the theoretical domain, where transcendental

[49] Consider, for example, McDowell's influential statement of this thought:

> [J]udging, making up our minds what to think, is something for which we are, in principle, responsible—something we freely do, as opposed to something that merely happens in our lives. Of course, a belief is not always, or even typically, a result of our exercising this freedom to decide what to think. But even when a belief is not freely adopted, it is an actualization of capacities of a kind, the conceptual, whose paradigmatic mode of actualization is in the exercise of freedom that judging is (1998, 434).

[50] Cf., for example, Allison (1996) and also Pollok (2017).

apperception is said to facilitate the recognition of a norm or norms required for rational agency. Second, Kant's characterization of the understanding as a faculty of judgment is liable to impress a reader who approaches the text antecedently convinced that Kant operates with an epistemic conception of judgment, for now it will appear that the defining act of the understanding is a paradigmatic expression of our cognitive agency. Finally and relatedly, many commentators hold that Kant takes the understanding to operate under substantive cognitive and epistemic *norms*, embodied in the laws of the understanding itself, which he takes to be presented and studied by the science of logic.[51]

It is not surprising, then, that a number of prominent commentators have found in Kant's doctrine of the spontaneity of the understanding the resources to develop an interpretation of his account of cognition in terms of cognitive agency. On this view, self-consciousness is central to the capacity for judging because this kind of consciousness is not only the basis for self-ascribing contents; it is also an expression of the endorsement of the proposition that is thereby ascribed. More precisely, the act of endorsing is at the same time the act of ascribing to the self those contents, and since the endorsement takes place through the recognition of certain rules, it thereby confers normative significance on the act of self-ascription. Here are a few examples of this kind of view expressed in the literature on Kant's conception of spontaneity:

> 'I think' expresses my endorsement of the … proposition, backed by my reviewing the reasons for asserting it in the first place: combining the pieces of sensory information available, comparing them to others, forming recognitional concepts, and coming up with a proposition 'this is a tree,' which is now available as a premise for use in further reasoning. Note also that 'I think' may have different degrees of epistemic force. I may mean, cautiously, 'yes, it seems to me this is a tree' or, more strongly, 'yes, I'm sure it is a tree!' or some intermediate statement such as 'I have no reason to think it is anything other than a tree.' These different degrees of epistemic force depend, according to Kant, on the relation between the

[51] For *normativist* readings of Kant's conception of logical laws, which I will discuss in detail in Section 4, see Leech (2017), Longuenesse (2005), Lu-Adler (2017), and MacFarlane (2002).

> particular proposition 'this is a tree' and other propositions that belong to the system of knowledge available to me to back up my statement, whether or not I am, currently, consciously assessing the truth of those propositions. (Longuenesse 2017, 27)
>
> Unless the mind could think, that is, represent to itself or recognize the rule-governedness of the unification of its representations, it could not affirm the objective validity of this unification. It would remain a merely contingent, causally conditioned connection of mental states without any epistemic significance. (Allison 1996, 60)
>
> Now consider the character of the subject's awareness in making such a judgment. According to the account just given, her judging … expresses her taking this combination of representation to be required in any sound judgment on the matter, in virtue of the standards contained in the categories of substance and accident. And precisely in taking this to be how anyone ought to judge the matter, *she herself* thereby judges" (Boyle 2015, 24)

For all of their differences, these views share two important commitments. First, all of these authors agree that when Kant characterizes the understanding as spontaneous, he has in mind that the subject determines her judgment in accordance with certain rules that function to unify the representation in a way that is *correct*, or *appropriate*, or *required* for objective representation. The subject is therefore responsible for the judgment so formed, and the activity of judgment is an expression of her agency through her recognition of what is required of her in forming it. Second, they claim that when the subject forms judgments in accordance with these rules, the judgements acquire a certain epistemic significance for her; that is, her judgments express some form of commitment to the truth of the content of the proposition. Boyle and Allison therefore assimilate or closely associate judgment and belief, whereas Longuenesse wants to incorporate a broader spectrum of 'epistemic force' that could express attitudes weaker than belief.

For all of these commentators, these judgments are self-conscious in the sense that they are endorsed by the subject, and the subject has some kind of (implicit or explicit) grasp of the basis for the judgement or of what is required for forming the judgment. Now, there are different ways

of spelling out exactly what this consciousness consists in. McDowell (1994), for example, argues for a kind of explicit reflection on the basis for assenting to the truth of a judgment, which requires that the subject reflects on the reasons that they have for accepting it:

> Part of the point of the idea that the understanding is a faculty of spontaneity – that conceptual capacities are capacities whose exercise is in the domain of responsible freedom – is that the network [of interlinked capacities], as an individual thinker finds it governing her thinking, is not sacrosanct. Active empirical thinking takes place under a standing obligation to reflect about the credentials of the putatively rational linkages that govern it. (McDowell, 1994, 12)

Allison, too, sometimes expresses the idea in these terms. He says, for example that

> [o]n a Kantian account, the appropriate causal history of one's beliefs … is not sufficient for understanding. This is because understanding that *p* requires not merely arriving at one's belief that *p* by the correct causal route…; it also requires taking or recognizing their prior beliefs as warranting the belief that *p*. (Allison 1996, 63)

Allison nevertheless qualifies this view to suggest that there is no explicit representation of the basis for the judgment that occurs separately from the judgment itself. Instead, he claims that "apperception is not another act that the mind performs in addition to the representation to itself of objective connection. Consequently, it can consist only in the consciousness of the act of representation itself; its recognition of the nature of its own act" (1996, 61). On Allison's view, it is simply part of the nature of the act of judging that the judging subject is conscious of her activity of judging and of the norms or rules that govern that activity. Boyle (2015) takes up this suggestion of Allison's and develops it in more detail:

> In the normal, unreflective case, her focus is not on what she herself is doing, but how it is *correct* to combine a certain manifold of objectively-significant representations. In making a determination about this, she thereby judges, and may on reflection 'accompany her judgment with the *I think*', but her primary awareness is not a spectatorial awareness of some extant representational act, but an agential awareness of resolving a certain question. (Boyle 2015, 26)

Though we see disagreement between these authors about whether the self-consciousness in question is explicit or implicit, they nonetheless agree that it is because an agent is self-conscious that she is capable of regulating her activity in accordance with the rational norms governing judgment.

So far, we have been characterizing the cognitive agency reading without reference to the primary text, but it would be beneficial to see how one might work out the view as a reading of Kant's claims about spontaneity and self-consciousness in the Analytic. To this end, I want to consider a recent contribution to the discussion of cognitive agency in Kant by Thomas Land. In addition to offering a systematic argument for the position, Land's contribution is valuable because it highlights certain features of this kind of account, like the epistemic conception of judgment, that others often assume without argument, and that I will discuss in more detail in section 3.

Land (2021) finds in Kant's theory of judgment a view from the contemporary philosophical literature on the nature of belief that he terms *agentialism*. Land defines agentialism as the conjunction of two claims: according to agentialism, belief/judgment[52] is 1) spontaneous, and 2) constitutively self-conscious. Land uses agentialism as a device for interpreting Kant's conception of spontaneity: on this understanding, a cognitive subject acts spontaneously insofar as she is self-determining with respect to her beliefs. Thus, like many commentators, Land identifies the spontaneity that Kant attributes to the understanding with the kind of freedom required for agency:[53] on his view, this freedom or spontaneity is exercised when the agent grasps the normative constraints on forming her beliefs. The result of this freedom is that "when her doxastic

[52] Note that, for the purposes of the argument in the paper, Land does not distinguish between belief and judgment. These terms can be used interchangeably and are understood to involve the agent's assent to a proposition.

[53] Aside from Land, this conception of spontaneity is also present in Allison (1990), Boyle (2015), Engstrom (2006), Kohl (2015), McDowell (1994, 1998), McLear (2020a), Pippin (1987), Pollok (2017), and Valaris (2017).

capacities are functioning properly, a thinker believes that *p* if and only if she thinks she should believe that *p*" (3138). In such cases, belief is constitutively self-conscious because having beliefs requires "an ability to regard one's beliefs as subject to certain norms, or standards of correctness" (ibid).

At the basis of Land's account is a particular interpretation of the role of the unity of self-consciousness or apperception in our cognition. He takes the fact that Kant names the 'principle' of this unity as the 'highest principle of all use of the understanding' (B136) to imply that the understanding is *guided* by 'the rules of thinking'. These are the rules studied and presented in a science of logic, which for Kant encompasses both general logic and transcendental logic. One set of rules that guides our thinking about objects is therefore, on Land's view, the categories of the understanding themselves. Land then explains that "a thinker is guided by the rules of thinking [if she] … judges that *p* (at least in part) because she comprehends that judging that *p* is what conformity to the rules of thinking demands" (3144). Thus, the unity of self-consciousness itself functions as a requirement or a constraint that makes the rules of thinking binding. Land therefore claims that the thinker must exercise her faculty of understanding (that is, judge) in a certain way, namely, in conformity with 'the rules of thinking', so as to achieve the kind of unity that is required for judgment or cognition. She does so, moreover, insofar as she recognizes this unity to be required, that is, as a substantive normative constraint on her thinking and judging, and this is apparent to her *qua* self-conscious, rational being. Land thus claims that judgment must be self-conscious; that is, the thinker must think of the act of judging first-personally, and she must represent the activity of judging "under a description that gives rules for judging" (presumably, again, the categories). In that sense, she must also (at least implicitly) know what the rules are.

Land insists, however, that none of the rules are representations that form part of the

content of the judgment; they are, rather, determinants of its *form*.[54] He therefore thinks that self-consciousness is consciousness of the form of the understanding, which is just the consciousness of the rules of thinking and their bindingness on the thinker. On his view, then, the thinker never explicitly represents the rules as such in judging.[55] Rather, it is only in the subsequent assessment of the judgment that sensitivity to the rules could show up, as a disposition to change or withdraw assent on further reflection.[56]

This brief overview does not do full justice to the richness of Land's account, but it reflects, I think, many of the motivations for bringing the resources of the cognitive agency view to bear on Kant's project in the Analytic, and especially his account of the spontaneity of the understanding and the role of the unity of apperception. Land is making a case for reading Kant's project in the Analytic as an attempt to provide the normative framework that is supposed to be at the basis of the activity of judgment. It is Kant's conception of transcendental self-consciousness

---

[54] This is important, for he wants to avoid a problematic regress that arises if representation of norm-conformity is required for any act of thought. The worry is that this representation is itself a judgment that the subject needs to make – a second-order judgment that the first-order judgment in fact meets the relevant rules. We can avoid this, according to Land, if normativity does not require the representation of a distinctive content, but is instead expressed in the act of judgment itself. It is, as Land puts it "an aspect of taking the attitude of believing to a propositional content, as opposed to an aspect of a different content" (2021, 3147).

[55] While this is definitely Land's commitment, I am not entirely sure that it is consistent with his insistence that the activity of judging must be represented under a description that gives a rule for judging:

> For the rules of thinking to guide her activity, she must think of the activity as one she is engaged in. She must think of it first-personally. Second, it is already implicit in this that she must represent her activity. More precisely, she must represent her activity under a description that gives the rules of thinking application. (2021, 3145)

It is presumably only if I represent the activity *explicitly* as falling under one description rather than another that this requirement can be met, but in that case the rule must be part of the content of the representation in question.

[56] On Land's view, self-consciousness is simply a kind of mastery of the rules of thinking, and this mastery consists not in any theoretical knowledge, but rather in a skill or a kind of practical knowledge, which expresses itself in the exercise of the faculty of understanding itself:

> Possession of such knowledge is manifested in, for instance, a thinker's recognition of certain questions or demands that thematize her act's conformity to a relevant norm as appropriate as well as in her readiness, in non-defective cases, to withdraw her judgment consequent on her coming to think that it violates a norm. (2021, 3147)

that is meant to facilitate a subject's recognition of the normative constraints on judgment and her ability to regulate her activity in accordance with them.

It is important to note that Land's view, and all views like it, ascribes to Kant the *epistemic* conception of judgment introduced above, according to which judgments not only function to combine representational contents but also constitute acts of *assent* to those contents. It is because judgments constitutively involve acts of assent, in which the subject is responsive to the norms that govern judgment, that we can recognize judgments as expressions of cognitive agency. In the next section, as the first stage in our critical engagement with the cognitive agency reading, I will call this epistemic reading of Kant's account of judgment into question.

Section 3: Kant's Account of Judgment

The majority of commentators, regardless of whether they accept the cognitive agency interpretation of spontaneity, attribute to Kant the epistemic conception of judgment.[57] Indeed, this conception of judgment is so pervasive that most commentators assume that it is Kant's view without argument. According to this conception, as Land puts it, judgment is "not merely the entertaining of a proposition, but rather an act that involves the taking of an attitude regarding the truth of the proposition" (Land 2021, 1341). On this broadly Fregean conception, judgments constitutively involve an act of assent or dissent in addition to any other function that they may have.[58]

[57] Allison (1996, 2004), Engstrom (2006), Kohl (2015), Land (2021), McDowell (1994, 1998), Pollok (2017).

[58] As will become clear below, many scholars take the epistemic function of judgment – the act of assent or dissent – to be an essential feature of judgments but not the only function, for they recognize that on Kant's account judgments also perform the more elementary sematic function required for the unity of the proposition itself that is the target of the ultimate epistemic attitude.

Though a wide variety of readers assume the epistemic conception, however, we have seen that it plays a particularly important role for cognitive agency readings. It is because judging constitutively involves an act of assent that it is assessable in light of rational norms, and it is precisely because judgment is norm-governed in this way that it is supposed to constitute an exercise of cognitive agency. In this section, though, I will argue against the epistemic interpretation of Kant's account of judgment, and, by implication, against any view that depends on such an interpretation. In its place, I will characterize an alternative, non-epistemic account of judgment, which, I will claim, better fits with the passages in which Kant provides an explicit articulation of what judgment is. I call this the *semantic conception of judgment*; on this view, a judgment is simply an act of unifying representations to form propositional contents and does not additionally involve an act of assent to the content so formed. We will see that the semantic interpretation does a better job than the epistemic conception of making sense of the relationship between Kant's account of judgment and his broader epistemology, and offers a more natural reading of the various passages in which Kant explicitly characterizes the nature of judgment.

Reason to doubt the epistemic interpretation initially emerges when we consider the textual place of Kant's theory of judgment and take note of the fact that he develops his theory of epistemic attitudes such as 'knowing' or 'believing' in a different context altogether. Due primarily to the work of Watkins and Willascheck (2020), and also Tolley (2020), it is now widely accepted that Kant clearly distinguishes between cognition [*Erkenntniss*] and knowledge [*Wissen*] and discusses the latter, together with related epistemic attitudes of the subject, in the third section of the Canon of Pure Reason (A820/B848), and not in the part of the *Critique* in which he develops his theory of judgment and cognition, namely, the Analytic.[59] As these commentators have convincingly

[59] The following is a helpful summing up of the distinction between cognition and knowledge, as Watkins and Willaschek understand it: '[w]hile cognition is a conscious representation that is characterized by its

shown, it is important to keep these notions distinct, because cognition is 'upstream' from fully fledged epistemic states like knowledge. Consider, for example, Tolley's expression of these points:

> [C]ognition cannot be identified with what Kant means by '*knowledge* [Wissen]', i.e., the act of 'holding true [*Fürwahrhalten*]' a judgment or proposition which is true, where this holding-true is done with sufficient 'grounds [*Gründe*]' (roughly: justified true belief) … The progression to cognition does not include holding-true as one of its stages, nor does cognition itself include holding-true as one of its conditions. In this last respect, my approach will contrast with the more familiar *epistemological* way of explaining the significance of cognition for Kant, which takes cognition itself to be a mental state or act closely akin to knowledge, and takes Kant's analysis of cognition and its limits in the first *Critique* and elsewhere to be, first and foremost, discussions of the conditions under which we are justified in holding certain claims to be true. I will argue, instead, that Kant's analysis of cognition consistently places it at a much more psychologically elementary level in our mental lives, characterizing it as a distinctive form of consciousness of a real object by way of a specific kind of combination of representations. For cognition to obtain, only this specific form of consciousness of an object must obtain; nothing about the object (or anything else) needs to be claimed (held-true) at all—which also means questions of justification for such attitudes do not (yet) arise. (Tolley 2020, 3216)

If we conflate these notions of cognition and holding-true, then we cannot understand Kant's account of the more basic state of cognition on its own terms, for we are bound to then introduce notions of justification and related epistemic notions that are not obviously helpful for understanding Kant's concerns in the Analytic. Now, while the distinction between cognition and knowledge has been widely acknowledged in recent years, the implications of this distinction for Kant's conception of judgment have not been sufficiently appreciated. Because Kant explicitly

representational content (conceptual determination of an object), semantic features (successful reference to an existing object) and object-involving character (awareness of the existence and features of objects), knowledge is a propositional attitude that is defined in terms of a particular kind of epistemic justification' (Watkins & Willaschek 2020, 3209). For an earlier insistence on the distinction between knowledge and cognition, cf. also Smit (2000).

takes up the specifically epistemological states that we commonly associate with propositional attitudes independently of his theory of judgment as developed in the Analytic, we should be suspicious of any account that collapses the distinction between judgment and the forming of these attitudes. As I will show, Kant's explicit characterization of judgment is in terms of representing a certain kind of *unity* among our representations; the propositional attitudes that Kant describes as different ways of 'holding-for-true' (*Fürwahrhalten*) are, by contrast, not concerned with this more elemental unity of the representation. Instead, Kant's discussion of holding-for-true explicitly concerns the nature of the subject's assent to judgments that have different 'subjective' and 'objective' grounds of unity.[60] Thus, on my reading, any holding-for-true is an epistemic act that is downstream from and presupposes the more elementary semantic act of unifying contents in which judgment consists.

Now, to this general line, the cognitive agency reader might respond that the doctrine of holding-for-true, though only explicitly developed in the Canon, is really already implicit in Kant's discussion of the Analytic, and is simply an elaboration of an aspect of judgment that Kant does not develop fully in that context. As we have already seen, on the cognitive agency view, some kind of assent is built right into the very act of judgment itself. Since the act of judgment already involves the subject's (implicit) awareness of how she must or should unify the representational contents that feature in a particular judgment, the act of unifying the contents and the act of assenting to the proposition thereby established are one and the same act, and can be distinguished only logically or conceptually. In the Analytic, the proposal would then go, Kant's focus is on the content-generative feature of judgment, and in the Canon, his focus is on the epistemic feature of judgment. In what follows, I will show that there is no compelling textual basis for such a proposal,

[60] For discussion, see especially Chignell (2007), and also Watkins and Willaschek (2020).

and that, to the extent that proponents of the cognitive agency view rest their reading on this conception of judgment, they owe us a more direct argument for this view than they have provided hitherto.

The first point to make is simply that Kant's explicit characterizations of judgment in the Analytic never explicitly mention the epistemic dimension that is essential to the cognitive agency reading. In fact, it is difficult to find this epistemic dimension explicitly articulated in any of his published writings. Of course, this is not by itself a reason think that the epistemic interpretation is wrong, but the point I would like to make here is that if we do not approach the text antecedently convinced of the cognitive agency reading, we will not find any passages that speak directly in favor of the epistemic conception of judgment.

When we consider Kant's explicit pronouncements on the nature of judgment, we see no clear evidence that he takes it to be identical to any kind of commitment to the truth of a proposition. Instead, as we will see, Kant's characterizations suggest that a judgment is simply the *representation of a unity* of representations or cognitions. In the context of the Analytic, it is not clear that a judgment is anything over and above a determinate way of putting or holding different representational contents together so as to represent a necessary connection between those contents, such that they can be thought as combined in an object. Judgment, on this kind of view, is an operation that first forms truth-evaluable contents, but it does not additionally involve any endorsement of such contents. This is more appropriately seen as a semantic (rather than an epistemic) operation.

Support for the semantic reading comes from the *unifying* role that Kant consistently attributes to the activity of judgment. This role is particularly emphasized in various *Reflexionen*, as in the following:

> Judgment is a cognition of the unity of given concepts (*R*3042; 16:629)
>
> A judgment is the representation of the unity in the relation of cognitions. (*R*3044; 16:629)
>
> A judgment is the unity of a concept out of the relation (connection) of different concepts. (*R*3045; 16: 630)

In the *Prolegomena*, Kant says that "the unification of representations in a consciousness is judgment" (4:304). Though slightly less explicitly emphasized in the first *Critique*, the unifying role of judgment is also key to Kant's claim, in the B-Deduction, that the representations or concepts that make up a judgment must belong to one consciousness, and that judgment itself is "nothing other than the way to bring given cognitions to the objective *unity* of apperception" (B141, my emphasis). The semantic reading, which understands judgment simply as the operation that unifies representations into truth-evaluable contents, has a natural way of accommodating this unifying role. In contrast, the epistemic reader needs to find a way of casting this unifying operation in epistemic rather than semantic terms.

Now, the cognitive agency reader can tell a story that subsumes the unifying function of judgment under its epistemic function. According to the epistemic conception of judgment, the bringing of 'given cognitions to the unity of apperception' must be understood, ultimately, in terms of the subject's acceptance of certain propositions into her network of beliefs or cognition. Typically, this kind of interpretation will count the endorsement of this propositional content as the act of ascribing the propositional contents to the thinker, thereby bringing it to the unity of apperception. Of course, different accounts have different interpretations of how on Kant's account the endorsement is meant to occur, but the majority of commentators will want to say that the endorsement is immediately engendered by my consciousness of combing contents in the *correct* way, or in the way that is *required* by the rules of the understanding (the categories). Consider, for

example, Pollok's (2017) expression of the position:

> The 'I think' is not a second act, and not even a second-order act of our empirical mind. It rather stands for the imputability of this very same empirical consciousness of a [representation] tree. Without this possibility of an accompanying 'I think,' the empirical consciousness of a tree could not be identified as this representation. The transcendental apperception is the authority required for any empirical self to appropriate that representation, so to speak. (2017, 64)

Thus, we see that the cognitive agency reading does have the resources to give an account of judgment that combines its semantic function with an additional epistemic function. The intuitive thought is that, once one understands Kant's account of self-consciousness as ascribing contents to oneself as a thinker, then it is natural to think that the self-ascriptions cannot be neutral. Once I am in a position to "accompany all my representations" with the "I think" (B132), I also implicitly or explicitly assent to the propositions expressed by the combination of those representations in judgment. Every act of combining is at the same time an act of affirming the proposition that is the result of the combination. On this view, we need simply to understand the significance of Kant's account of spontaneity and self-consciousness to see that in judging we are not simply unifying contents is some kind of 'mechanical' way, but that the subject is really present and aware of her activity as responsive to certain norms. These are rules that the subject grasps a priori and that therefore guide her cognitive activity right from the outset. She can, therefore, appropriately be held accountable for the judgments. Once we appeal to this normative conception of the understanding, it is not difficult to see how we can provide a systematic account of judgment as already incorporating the epistemic dimension discussed more explicitly in the Canon.

Now, again, my aim here is not to question the philosophical coherence of this view. My point is rather that if we do not approach the text antecedently convinced of this sort of story, we will be hard-pressed to find anything in Kant's explicit discussion of the nature of judgment that would provide any obvious support to the epistemic conception of judgment on which the

cognitive agency reading depends. The question is not whether a tendentious reading of these passages is available; the question is whether there are any passages that speak directly in favor of the epistemic conception of judgment, and none of Kant's explicit characterizations of judgment appears to do so.

Nevertheless, we do find in Land's recent contribution one of the few attempts to spell out some textual basis for the epistemic conception of judgment, and it is worth considering the passage that Land adduces. Land's main argument for attributing the epistemic conception to Kant appeals to Kant's account of the modality of judgment, according to which all judgments are either problematic, assertoric, or apodictic. Here is Land:

> It might be objected that Kant does not apply the term 'judgment' to acts that carry assertoric force, since he uses a different term, 'holding true' (*Fürwahrhalten*), for these. Although I cannot make a full case for this here, I follow Michael Wolff in holding that, for Kant, a judgment is not merely the entertaining of a proposition, but rather an act that involves the taking of an attitude regarding the truth of the proposition. This is supported by the fact that Kant includes among the formal features of judgment what he calls its modality: every judgment is either problematic, assertoric, or apodictic. The modality "contributes nothing to the content of the judgment" (A74/B100), but instead concerns the thinker's attitude towards this content. (2021, 3141)

Land's argument proceeds by highlighting the fact that Kant thinks of the modality of judgments as a purely formal feature: modality is, as Kant puts it, "a quite special function of [judgments], which is distinctive in that it contributes nothing to the content of the judgment" (A74/B99). Given that modality does not shape the content of a judgment, Land concludes that the modality of a judgment concerns instead the thinker's *attitude* towards the relevant content. Kant continues the passage just cited by saying that the modality of judgment instead "concerns only the value of the copula in relation to our thinking…" (A74/B100). Citing this, Land claims that the 'relation to our thinking' that Kant is referring to here is a matter of a thinker's attitude to the proposition (the content). He then provides the rest of the passage in support. I quote it in full below since my

argument will depend on the sentence immediately before the selection that Land quotes (indicated by '*'):

> The modality of judgments is a quite special function of them, which is distinctive in that it contributes nothing to the content of the judgment …, but rather concerns only the value of the copula in relation to thinking in general. *Problematic judgments are those in which one regards the affirmation or denial as merely possible (arbitrary). Assertoric judgments are those in which it is considered actual (true). Apodictic judgments are those in which it is seen as necessary. (A74/B100)

Having quoted the passage, Land then claims that "[t]his suggests that what corresponds to the contemporary notions of belief and judgment in Kant's framework comprises, at least, assertoric and apodictic judgments" (2020, 3141). Unfortunately, Land does not discuss his analysis of this passage in detail, but I take it that he thinks that it is obvious that Kant's talk of 'regarding', 'considering', and 'seeing' the judgment in a certain way refers to an epistemic attitude the thinker might take towards the proposition expressed by the judgment.[61] If modality is to be explained in terms of the epistemic attitude of the subject, and on the plausible assumption that all judgments have some logical modality, it follows that all judgment involves the taking up of an epistemic attitude.

It is, however, difficult to understand exactly how the modality of a judgment — whether the judgment is problematic, assertoric, or apodictic — is supposed to be a function of the epistemic relation between the content or proposition expressed in a judgment and the subject. There are, I think, two possible ways that Land could be envisioning this to work. First, he could take the modality of the judgment to depend on the actual epistemic attitude that a particular subject takes to the proposition in question. The second possibility is that he takes the modality to depend on the epistemic attitude that the subject *ought* to take to the proposition. But, as we will now see,

[61] Land attributes this pattern of argument to Michael Wolff (1995).

both of these alternatives are problematic. The problem with the first line is philosophical: it makes the modality of judgments radically subjective. The problem with the second is exegetical: it both overreads and misreads the passage. First, if Kant had in mind the kind of normative relation that this reading requires, we would expect the passage to reflect this normative dimension more explicitly. Second, Kant locates the 'affirmation and denial' under discussion *within* the proposition, not in a thinker's relation to the proposition. Let me briefly expand upon the shortcomings of each of these two possible ways of reading the passage.

If we insist that the relation is one that obtains between the content of a judgment (a proposition) and some specific thinker, then we deny that the modality of a judgment is an objective feature of it, a consequence that I think it is clear we would want to deny. To illustrate, consider the proposition "snow is white". Suppose that some thinker, *X*, assents to this proposition. Then, on Land's view, the modal force of this affirmative categorical judgment is assertoric. Suppose now that thinker *Y*, who has never seen snow, has no opinion on the matter. On Land's account, this same affirmative categorical judgment now has a different modal force — it now seems that it is a problematic rather than an assertoric judgment. The modality of the judgment can therefore not be an objective feature of it since it is one that varies with given thinkers' attitudes. But if the modality of a proposition shifts with the thinker's attitude toward that proposition, it would be impossible to be mistaken about a proposition's modality. This implication, however, is inconsistent with a core tenet of Kant's criticisms of rationalist metaphysics in the Transcendental Dialectic. There, Kant argues that there is a set of metaphysical judgments generated by reason that are by their nature problematic, given how they are related to thinking. These are the synthetic a priori judgments that reason generates about the nature of the soul, the world-whole, and God. The specific epistemic attitudes of individual thinkers towards

these propositions do nothing to alter their character as problematic propositions, and this strongly suggests that for Kant the modal status of judgments is fixed independently of particular thinkers' attitudes toward their propositional contents. Thus, unless we want to ascribe to Kant a view of modality that makes it radically subjective, we must take seriously the fact that Kant does not make modality relative to any actual thinker, but to 'thinking in general'.

The second reading avoids this consequence but, as we will see, has problems of its own. On this view, the modality of a judgment consists in how a subject *ought* to regard the proposition expressed in it and hence is not liable to change with variations in how particular thinkers *in fact* regard it. Accordingly, the relation to 'thinking in general' is a normative relation to the rules that govern thinking as such. Yet a problem with this reading emerges when we look at how Kant sets up the selection of the passage that Land quotes, in which he distinguishes the three kinds of modality. It is not simply that there is no hint of any normative language in the way that Kant articulates the relations that constitute the different modalities of judgment; Kant also makes it clear that 'the assertion or denial' under discussion is not meant to express any attitude of the subject or thinker. Instead, a careful reading makes it clear that the 'assertion or denial' in question refers to the copula; that is, it is meant to be an assertion or denial of the predicate to the subject concept. This affirmative or negative *copula* is what Kant refers to as "the value of the copula" (A74/B100), and it is this value itself that is related to thinking. Land's reading requires us to read the 'assertion or denial' as an act of assent or dissent on the part of the subject to the proposition expressed, thus as a relation *between* the thinker and the proposition, when in fact Kant is referring to a logical relation *within* the proposition itself, between subject- and predicate-concepts. Kant claims that it is the value of the copula that is related to thinking, and not the proposition itself. The modality of the judgment therefore depends on whether this copula, which

determines a connection between two representations, expresses a connection that is merely possible, one that is actual, or one that is necessary.[62]

Building on this point, the semantic reading of judgment has a straightforward way of understanding Kant's theory of judgmental modalities without invoking any epistemic notions. It is undoubtedly important that a reading of Kant's theory of modality respects the claim that modality is not a feature of the content of judgment, and while Land's account does respect this desideratum, the semantic reading is also able to do so, and in a way that fits much more closely with the text. Rather than thinking of the modality of a judgment as being determined by a subject's propositional attitudes towards it, the semantic reading can think of the modal features of a judgment as reflecting facts about its *origin*. Consider the following footnote, which Kant adds to his discussion of the modalities of judgment:

> It is just as if in the first case thought were a function of the understanding, in a second of the power of judgment, and in the third of reason. (A74/B100n)

This suggests a reading on which the modal status of a judgment is determined by the cognitive faculty by which it was produced. If the concepts in a judgment are combined by the faculty of understanding, the resulting proposition is problematic; if they are connected by the power of judgment, the resulting proposition is assertoric; and if they are connected by reason, the resulting proposition is apodictic. I conclude, then, that there is a straightforward way of making sense of Kant's theory of modality without appeal to epistemic attitudes.

[62] Robert Hanna makes a related criticism of epistemic accounts of judgmental modality in his Stanford Encyclopedia of Philosophy entry on Kant's theory of judgment (2017).

Section 4: Normativity

In the previous section, I argued that the epistemic interpretation of judgment on which the cognitive agency reading depends is textually under-motivated and that an alternative, semantic, conception of judgment is much better placed to make sense of Kant's explicit claims about the nature of judgment as well as the relationship between his theory of judgment and his broader epistemology. In this section, I take a different approach by calling into question the way that the cognitive agency view conceives of the laws of the understanding. If we read Kant's doctrine of the spontaneity of the understanding as amounting to the claim that exercises of the understanding are manifestations of cognitive agency, then we also need to read him as holding that exercises of the understanding are governed by norms in light of which they admit of evaluation,[63] and these norms, rather than being externally imposed, are themselves given by the laws of the faculty of understanding. The cognitive agency reading thus presupposes a *normativist* account of the laws of understanding, on which such laws create the normative dimension along which exercises of the faculty can be assessed. But this normativist reading, I will argue, is deeply problematic as an interpretation of Kant because it is incompatible with core tenets of his faculty psychology. In making this argument, I provide support for an alternative, *constitutivist* reading of the laws of understanding, which pairs naturally with the semantic conception of judgment motivated in the previous section.

[63] Lu-Adler (2017) draws a distinction between *evaluative* and *imperatival* norms: whereas evaluative norms govern the assessment of particular thoughts, imperatival norms 'prescribe how an agent ought to conduct herself in her epistemic activities' (2017: 207). The primary notion I am interested in here is that of evaluative norms: normativism, as I understand it, holds that judgments, as products of the exercise of the understanding, are subject to assessment according to the norms given by the laws of the understanding. Presumably, though, the normativist will also appeal to the second notion: when a judgment is deemed to be defective in light of an evaluative norm, it will presumably have been generated through the agent's failure to conform to an imperitival norm in her epistemic activities.

We can start by briefly considering how the issue of normativity arises in the context of Kant scholarship. Both in his general and transcendental logic, Kant expounds laws of the understanding; he presents these laws in two tables, the table of the logical functions of the understanding in judgments (A70/B95) and the table of categories (A80/B106), the latter of which is meant to be derived in some way from the former (A79/B105). The laws presented in these tables are meant to govern the activities of the understanding that pertain to judging or thinking, and to cognition, respectively. Now, the question is whether these laws are *constitutive* or *normative* with respect to the faculty of understanding. *Constitutivist* readers, following Tolley (2006), hold that the laws of the understanding are constitutive of the understanding, such that nothing could both violate these laws and count as an exercise of the faculty. On this picture, the laws of understanding relate to the faculty much as the laws of physics relate to bodies: just as nothing could be a body and not be subject to the laws of physics, so nothing could be an exercise of the understanding and not be subject to the laws of understanding. *Normativist* readers, in contrast, hold that the laws of the understanding function as norms that create a dimension along which exercises of the understanding can be evaluated. Key to the normativist reading is the idea that an exercise of the understanding *could* fail to conform to the laws of the understanding: such exercises are precisely the ones that will be evaluated as bad or defective in light of the normative standard that the law establishes. On this kind of reading, the relation between the laws of the understanding and the exercises of the faculty should be modeled, not on the relation between the laws of physics and bodies, but rather on the relation between the moral law and a human subject. A human subject is certainly capable of violating the moral law, and insofar as they do, their conduct is judged to be bad; in the same way, the human understanding is capable of violating the

laws of general and transcendental logic, and when it does so, the relevant exercise of the understanding is accordingly to be deemed defective.

Much of the basis for the normativist interpretation lies in Kant's remarks about the nature of the science of logic. Commentators who want to claim that, for Kant, the laws of the understanding are normative, tend to focus on Kant's claim that logic, which is the science of the laws of the understanding, does not offer a descriptive psychological account of how the understanding in fact functions. In the following passage from the Jäsche Logic, Kant explicitly rules out that the laws of the understanding can be studied by any observation of how the understanding in fact functions:

> [s]ome logicians presuppose *psychological* principles in logic. But to bring these sorts of principles into logic is just as confusing as drawing morality from life. If we were to draw our principle from psychology – that is, from the observations of our understanding – we would merely see *how* thinking occurs and how it *is* under various subjective hindrances and conditions. This would lead to the cognition of merely *contingent* laws. But in logic it is not a question of *contingent* but of *necessary* rules; not of how we do think, but *of how we should think*. (*JL* 9:14, final italics mine.)

Now, at face value, this is a pretty compelling statement in favor of the normativist reading: Kant appears to be saying that the laws discovered by logic pertain to how we ought to think, just as the normativist reading holds. The passage appears to rule out any account on which the laws of logic merely describe the functioning of the understanding.

Below I will argue, however, that all Kant wants to rule out here is any *empirical psychological* investigation of the understanding as a method for studying and discovering its laws. This does not rule out constitutivism, because the laws that flow from the essence or nature of the understanding as such are not discovered empirically, but rather, as Kant explains in the Analytic, through a transcendental and a priori investigation of the faculty itself. Once we recognize that the investigation of the essence or nature of the understanding is not an empirical one, we can maintain

that the anti-psychologist claims in the above passage do not threaten the constitutivist position. Let us consider, then, in what sense Kant's project in the Analytic may be conceived of as an investigation into the essence of the understanding and how this project is related to the broader tradition of faculty psychology within which Kant is operating.

To do so, it is useful to discuss the Leibniz-Wollfian conception of powers since Kant explicitly characterizes the understanding as a power that functions in accordance with certain laws that are essential to its nature. Consider, for example, the claims that open the Jäsche Logik:

> 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 that we follow, *unconscious* of them at first, … Like all our powers, *the understanding* in particular is bound in its actions to rules, which we can investigate. Indeed, the understanding is to be regarded as the source and the faculty for thinking rules in general. (JL 9:11)

This passage is notable for two reasons. First of all, Kant does not seem to be drawing a distinction in kind between the laws or rules of nature and the rules that govern the understanding.[64] Second of all, we also see Kant endorsing a broadly Leibnizian conception of faculties and powers. As we will see, this conception of powers invites a constitutivist conception of the laws and rules that govern exercises of the understanding, just as the comparison with the laws of nature in this passage would suggest.

The general conception of a power that Kant inherits from the Leibnizian tradition is characterized in Baumgarten's exposition of the metaphysics of this tradition in his *Metaphysica*,

[64] And note that in an important passage in the third *Critique*, Kant goes further than this and explicitly *models* the relationship between laws of understanding and the faculty of understanding on the relationship between laws of motion and matter: "The universal laws of the understanding, which are at the same time laws of nature, are equally necessary to it (though they have originated from spontaneity) as the laws of motion are to matter; and their generation presupposes no intention with regard to our faculty of cognition" (*CPJ* 5:187). Like the passage above, this one also invokes a key notion from the metaphyics of powers that underlies Kant's faculty psychology — in this case, the technical notion of 'nature', which is alluded to in Kant's description of the laws of understanding as *natural* laws.

the standard metaphysics textbook in late 18th century Germany, which Kant himself uses as the basis of his metaphysics lectures throughout his career.[65] According to the Leibniz-Wollfian conception that Baumgarten adopts, a power is that which grounds the inherence of an accident in a substance.[66] Kant accepts this conception, but whereas Baumgarten then identifies this ground or power with the substance itself, Kant takes the ground, or power, to be a *relation* between the substance and the accident, which, therefore, is not reducible to either: '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' (*MM* 29:771). Nevertheless, while Kant rejects Baumgarten's reduction of power to substance, he accepts the view of power as the basis or locus of causal activity.[67] I will discuss this view in much more detail in the next chapter, but for present purposes, note that on Kant's view a substance *acts* causally by grounding the inherence of an accident though an exercise of its powers. This grounding occurs either exclusively through a substance's own power or powers, or in combination with the power or powers of another distinct substance. The substance thus acts — it *actualizes* an accident (that is, literally, makes the accident actual) — through the exercise of its various powers.

Now, it is important that every substance has an essence or nature.[68] My focus in what follows will specifically be on what Kant calls *formal* nature. Kant distinguishes formal nature

[65] As we saw in the Introduction, Kant does refines some of the framework of *Metaphysica* to better fit his Critical project, but the revisions he incorporates do not suggest a wholesale rejection of the basic Leibnizian metaphysical framework, or of the key metaphysical concepts — substance, accident, power, faculty, activity, spontaneity, passivity, to name a few — within the Critical system. See Tolley (2021) for a recent contribution that also makes the case for thinking that Kant's philosophy of mind is continuous with the powers-based metaphysics of the rationalist tradition.

[66] Cf. *Metaphysica* section 191; 197.

[67] See, for example, Watkins (2005), Oberst (2019).

[68] For my purposes here, the distinction between essence and nature is not particularly important. Michael Oberst (unpublished ms) has argued that Kant's conception of real as opposed to logical essence is equivalent to nature, whereas Stang (2016) argues that the concept of real essence is broader than that of nature. For our purposes in what follows, I will assume here that Oberst is right, and when I say 'essence' I will mean real essence.

from material nature, which, unlike formal nature, is common to all things: "it is the whole of all natures, the unity of all connected nature, in short, the general nature of all things, or that which all things have in common with one another" (*MM* 29:934-5). Whereas material nature is common to all things, formal nature, by contrast, is that which distinguishes one thing from others and makes it the thing that it is, for which reason Kant compares his conception of formal nature to Aristotle's conception of substantial form (*MM* 29:935). Kant characterizes formal nature as follows:

> [Formal] nature is the first general inner objective principle of all that which belongs to the existence of the causality of a thing... Nature is the first principle, and thus subalternate to no other… it is also an inner principle which is met in things themselves. (*MM* 29:935)

Kant also says that nature is the "internal first ground of that which belongs to the actuality of a thing" (*ML1* 28:215). Kant takes the nature of a substance to be composed of powers, and he takes these powers to be interconnected: "to nature belong not merely powers, but also the connection of these with one another" (*MM* 29:934). The formal nature of a substance is thus an integrated system of powers, and these powers constitute the inner principles of the substance's (causal) activity.

The inner principle of a substance determines which accidents could inhere in a substance that acts in accordance with it. It is therefore a *law* that governs the activity of some power of a substance and is expressed directly in any exercise of the power. Since the inner principle of a substance is given by its nature, Kant accordingly says that "[n]ature is the existence of a thing insofar as it is internally determined according to general laws" (*MM* 29:934), and, again, '*because* all things have a nature, nothing is by accident. Everything is according to laws" (*MM* 29:935, my emphasis). It follows, then, that every substance acts in accordance with its own nature, and for a substance to act in accordance with its own nature is for its acts, as expressions of its powers, to

be governed by a law that is given by its nature. Every power of a substance, insofar as its exercise is governed by an inner principle or law, is therefore expressive of that substance's nature or essence.

The laws that stem from the nature of a substance, and that are expressed in the exercise of its powers, have a special constitutive status. Since the formal nature of a substance, like an Aristotelian substantial form, is what gives that substance its specific identity, a substance could not fail to act in accordance with these laws without losing its identity as the very substance that it is. Similarly, on the view developed here, the exercise of a particular power expresses the law or inner principle that constitutes its essence and makes it the specific power that it is, such that a power could not fail to conform to the laws that derive from its essence.

With this background in place, we can ask whether the laws that govern the understanding are likewise determined by its essence or nature. If those laws *are* determined by the nature or essence of the understanding, then they must be constitutive of the understanding, such that for anything to count as an exercise of the understanding it must express those very laws.[69] Now, Kant claims not only that "*the understanding* in particular is bound in its actions to rules, which we can investigate"; he also claims that "the understanding is to be regarded in general as the source and the faculty for thinking rules in general" (*JL* 9:11 – 12). This conception of the understanding as the *source* of rules suggests that the rules of the understanding derive from its nature, and as we

[69] Tolley (2006) recognizes this connection between constitutive and essence-expressing laws in the following passage:

> [n]either a Leibnizian nor a Kantian 'understanding' is 'free' to adhere to any other law (not to be 'lawless'). It is simply that which accords with the logical law; *the 'essence' of understanding is wholly expressed by logical laws*. These laws are not things that either Leibniz or Kant suggest that the understanding 'ought' to live up to, or 'ought' to act in accordance with, but are rather laws which articulate the very form of 'being' of the understanding itself. (2006, 388, my emphasis)

have just seen, no exercise of a power can diverge from the laws that express the nature of that power. And this account of the origin of the rules of the understanding reappears in the context of the Analytic, for there Kant characterizes the project of Transcendental Philosophy as one of analyzing not the concept, but "the faculty of the understanding itself, in order to research the possibility of *a priori* concepts by seeking them in the understanding as their birthplace" (A65/B90 – A66//B91). Once again, this passage suggests a status for the laws of the understanding that, given Kant's faculty psychology, would strongly recommend a constitutivist reading.

Now, one strategy the normativist might want to take at this point is just to deny that the laws of the understanding derive from its essence. I do not think that this is a particularly promising strategy; if the laws of the understanding do not express the essence of the faculty, then the normativist reader owes us an alternative account of their source as well as Kant's method of discovering them, and giving such an account would involve the normativist reader in some formidable textual challenges. If the laws of the understanding do not have their origin within the understanding itself, then they are presumably legislated to the understanding by some external source. But it is hard to square this conception of the origin of the laws with Kant's chosen method for discovering them. For Kant claims to discover the laws of the understanding in the Transcendental Analytic, in which, he says, 'we isolate the understanding… and elevate from our cognition merely the part of our thought that has its origin solely in the understanding' (A62/B87). And recall that he says that he will "research the possibility of *a priori* concepts by seeking them *only* in the understanding as their birthplace" (A66/B90, my emphasis). It is very hard to see how Kant could hope to discover the laws of understanding through this method of isolation and analysis if he thought that those laws had a source outside the understanding.

An alternative strategy for the normativist reading would be to concede these points, and

accept that the laws of the understanding are expressive of its essence, but nevertheless insist that we can make room for a normative dimension to the rules of the understanding that does conflict with their constitutive status. On this view, we can acknowledge the constitutive status of the laws of understanding while nevertheless making room for the normative assessment of the acts of the understanding by thinking carefully about the different contexts in which the understanding functions. Boyle (2020) has recently made an argument to this effect, and it is worth considering whether or not this attempt to reconcile the two conceptions might be sufficient to preserve the reading of the spontaneity of the understanding in terms of cognitive agency.

Boyle (2020) emphasizes the fact that we can consider how the understanding functions in two different contexts. We can consider how it functions in isolation from any other cognitive faculties, or we can consider how it functions in relation to these other faculties. Once we pull these two contexts apart, Boyle argues, we will see that conceiving of the rules of the understanding as constitutive does not conflict with the claim that these very rules are also normative:

> The point I want to emphasize about Kant's view of these matters is that it does not oppose prescriptive laws to laws that govern the actual functioning of a power. On the contrary, the very laws that describe the *proper* functioning of a power are the laws that determine its *actual* functioning *if nothing interferes*: they are laws that describe how, as we might put it, that power is, in itself, disposed to act. Thus if logic is the science concerned with the laws of the understanding, it is concerned with laws that do not merely prescribe *to* the understanding but flow *from* the nature of this faculty. (2020, 129; my emphasis)

While I agree with Boyle's claim that there is a way to make room for a normative dimension to these rules, the problem is that Boyle is not being specific enough about the different contexts in which these rules are studied. If it turns out that the context in which these rules show up as norms is not relevant in the Analytic, it would be hard to see how this normative dimension could help us to make sense of the claims that Kant makes there about the nature of the understanding.

To see this more clearly, it worth considering the conditions under which the understanding

can violate its own laws, for it is these conditions that Boyle appeals to in order to open up the possibility of a normative role for them. His starting point is Kant's claim that

> [n]o natural power [*Kraft der Natur*] can of itself depart from its own laws. Thus neither the understanding by itself (without the influence of another cause), nor the senses by themselves, can err. (A294 / B350, cf. *JL* 9:53–54)

What is suggested by this passage is, of course, that the understanding *can* err if it is influenced by some external power. Thus, on Boyle's view, the rules of the understanding are constitutive with respect to the understanding considered in isolation but normative for a subject whose understanding functions in combination with other powers that might interfere in various ways with its proper operation.

We should note, however, that Tolley (2006) already recognizes scope for this kind of normative dimension but insists that if we look carefully at Kant's distinction between pure and applied general logic, then we will see that normativity is altogether absent from the former. Kant explains the contrast between these disciplines as follows:

> In [pure general logic] we abstract from the empirical conditions under which our understanding is exercised, e.g., from the influence of the senses, from the play of imagination, the laws of memory, the power of habit, inclination, etc., hence also from the sources of prejudice… A general logic, however, is then called applied if it is directed to the rules of the use of the understanding under the subjective empirical conditions that psychology teaches us. (A53/B77)

It is only in applied general logic that Kant takes into account the empirical conditions under which the understanding can err at all, which means that it is only from the perspective taken in applied logic that the laws of general logic could be viewed as norms. A pure general logic, in contrast, abstracts entirely from this set of empirical conditions, which means that from the perspective taken in pure general logic, the laws of the understanding could not take on any normative

significance for exercises of the faculty.[70]

Now, Kant distinguishes applied logic not only from pure general logic, but also from transcendental logic. While there is debate about whether transcendental logic should be considered a general or a particular logic (for discussion, see Tolley 2012), what is not in question is that transcendental logic is no less pure than pure general logic. Thus, transcendental logic, no less than pure general logic, abstracts from the subjective empirical conditions under which the understanding can err. From the perspective taken in transcendental logic, then, just as from that taken in pure general logic, the rules of the understanding could not function as norms. Now, Kant's primary interest in the Transcendental Analytic is in the laws of transcendental logic, which means that the transcendental subject discussed in the context of the Analytic is not the same as the empirical subject studied in applied general logic. It is only this empirical subject for which the laws of the understanding function as norms, and hence only the empirical subject whose conformity to those laws could amount to an exercise of cognitive agency. Therefore, although Boyle is correct to claim that we can preserve some normative function for the rules of the understanding, he has only shown, as Tolley claims, that "for Kant, normativity is at best an externally conferred, rather than an essentially inherent property of logical law" (Tolley 2006, 375). Once we concede this point, it is unclear that Boyle's proposal allows us to understand Kant's account of the understanding in the Transcendental Analytic in terms of the epistemic normative framework that the cognitive agency view requires.

[70]In aligning the distinction between pure and applied logic with the distinction between a theoretical perspective from which the laws of understanding have constitutive status and one from which they have normative status, I am in agreement with Merritt (2015; 2018).

Conclusion

In this chapter, I have argued that one prominent model for interpreting Kant's doctrine of the spontaneity of the understanding, which appeals to the notion of cognitive agency, cannot succeed. Having outlined the motivations and presuppositions of this cognitive agency reading in Sections 1 and 2, in Section 3 I raised the first objection to the reading, arguing that it attributes to Kant a conception of judgment that is foreign to the project of the Analytic. This *epistemic* conception of judgment cannot be Kant's, I argued, because it would appear that Kant sets aside a more elementary content-producing role for judgment, which is independent of the epistemic assessment of those contents. I therefore briefly sketched an alternative *semantic* conception of judgment better suited to capture its unificatory function,[71] and I showed how this conception does not obviously require that we import any normative notions, such as agency, accountability or responsibility into Kant's theory of understanding. Next, in Section 4, I suggested that the normative epistemic framework that the cognitive agency view relies on does not adequately take into account the historical context in which Kant develops his view of the understanding as a faculty or power of the soul. Once we take the basic framework of the Leibniz-Wolffian conception of powers into account, I argued, it would appear that Kant thinks of the laws of the understanding in straightforwardly constitutivist terms.

Nevertheless, I do not take my argument against the cognitive agency reading at this point to be conclusive, partly because we do not yet have any real grasp of what an alternative to such a reading would look like. My aim in the rest of this dissertation is to articulate and motivate just such an alternative, and to that end, I turn in the next chapter to developing an alternative conception of the spontaneity of the understanding. The conception of spontaneity that I attribute

[71] I say much more about how judgment achieves this crucial unificatory function in Chapter 4.

to Kant is continuous with the Leibnizian tradition and it further supports the claim that Kant does not conceive of spontaneity as any kind of *freedom* about what to believe or how to judge.

CHAPTER 3

Spontaneity and Teleology

As we saw in the previous chapter, by far the most common strategy in the literature for making sense of the Kantian conception of spontaneity in the theoretical domain is to appeal to the notion of *cognitive agency*. According to advocates of this strategy, the spontaneity of the understanding is best understood as a kind of freedom or autonomy that the agent has with respect to what she holds true or what she believes, or more broadly in judgment. In the present chapter, in the course of introducing my own positive proposal, I engage with an alternative view, which is meant to constitute a corrective for the more metaphysically committal or 'loaded' cognitive agency views. This kind of view, which I call *deflationist*, has wanted to resist making any substantive metaphysical claims about the cognitive subject and her freedom by suggesting that *all* that is meant by spontaneity in the context of Kant's theoretical philosophy is a kind of 'structural creativity' in the mental operations of the understanding.[72] On this kind of view,

[72] See, for example, Hanna: "A cognitive faculty is spontaneous in that whenever it is externally stimulated by raw unstructured sensory data as inputs, it then automatically organizes or "synthesizes" those data in an unprecedented way relative to those inputs, thereby yielding novel structured cognitions as outputs (B1–2, A50/B74, B132, B152). So cognitive spontaneity is a *structural creativity* of the mind with respect to its representations" (Hanna, *Kant's Theory of Judgement*). Now, in presenting the deflationist view as an *alternative* to the cognitive agency view, my claim is not that the cognitive agency view cannot also account for the structural creativity. One could (and many do) endorse a view according to which it is the activity

spontaneity can be understood simply in terms of the fact that the understanding can 'hold together' or 'structure' representations in a way that is not *given* in sensibility, which is therefore novel or creative relative to what is already cognitively available.[73] The deflationist reading has been especially appealing to psychological readers such as Kitcher, since this conception does not interfere with the project of explaining the operations of the understanding in the same phenomenal causal terms that we use in the context of natural science.

In resisting what they take to be the problematic metaphysical commitments of cognitive agency views, deflationist readers treat the understanding's activity as *automatic*, *involuntary*, and *mechanical*.[74] Moreover, most commentators seem to take these characterizations to be at least extensionally equivalent if not synonymous. To the extent that they do so, they will, therefore, want to insist that the understanding is, in some or other sense, mechanical, for any rejection of this commitment would seem to embrace exactly the kind of freedom that is essential to the cognitive agency view. Consider, for example, Susan Neiman's (1994) forceful expression of this mechanistic conception of the understanding:

> The categories of the understanding are spontaneous in that they are applied to, rather than derived from, experience […] The understanding provides the conditions under which anything can be perceived as an object, yet it does so in a

---

of the agent that is responsible for structuring the representational contents in ways that are not given in sensibility, but also make the additional claim that in so structuring the content, the agent is *also* making up her mind about what to think (that is, whether to accept the propositional content generated). Thus, my point in setting up the two views as alternatives to one another is not to claim that the cognitive agency view cannot acknowledge that structural creativity plays any important role in Kant's account of spontaneity. The contrast is rather that for deflationist readers, unlike for cognitive agency readers, the spontaneity of the understanding is entirely accounted for by the novel structural features that the understanding brings about and not by any special metaphysical fact about the agent or her freedom.

[73] I take it that Hanna (2011), Kitcher (1990, 2011, 2017), and, as we will see, Neiman (1994), as well as Allais (2015), Bell (1987), Brook (1994), and Waxman (2014) all sign up for some version of this view.

[74] See, for example, Neiman (1994): "In later works, the mechanical nature of this process [of the understanding] becomes increasingly explicit. ... The fact that the workings of the understanding are automatic is the ground of the fact that its concepts are fully determinate: there is no need to invoke an external decision procedure to determine whether, and how, a category of the understanding is to be applied" (1994, 49).

> way Kant describes as reactive, dependent on the sensible world as the occasioning cause of its activity (VII; 71). Its operations are *routine, automatic, and mechanical*: understanding 'merely spells out appearances according to a synthetic unity, in order to read them as experience' (A314/B371). (1994, 49; my emphasis)

Now, admittedly, not all commentators express this commitment quite as explicitly as Neiman does here, but I take it to be a standard commitment of deflationist views such as Neiman's that there is some sense in which the operations of the understanding are simply part of the causal order of nature.[75] To the extent that this is so, one would expect that the functioning of this faculty is explicable in the same causal mechanistic terms by which we explain natural phenomena in general.

While I agree with the psychological reader that the operations of the understanding should not be understood in normative epistemic terms, in this chapter I will argue against the deflationist commitment that the understanding functions mechanically, by showing how the notion of spontaneity fits into a set of closely related notions, including *activity*, *substance*, and *power*, all of which are, as we have seen, central to Kant's broader metaphysical conception of causality and his faculty psychology. I will show that the conception of spontaneity that emerges in this context is in fact incompatible with mechanism, and therefore, since Kant characterizes the understanding

---

[75] It is worth noting that in characterizing relative spontaneity, Sellars (1970) suggests an *analogy* rather than *identity* between the mechanism of nature and the operations of the understanding, and he therefore, presumably, does not mean that the understanding is mechanical in a strict sense.

> the spontaneity of which we are conscious is, though not a *sheer* passivity, nevertheless a passivity in that the inner development is set in motion by a foreign cause and follows a routine. In the awareness of noumenal activities of synthesis we would encounter simply another example of a cause the causality of which is caused. If this were all that the spontaneity of the noumenal self amounted to, then although it would not be part of the phenomenal nature of outer and inner sense, it would be *like* an object in nature and might be called a noumenal mechanism. (1970, 24)

It is less clear whether Neiman's conception is metaphorical because she seems to explicitly to want to rule out that the understanding can act for ends (see, for example, Neiman 1996, 62). Regardless of whether or not the claim is ultimately metaphorical, I will argue that it is misleading to conceive of the function of the understanding in these terms because it in fact functions teleologically.

as spontaneous, it follows that the understanding cannot function mechanically. I thus argue that we can avoid *both* this mechanical view of the understanding *and* the cognitive agency view, and my ultimate aim in this chapter is to develop a third alternative by appealing to the traditional Leibniz-Wollfian conceptual framework that Kant adopts and extends in articulating his conception of spontaneity. It is to this conception of spontaneity, I show, that we must appeal in order to make sense of Kant's claim that the understanding is a spontaneous faculty.

My treatment of spontaneity as a central concept in Kant's broader metaphysical conception of causality is in stark contrast to the general trend in the literature to take Kant's theory of cognitive spontaneity to be explicable purely in terms of his theory of judgment.[76] I will argue that by setting Kant's account of spontaneity against the theoretical backdrop of his metaphysics of causation, we are in a position to appreciate more fully a distinctive and often overlooked feature of his theory of the understanding — namely, that Kant conceives of the understanding in teleological terms.[77]

The chapter is divided into two main sections. In Section 1, I will draw on Kant's metaphysics of causality in order to distinguish two kinds of activity — *passivity*, on the one hand,

[76] There are some exceptions here — most prominently, Sgarbi (2012) and Smit (2009). Dyck (2016) has also made an important contribution by emphasizing the metaphysical framework in which Kant develops his account of spontaneity, though he focuses primarily on Kant's pre-critical philosophy. See also Ameriks (2000) and Dyck (2014) for accounts of how Kant's metaphysical framework of causal powers informs his rational psychology. Most recently, Tolley (2021) emphasizes the metaphysical context of Kant's account of mental powers in general.

[77] The claim that the understanding functions teleologically is not unprecedented. Béatrice Longuenesse (1998) develops an account on which we cannot make sense of the understanding's activities of synthesis without positing judgment as its goal. Karl Schafer (2020) recently argued that generalizing Longuenesse's account of the teleological unity of the understanding enables us to see the faculties of the mind as forming a teleological system, but Schafer departs from Longuenesse in drawing the unifying principle of this system from the aims of highest cognitive faculty, namely, reason, rather than the lower intellectual faculties. Both Bernt Dörflinger (1995, 2000) and Courtney Fugate (2014) also develop accounts of the teleology of the intellect as a whole and of understanding as a component thereof; but no one has attempted to argue that the understanding is teleological by focusing, as I do in this chapter, exclusively on Kant's conception of spontaneity.

and *self-activity*, or *spontaneity*, on the other — and show how each of them is a specification of a common genus, *activity*. I will then utilize this distinction to draw an important further conclusion, namely, that each of these kinds of activity is subject to a distinctive form of explanation. Whereas beings that are wholly and exclusively passive are intelligible through mechanistic explanation, we will see that beings that also have spontaneous powers cannot be explained except in *teleological* terms. With these results in place, we will turn, in Section 2, to a more detailed investigation of the notion of spontaneity as it characterizes the human cognitive faculty. I conclude by suggesting that my reading can offer a genuine third way between the existing interpretative options: the position has the resources to resist the mechanistic conception of understanding implicit in the deflationist view without collapsing into a variant of the cognitive agency view.

Section 1: Kant's Conception of Activity

We can start by rehearsing some of the distinctive features of Kant's metaphysical theory of causation. As has been appreciated in recent literature on the subject, the basic framework of this theory is developed in terms of an account of causal powers.[78] For orientation, it is useful to provide a very brief overview of the main components of Kant's account as it is developed by Eric Watkins (2005). Watkins argues that Kant employs a broadly Aristotelian framework that is widely accepted in the metaphysical tradition of his rationalist predecessors. Kant develops this framework, however, by articulating a novel conception of causal powers, according to which they are irreducible to either substances or their determinations; instead, they constitute "primitive

[78] See Ameriks (2000), Dyck (2014), Indregard (2014, 2017), Oberst (2019), Smit (2009), Tolley (2021), among others. For some concerns about this trend in the literature, see Kitcher (2016) and Kreines (unpublished).

relation[s] 'in between' substances and their determinations" (Watkins 2005, 13). The exercise of a causal power, which establishes the relevant relation between a substance and its determination, depends on what Kant calls a *ground*. As Watkins puts it, "[t]he basic idea of a model of causality for which grounds are central is that one substance determines the successive state of another by means of an unchanging ground that is part of its essential nature" (244). What is particularly important for us is that on Kant's view these grounds are *activities* (cf. Watkins 2005, 231).

1.1 Activity in general

We can get a better grasp of Kant's notion of an activity by considering the basic distinction between a substance and an accident. Kant draws the distinction in the *Methaphysik Mrongovius* as follows: "That which exists without being the determination of another is substance; that which exists only as determination is accident" (*MM* 29:770).[79] On this basis, we can identify substance as an existing thing. Kant says that it is what remains even when we "leave aside all accidents" (*MM* 29:771). We cannot form any positive concept of such a being, as we have 'no acquaintance with the substantial, i.e., the subject, in which no accidents inhere…' (*MM* 29:771); instead, we cognize a substance only through its determinations or accidents. Since "accidents cannot exist other than in the substance" (*MM* 29:796), the substance contains a ground for the existence of an accident. In order for a substance to exist in one rather than another way, that is, for it to have a particular determination, it has to 'actualize an accident', and this actualization is, or expresses, the causal power of the substance. Thus, the relation between a substance and an accident, such

[79] A note on methodology: in this section and throughout, I will be relying quite heavily on the transcripts of Kant's metaphysics lectures. Where possible, I build my case exclusively on transcripts from the Critical period and published works; where I have appealed to earlier transcripts from the pre-Critical period (such as the *Metaphysik L*$_1$), I have supplemented these textual materials with published material or transcript material from the Critical period.

that the accident exists 'in' the substance, or 'inheres' in it, is the result of the exercise of the causal power of the substance in which the accident inheres.

With this distinction between substance and accident in place, we are in a position to understand Kant's most basic conception of activity. Consider the following passage, again from the *Metaphysik Mrongovius*:

> A substance, insofar is it contains the ground of that which belongs to the being of one thing, acts <*agirt*; G: *handelt*>; ... substance, insofar as its accidents inhere, is in action, and it acts insofar as it is the ground of the actuality of the accidents… (*MM* 29:773)

Here, Kant defines activity in general (the genus) as that in a substance which grounds the inherence of an accident.[80] Because accidents cannot exist otherwise than *in* a substance, the substance must contain the ground of the accident that belongs to it. As I've noted already, such a ground is an activity. Activity, in general, is therefore the explanation for why a substance exists in a particular way, that is, why it has a particular accident or is in a particular state.

It is important to note that Kant does *not* take activity and passivity to be mutually exclusive concepts.[81] Instead, as I will show, Kant treats activity as a genus that admits of division into two species: the first is *passivity* or *suffering,* and the second is *self-activity* or s*pontaneity*. Both passivity and self-activity, as species of the genus, activity, can be understood in terms of the

[80] At this point, one might worry that defining activity *in general* in terms of a ground of the inherence of an accident in a substance is too restrictive, for inherence is not the sole relation for which activities are grounds. The most obvious candidate to be included is the causal relation. Thanks to Brian Tracz for raising this concern. While I agree that causal relations are also grounded by the activities of substances, I nevertheless think that such a relation obtains only when a substance grounds the inherence of an accident *through an external power*. The accident can then be considered as an effect of the external power on the substance in which the accident comes to inhere. If this line of thinking is on the right track, then it is correct to say that activity in general is what grounds the inherence of an accident in a substance; the question of whether the activity is causal concerns *how* the accident comes to inhere in the substance. The causal relation specifically relies on the exercise of an external power, without which the actualization of the accident is not possible. This will become clearer in what follows.

[81] As we saw in the Introduction, Kant departs from Baumgarten on this point.

general conception of activity as that which grounds the inherence of an accident in a substance. What distinguishes the two kinds of activity is *how* the substance grounds the inherence: either a substance can ground the inherence of an accident through an *external* or *outer* power, in which case it is passive, or the substance can ground the inherence of an accident though its own *internal* or *inner* power, in which case it is spontaneous or self-active. I turn now to a more detailed discussion of each of these two activities.

1.2 Passivity

I start by considering the first kind of activity, which Kant also identifies, somewhat paradoxically, as passivity or suffering. Despite the fact that there seems to be a logical opposition between the characterization of something as active and the characterization of something as passive, Kant in fact takes passivity to constitute the first of two kinds of activity: "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" (*MM* 29:822). In the case where a substance suffers, it has a power that Kant calls *receptivity*. It is crucial that such a power cannot actualize the accident independently of an external power; that is to say, it is not 'externally sufficient' (*MM* 29:824) for the change to take place: it needs something from outside itself, another power, to actualize the accident:

> What then is genuine passivity <*passio*>? 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)

The substance acted upon has a power that is 'determined' by the power of the substance that acts upon it. The power of the substance acted upon is thus "externally insufficient" (*MM* 29:824): it

needs to be determined by an outer power because it is not sufficient by itself to actualize the accident; it is therefore a receptive power, which must 'receive' a determination from outside in order to actualize the accident.

It is, nevertheless, important to note that the suffering itself is an *activity* of the substance that is being acted upon, for it is only in virtue of the suffering that the accident inheres in the substance, and, recall, activity in general is that which explains the inherence of an accident in a substance.[82] Kant illustrates how exactly this is meant to work with the following example:

> 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. From the union of one substance with another an effect comes about, namely, the representation of the trumpet sound. (*MM* 29:822)

First, we should note that *hearing* the trumpet sound is meant to be an example of suffering or passivity. The example illustrates that the *receptive* power of representation (in this case of a sound, so a sensible power) must nevertheless be active, for if it was not, I could not represent (or hear) the sound of the trumpet. This is still an example of *suffering*, for, as Kant says, the "trumpet sound inheres in me through an external power". The power of representation is the ground of the *inherence* of the accident but it is not by itself sufficient to represent the sound of the trumpet (the accident): something *outside* of the power of representation is required for me to actualize this accident in myself, namely, the sound itself that is generated by the trumpet. I couldn't *hear* the sound of the trumpet without an external power, the sound itself, affecting my sensible (receptive) powers.

To recapitulate briefly: Despite the appearance of a contradiction, passivity is really a kind of activity insofar as it is the sufficient internal ground of the inherence of an accident in the

[82] See *MM* 29:823

suffering substance. As Kant's example illustrates, if a substance does not exercise a receptive capacity, then no external power can act on it such that an accident comes to inhere in it. It is only in the conjunction of the external and internal powers that 'an effect', the actualization of an accident in the substance, comes about. Since the inherence of the accident in the substance depends essentially on this receptive power, the exercise of such a power is an activity, albeit an activity that constitutes suffering. The activity consists in nothing more than receiving a determination, but it is not for that reason not active, for it is still the internal ground of the relation between the substance and the accident, that is, it is still the ground of the fact that the accident inheres in the substance. It is nevertheless also the case that the internal ground is determined by an external power, which thereby constitutes the cause of the state of the affected substance.[83]

1.3 Spontaneity

We have seen that the first kind of activity, passivity or suffering, is externally *insufficient* for the inherence of an accident in the suffering substance because the substance still needs, beyond its internal power, an external power to actualize the accident. It is nevertheless internally sufficient since, given that the external power is exercised, the substance grounds the inherence of the accident through the exercise of its receptive capacity. With this account of passivity in hand, we

[83] All substances have internal grounds or principles for the inherence of accidents. These grounds are either externally *sufficient* to ground the inherence or they are not. If they are not, then they depend on an external principle; if they are, then they do not depend on an external principle but exclusively on an internal principle. Kant explains the notion that substances can have internal grounds that are not externally sufficient to realize an accident as follows:

> 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 a power. Power that suffices internally but not externally is dead power; that which suffices internally and externally is living power. (*MM* 29:824)

Accordingly, an internal ground that can bring about a change in a substance is a power. An internal ground that is not sufficient and 'to which something more must be added' for the accident to inhere is a receptivity.

are now in a position to see that the real contrary of passivity is not activity *per se*, but rather what Kant calls *self*-activity or spontaneity. Unlike passivity, this second kind of activity is determined, not through an outer or external power, but from an inner principle. As Kant says, "[T]he act <actus> of spontaneity cannot proceed from an outer principle" *(ML1* 28:285; cf. *MVM* 28: 448; *MFNS* 4:544). That simply means that the inherence of the accident in the substance does not depend on any *external* power, and since, as we have already seen, the inner principles of a substance constitute its nature, the change in the substance is determined by its nature.

We can therefore distinguish two ways in which a change in the state of a substance comes about. Either the change is caused by an external power or the change is grounded in the power itself, in its nature or essence. In the former case, Kant says that the activity proceeds from an *outer* or *external* principle that affects the substance, and we have seen that this constitutes the first kind of activity, suffering. In the latter, the activity proceeds from an *inner* or *internal* principle, and it constitutes the second kind of activity, spontaneity. In this context, then, a principle is what determines a change in the state of a substance. Whether a principle is external or internal depends on whether the determination is a cause outside of the power of the substance that is changed, in which case it is external, or whether the determination is grounded exclusively by the nature or essence of the power of the substance in question, in which case it is internal.

1.4 Matter vs. Living Beings

What I want to do now is to start motivating my claim that beings with spontaneous powers are subject to a different form of explanation than beings with merely passive powers. And in order to see this point it will help to start by considering more closely the kinds of beings that Kant takes each of the activities to be characteristic of. There is a specific kind of substance that Kant thinks

can be characterized *exclusively* in terms of passivity, namely, phenomenal substance <*substantia phaenomenon*> or matter. By contrast to matter, which is merely passive, and which Kant consequently characterizes as 'lifeless' (*MFNS* 4:544; *MM* 29:913), or sometimes also as 'dead' (*MM* 29:824), the kinds of being or substances that are spontaneous, or self-active, are characterized as alive or 'animate'. I discuss each of these kinds of beings in turn.

Take, first, the kind of substance that Kant thinks is exclusively and wholly passive. Kant says that

> [a]ll matter is lifeless, has no faculty for determining itself, and the principle of life is something other than matter. *For every matter remains in motion or at rest until it is altered by something else. Matter thus has mere receptivity or passivity*. (*MM* 29:913, my emphasis)

Matter is passive because all of its determinations, motion and rest, must be explained through external powers.[84] What the passage we just read makes clear is that matter is merely passive or receptive insofar as it lacks a faculty or a power for self-determination. Kant identifies two primitive powers of matter, attracting and repelling powers (*MM* 29:822; *MFNS* 4:499; 4:508), which serve to explain motion and rest. It is crucial to note that these are powers that matter has

[84] It is worth emphasizing once more that passivity is simply one kind of activity, specifically an activity of a substance that requires the exercise of an external power to alter its state. Thus, when matter is characterized as merely passive, Kant does not deny that it is active, but only that it is self-active or spontaneous. In the passage just quoted Kant claims that 'matter has no faculty for determining itself'. He is therefore denying that matter can move or stop moving by itself. Once we fully appreciate the fact that Kant's characterization of matter as wholly passive does not prevent us from accepting that material bodies are *active*, it is possible to make sense of the causal powers of material objects and of the dynamic interaction between material bodies despite Kant's characterization of matter as wholly passive. For example, I can explain the fact that a billiard ball (A) starts to move by appealing to the fact that another billiard ball (B) in motion collided with it. It is natural to want to say that the first billiard ball (A) is passive and the second (B) active insofar as the latter determines the movement of the former. This is fine, if we bear in mind that the 'activity' of the latter (B) is also only the exercise of a passive or receptive power. The second billiard ball (B) is also passive, since it did not start moving by itself, but required, once again, an external power to determine its motion and so on *ad infinitum*. It is only if we consider the activity of (B) to be incompatible with the exercise of a merely passive power that the characterization of matter as wholly passive seems problematic.

only relationally; that is, no body can begin to move or stop moving by itself, but always only in relation to another body that exercises a repulsive or attractive force on it. It is only if we conceive of the powers of attraction and repulsion in these terms that Kant can maintain that matter is passive, that is, that all movement and rest of matter is the result of an *external* power acting on a receptive capacity. So Kant holds that matter "has no faculty for determining itself" and therefore is always only determined by an external power. Kant calls such powers, insofar as they play a determining role in bringing about a change in the state of a substance, an *external principle*.

Consider now spontaneous beings. In the following passage, we see clearly that the spontaneity of living beings consists exactly in the fact that such beings *are* self-determining:

> Now because all matter is lifeless […], everything that belongs to life cannot come from matter. The act <actus> of spontaneity cannot proceed from an outer principle, i.e., there cannot be outer causes of life, for otherwise spontaneity would not be in life. That lies already in the concept of life, since it is a faculty for determining actions from an inner principle. (*ML1* 28:285; cf. *MFNS* 4:544)

A being can be said to be self-determining insofar as no external power plays a determining role in the changes of its state. Instead, the power of the substance brings about a change in its state in accordance with its own *inner principle*. [85]

[85] Cf. *ML1* 28:275. It is important to note that, despite the fact that matter is lifeless, Kant is not therefore saying that life is immaterial. Kant insists only that '*everything* that belongs to life' does not come from matter, and, in particular, that life requires an inner principle or spontaneity that cannot be found in matter at all, and for that reason is also *separate* from it. In what follows, it will be important to keep in mind that all finite living things are nevertheless embodied material beings. As embodied beings, animals are *both* self-active and passive, and accordingly have both receptive and spontaneous powers. Since Kant does not reduce causal powers to substances, it is possible for one substance to have more than one, and also different kinds of, power (Watkins 2005, 261-2). Moreover, the spontaneous power can act as an external power on the passive power, so as to actualize an accident in the substance whose powers it is. Kant calls this kind of interaction between a spontaneous and a passive power in the same substance or subject, self-affection. Thus, we can distinguish *mere* matter from living beings, by noting that there is an inner principle in the living being (a spontaneous power), which can determine the receptive capacity of that same being and thereby actualize an accident in it. For very helpful discussion, see Indregard (2014).

1.5 Mechanism and Teleology

The distinction between the two kinds of activity that I've just outlined is important because, as we will now see, only one of these two kinds of activity, the activity that characterizes wholly passive beings, can be considered 'mechanical'. By contrast, Kant calls the activity that characterizes spontaneous beings 'animal' or 'practical', and as I will show, the explanation of such beings requires an alternative, teleological form of explanation.

Kant explicitly identifies the activity of passive beings, whose determinations, or changes of state, always depend on external powers, as 'mechanical'. And he contrasts this 'mechanical' activity with the second kind of 'animal or practical' activity, which determines a power from an inner principle, and which we have seen is characteristic of spontaneous beings:

> But there are still two kinds of activities: one of these is mechanical, and is produced by an external power; the other is animal or practical. Here the power is determined from an inner principle. The faculty for acting according to satisfaction or dissatisfaction is the practical, active faculty of desire. (*ML1* 28:254; cf. *MVM* 28:449)

In contrast to matter, to which Kant says only two basic powers, those of attraction and repulsion, can be attributed, living beings have, one and all, this 'practical, active faculty of desire'.

Now, spontaneous beings require an alternative kind of explanation because Kant restricts the mechanistic explanations that make up natural science to substances that are wholly passive. As we saw in the above passage, mechanical activity is 'produced by an external power'. The powers of matter, motive powers, are therefore exactly those suited to mechanical explanation. Kant says that "[M]otive powers cannot work otherwise than by outer causes, they are therefore also *determined only externally*…" (*MVM* 28:450, my emphasis). Mechanical explanations are, therefore, the only kind of explanation allowed in natural science conceived of as a theory of

material nature or phenomenal substance.[86] Kant says explicitly that we can never appeal to the second kind of activity, spontaneity, to explain any event in nature:

> If I want to explain an event in the world, and I derive this from the general laws of nature, then that is a natural event. In the world as a series of appearances, we *cannot and must not* explain any event from spontaneity <*ex spontaneitate*> (MM 29:862, my emphasis)

Now it is clear that everything in nature, insofar as nature is simply considered as a whole of appearance, and is therefore considered as phenomenal substance or matter, must be explained mechanistically. For the states and changes of bodies in space and time are always explained by reference to an *external* power, which, if that power is again to be allowed as an explanation of the first event, is also an event, which in turn can be explained in terms of an external causal power and so on ad infinitum. In fact, this is exactly how Kant conceives of natural or phenomenal causality:

> In all appearances of an event the causality of the cause of the event is itself an event. Now if all causes themselves have causes, then there is nothing in the world except nature. Now since there is nothing in the sensible world except events, we can go to infinity; everything that we will experience will still be either event or effect. For were it not an event, it would not be an object of experience at all. (*MM* 29:862)

We see that, on Kant's account, the scientific explanation of changes in phenomenal substance must always be given in terms of an external power, and, given that matter is entirely passive, every external power itself needs explanation in terms of a further external power. Each of these external powers must therefore be regarded as an effect. In science we must explain all determinations, that is to say, all events in the series of appearances that constitute experience,

[86] Cf. "We can and should be concerned to investigate nature, so far as lies within our capacity, in experience, in its causal connection in accordance with merely mechanical laws: for in these lie the true physical grounds of explanation, the interconnection of which constitutes scientific cognition of nature through reason." (*CPJ* 20: 236)

solely in terms of the first kind of activity, that is, as determinations of the receptive capacity of material substances by external passive powers.

We can now turn to the non-mechanical or 'practical' activity that characterizes spontaneous beings. It is useful to start by considering Kant's conception of life more carefully. Kant says that

> *[t]o live*, properly speaking, means to have a faculty for performing actions in conformity with one's representations. We call an animal *alive* because it has a faculty to alter its own state as a consequence of its own representations. (*MVM* 28:450) [87]

Because the spontaneity that characterizes living beings requires that something act in accordance with a representation, all living beings have the power of representation.[88] Notice, however — and this is a point that will be important later on — that the spontaneous power in the non-human animal is not the power of representation itself; instead, it is the power to *act in accordance with* representations that is said to be spontaneous. [89]

---

[87]Kant makes a similar point in the *Metaphysik Volckmann* (MVM 28:450) and also in *Metaphysik L1*: "[L]iving beings which act according to this inner principle must *act according to representations.*" (ML1 28:248) We also encounter this conception of life in his published work:

> Now we know no other internal principle in a substance for changing its state except desiring, and no other internal activity at all except thinking, together with that which depends on it, the feeling of pleasure or displeasure, and desire or willing. But these actions and grounds of determination in no way belong to representations of the outer senses, and so neither [do they belong] to the determinations of matter as matter. Hence all matter, as such, is lifeless. (*MFNS* 4:544)

[88] It is, admittedly, difficult to explain exactly how or why plants should be understood to have such a power. It is nevertheless possible to show that this characterization of spontaneity and of living beings is central to Kant's understanding of them (see previous fn). In any case, it is not impossible to consider plants to have some very basic representational capacities, given how they respond to their environments, even if we do not have a well worked out model of how that system of representation is meant to work. For a thorough discussion of Kant's views on plants, see Nunez (2021).

[89] At this point, it is important to clarify that we must still make a distinction between the human and the animal faculties of desire. One might think that to do so we must insist that the non-human animal's faculty of desire is passive in just the same way that its cognitive faculties are, but this would be a mistake, for the animal clearly has a power to change its state (to eat, for example) in accordance with its representations (hunger). We can accommodate the spontaneity of the animal's faculty of desire if it makes sense to think of spontaneity as a matter of degree. At least on the face of it, this seems something to which Kant is open (see, for example, *G* 4:452, A547/B575, *ML1* 28:267). What we can then note is that the non-human animal

Now if, as I have argued, mechanistic explanation is by its nature restricted to passive causal powers, it follows directly that spontaneity is outside the explanatory remit of the natural sciences. Kant explicitly accepts this conclusion in the case of animal life, for he criticizes thinkers like Descartes and Malebranche for holding that that animals "act only according to general laws of matter" and thinks that such a position is "utter foolishness" (*MVM* 28:449).

In the second part of the *Critique of the Power of Judgment*, Kant develops an alternative mode of causal explanation that is suitable for irreducibly spontaneous beings, namely, teleology.[90] This kind of explanation can make sense of a being that acts in accordance with its representations, for it is a mode of explanation that cites ends or purposes. We saw that when the activity of a

---

is less spontaneous than the human being, primarily because it not only can but *must* act in accordance with its sensible representations. As Kant puts it, "For a power of choice is **sensible** insofar as it is **pathologically affected** (through moving-causes of sensibility); it is called **animal** power of choice (*arbitrium brutum*) if it can be **pathologically necessitated**" (A534/B562). Now, it is clear in this passage that the human power of choice is pathologically affected but *not* necessitated. Consequently, the human power of choice is "not *brutum* but *liberum*, because sensibility does not render its action necessary" (A534/B562). Kant calls this "independence of the faculty of choice from **necessitation** by impulses of sensibility" practical *freedom* (A534/B562), and he tells us that this freedom is grounded on transcendental freedom, or what we might also call *absolute* spontaneity (*ML1* 28:267). There is no doubt that Kant takes non-human animals to lack absolute spontaneity altogether. What distinguishes the human faculty of desire is the fact that it is governed by practical reason, and is therefore absolutely spontaneous. It is not only that the human being, like the animal, has the power to act in accordance with (its sensible) representations; the human being can *choose* whether to act on those representations. Spelling out the exact distinction between animal and human desire requires a more thorough understanding of the distinction between relative and absolute spontaneity that I discuss further in Chapter 5. Thanks to Brian Tracz for questions regarding the nature of the non-human animal faculty of desire.

[90] Now, one might worry that I am overgeneralizing, and that at least some spontaneous beings, like crystal formations, are explicable mechanically. But although Kant sometimes insists that crystal formations must be explained by appeal to mechanical laws (*CPJ* 5:419), he nevertheless also appears to think that to we can identify a certain systematicity in the form of crystals that requires that we think of nature as a proceeding *technically* rather than mechanically in their formation (20:218). Accordingly, Kant says that

> [w]ith regard to its products as aggregates, nature proceeds mechanically, as mere nature; but with regard to its products as systems, e.g. *crystal formations*, various shapes of flowers, or the inner structure of plants and animals, it proceeds technically, i.e., as at the same time as an art (*CPJ* 20:218, my emphasis).

Kant goes on to identify the technique of nature as "the causality of nature with regard to the form of its products as ends" *CPJ* 20:219). Thus, although Kant is committed to finding mechanical explanations of phenomena as far as possible, at least some features of natural phenomena, like crystals, display a kind of organization that is not explicable mechanistically, and we might think it is exactly because they have those features that Kant conceives of them as, in some sense, spontaneous.

substance is passive, there is no problem about the possibility of offering a mechanical explanation of the changes of its state. Every explanation will appeal to the external cause that affected the substance and in so doing brought about the existence of an accident in it by triggering the receptive capacity through which the accident inheres. But when a substance acts spontaneously, there is no external cause that we can appeal to in order to explain the change of its state. Instead, we must appeal to something inside the substance that directs and produces the change. Rather than surrendering spontaneous action as rationally unintelligible, Kant adopts an alternative teleological mode of explanation that is precisely designed to explain how the changes of a substance flow from its inner nature. The nature of the substance is just the faculties and inner principles that we must appeal to in order to explain the change, and Kant conceived of these, in Leibnizian fashion, as inner strivings or endeavors that aim to actualize certain accidents; he says, "[b]etween faculty and power lies the concept of endeavor <*conatus*; G: *Bestrebung*> (*ML2* 28:565). Teleological explanation explains how the accidents of a substance result from these inner strivings and thus makes intelligible how a substance's states could be grounded by its inner principles.

In the case of animals, teleological explanation invokes the notion of pleasure (*Lust*) to make sense of how an animal could act in accordance with its representations, and hence for an end. Kant explains that pleasure arises through the satisfaction of an interest, and in the case of living beings this interest stems directly from the end that an organism has of surviving and flourishing (cf. *CPJ* 5:221). For present purposes, the full details of Kant's account of animal pleasure need not detain us;[91] the important point to note is that, by situating the animal's representations in the context of an end for the sake of which it acts, teleological explanation, in

[91] See Newton (2017) for helpful recent discussion.

contrast to its mechanistic counterpart, is tailor-made to explain the causality by which a being acts in accordance with inner principles.

What is important to emphasize, though, is that my claim here is not restricted to the case of animals. The claim is not simply that Kant takes living beings or animals to require teleological explanation, since there is no question that this is Kant's view. Instead, I have tried to show that living beings or animals must be explained teleologically *because* they are spontaneous. The following passage from the third Critique, where Kant says that the 'principle of ends' or teleology is a 'necessary maxim of reason', makes this connection between teleology and spontaneity particularly clear:

> it is […] necessary to conceive of a particular kind of causality […] that is not, unlike the mechanism of natural causes, found in nature, since to the receptivity […] of which matter is capable in accordance with that mechanism there must still be added the spontaneity of a cause (which thus cannot be matter) without which no ground of those forms [organized beings] could be given. (*CPJ* 5:411)

In this passage, Kant explicitly ties the need for teleological explanation to the existence of a spontaneous mode of causation that underlies the forms of organized beings. Teleological explanation is therefore required because there is a special kind of causality that we see manifested in nature, primarily in living beings, but that also, as we will shortly see, underlies the mind conceived of particularly as an intellectual power.

Section 2: Cognitive Spontaneity

Kant often characterizes the mind and its various faculties in teleological terms that directly pick up the account of teleological explanation developed in the third Critique. In the third Critique, he claims, for example, that each faculty that forms part of the mind has, as part of its

nature and so as a constitutive 'principle', an interest or a goal (cf. also A642/B670). We can think of this interest in broadly Aristotelian or Leibnizian terms as a striving to realize itself as the power that it is:

> To every faculty of the mind one can attribute an interest, that is a principle that contains the condition under which alone its exercise is promoted. (*CPJ* 5:120)

It should, therefore, not come as a surprise that Kant characterizes the understanding, too, in explicitly teleological terms. He says that "the understanding, … is aimed at an end which is necessary for it" (*CPJ* 5:187, my emphasis), and that we cognize "the lawful unity in a combination … as in accordance with a necessary aim (a need) of the understanding" (*CPJ* 184, my emphasis). In both of these passages, Kant makes it clear that this is an aim or end that the understanding has essentially and not merely accidentally or voluntarily, and this is exactly what we should expect given that the understanding is spontaneous and given the connection we have seen between teleology and spontaneity. In this context, it is also worth emphasizing that Kant characterizes the mind and not only the organism as alive: "the mind for itself is entirely life (the principle of life itself)" (*CPJ* 5:278); thus, it requires no stretch of the imagination to suppose that Kant would apply the same general form of explanation in his theory of mind as we have seen characterizes his account of animal life. Still, we do not yet have a purchase on precisely what a teleological explanation of the exercise of the cognitive faculty would look like, and how this mode of teleological explanation is related to the more familiar form of teleological explanation that applies to animal life.

To see more clearly how the notion of spontaneity we have been developing is relevant in the context of Kant's characterization of our higher cognitive faculty, it is helpful to start by considering what it is that Kant thinks sets us apart from non-human animals. As we saw in Section 1, Kant takes non-human animals to be spontaneous. That means that they are *alive*; they have the

power to act in accordance with their representations. But it does not follow from the fact that an animal has the power to act in accordance with its representations that its *representational* powers are themselves spontaneous. In fact, on Kant's account, the representational powers of non-human animals are always only passive. The spontaneity of the animal is not attributable to it in virtue of the nature of its representational powers; instead, the animal is spontaneous in virtue of its non-cognitive, desiderative power to *act in accordance with* its representations.

Moreover, Kant very clearly locates the passivity of animal representation in a specific feature of their representational powers — namely, that those powers are not *conscious*:

> Accordingly we ascribe to these beings (non-human animals) a faculty of sensation, reproductive imagination, etc., but all only sensible as a lower faculty, *and not connected with consciousness*. (*ML1* 28:277, my emphasis)[92]

It is this feature of animal spontaneity that is the key to distinguishing it from human spontaneity. For though the human, like the animal, has a spontaneous *conative* faculty, the human being *also* has conscious cognitive capacities, and these capacities, unlike the non-human animal's cognitive capacities, are also spontaneous. Humans, unlike animals, are *cognitively spontaneous*.

On my reading, Kant takes consciousness to be a special kind of spontaneity or self-activity; that is to say, it is a kind of activity that determines the state of the subject in accordance with a special kind of inner principle that is constitutive of the subject's nature as a *rational being* rather than as an *animal*. On Kant's view, the exercise of the higher cognitive faculties of the human being, reason and understanding, depends on this activity. We can therefore distinguish between cognitive spontaneity and the spontaneity that attaches to a being simply in virtue of being alive, by noting the difference between the principles that govern or determine the relevant powers

[92] Kant expresses this commitment in lectures throughout his career; see, for example, *MM* 29:888; *MVM* 28:449; *ML2* 28:594; *MD* 28:689.

of the being. Because *life* is characterized by spontaneity insofar as all living beings have a power to act in accordance with an inner principle, that is to say, have the power to act in accordance with their own representations, the activity is wholly *practical*; it is the kind of activity that a being engages in so as to prolong or sustain its life. The inner principle here is simply the nature of the animal itself, for that principle is what determines how the animal must act in order to survive. *Consciousness* is also characterized by spontaneity, but the activity it produces is *theoretical* rather than practical. In both cases, the self-activity of the substance actualizes an accident in it, that is to say, determines the state of the substance. But unlike the inner principle of the living being, consciousness is an inner principle that belongs to the nature of a rational being. In the first case, the state is something that the subject *does*: eating, or drinking, or doing whatever else is required to promote and sustain its life. Typically, we simply say that the subject acts, but, given the confusion that label would cause in the present context, it would be better to call these particular activities *doings*. In the second case, the state of the subject that is determined by its self-activity is not a *doing*, but a representational state, a *thinking*.

Consider the following passage, in which Kant draws a distinction between the representational powers of human and non-human animals:

> The irrational animal <perhaps> has something similar to what we call representations […], but which may perhaps be entirely different, but no cognition of things; for this requires understanding, a faculty of representation *with consciousness* of action whereby the representations relate to a given object and this relation may be thought. - However, we do not understand anything correctly <according to form> except that which we can make at the same time when the material for it would be given to us. Consequently, *understanding is a faculty of spontaneity in our cognition, a higher faculty of cognition, because it submits representations to certain a priori rules and itself makes experience possible*. (*APV* 7:141n; my emphasis)

What this passage makes clear is that it is in virtue of the fact that human beings have consciousness, a special self-activity of the understanding, that it is possible for them to represent the world objectively, that is, in accordance with rules and laws that make experience possible.[93]

As we have seen, on Kant's view non-human animals only ever represent something through an external power; thus, their power of representation is passive or receptive. But human beings have the power to represent according to an inner principle; that is in effect what it means to say that they are conscious or apperceptive. Now, we have seen that acting in accordance with an inner principle in the non-cognitive case involves acting in accordance with one's representations. We can find an analogy of this in the cognitive case, which would require us to think of consciousness as the faculty that makes it possible for us to represent things in accordance with an inner principle, where this inner principle is a certain formal organizing principle of conceptual representation. In this case, it is perhaps more appropriate to talk of an ability to represent 'in accordance with' a certain *form* of representation, where that form also constitutes the nature of the faculty of cognition. In the case of the understanding, as I will explain in more detail in Chapter 4, this will simply be the form of thought as Kant characterizes it in the Transcendental Analytic. It is clear that this form is itself the ground of a set of representations that have a special status: they are the *a priori* rules of thinking that Kant calls the categories.

As a result, not only can the human being act in accordance with representations that allow her to preserve, sustain and promote her life; she can also *represent* the world in accordance with

[93] It is important to emphasize that in denying that animals are conscious, Kant is not denying that there is 'something it is like' to be an animal. Kant's notion of consciousness, at least in this passage, is a long way from the contemporary notion of phenomenal consciousness. Consciousness, for Kant, is a pre-condition of conceptual intellectual representation, but it is not a condition of phenomenology. Therefore, in denying that animals are conscious, we need not read this passage as denying that animals can have phenomenal states with certain contents. For helpful discussion of animal consciousness in Kant, see Callanan (2020); Golob (2020); and McLear (2011; 2020b).

certain a priori representations.[94] These latter representations are rules or laws that allow her to understand and to make the world intelligible. While life is the inner principle of the conative faculties of all living beings, insofar as the promotion of life is the end or goal of the faculty of desire in animals, the inner principle of the cognitive faculty of human beings must be understood as an inner principle of intelligibility that allows the subject to represent the world as a whole of lawfully interconnected objects. This inner principle of intelligibility, I will go on to suggest in the next chapter, is nothing other than the pure consciousness or apperception through which a subject's representations are connected in accordance with rules. In the next chapter, I build toward a complete explanation of how self-consciousness is related to the possibility of representing the world objectively; for now, all I want to draw attention to is that we have a model for explaining what consciousness (or apperception) is. For Kant very explicitly identifies pure apperception with self-activity and spontaneity (B132), and as I have argued above, spontaneity is the activity of determining the state of a substance in accordance with an inner principle.

If we put these points together with the findings of the previous section, we see that Kant thinks that we must explain cognitive spontaneity as a causality in accordance with ends. That Kant conceives of living beings in teleological terms is not in any way controversial. It is also uncontroversial that Kant takes the faculty of reason to be a causality in accordance with ends. This is particularly obvious in his practical philosophy, in which it is clear that Kant takes reason to be a power to act in accordance with a very special kind of representation, namely, the moral law.[95] What has been less well documented by Kant commentators is that the causality that reason

[94] To be clear, consciousness is simply representing in accordance with a rule, and not in accordance with the representation of a rule. I am not suggesting that in order to consciously represent, for example, a tree, I need to represent to myself a rule in accordance with which I can then represent the tree. Instead, representing a tree consciously just *is* to represent the various features in a rule-like way.

[95] See the following passage from the *Groundwork*: "The will [practical reason] is thought as a capacity to determine itself to acting in conformity with the representation of certain laws. And such a capacity is only

manifests in the practical domain through its self-activity, and that is clearly teleological, is no less present in Kant's theoretical philosophy. Because the cognitive power is a spontaneous power, teleological explanation has no less secure a place in his transcendental psychology than it does in his practical philosophy or his theory of organisms.

Section 3: Sensibility and Spontaneity

Before closing, I would like to take the opportunity to engage with a natural challenge to the reading of spontaneity that I have been developing. In characterizing spontaneity as I do above, as an activity that brings about a change in a substance in accordance with an inner principle, one might worry that I inadvertently commit myself to the claim that sensibility, too, is spontaneous. Kant, after all, holds that sensibility has a certain *form*, which he characterizes as the *a priori* 'principles' of sensibility that can be studied in the Transcendental Aesthetic:

> In this investigation it will be found that there are two pure forms of sensible intuition as *principles* of a priori cognition, namely space and time … (A22/B37, my emphasis; cf. A21/B35)

In this passage, Kant is plausibly read as holding that changes in the state of a subject that are brought about through sensibility are determined in accordance with inner principles, namely, the pure forms of sensible intuition. If that is right, one might worry that on my reading of spontaneity, sensibility too must count as spontaneous. This would, of course, be a disastrous consequence for my view: given that Kant explicitly and repeatedly claims that sensibility is a passive receptive capacity, and not a spontaneous faculty, if my reading of spontaneity entails the denial of this

found in rational beings" (*G* 4:428). We will discuss Kant's account of practical reason in the Concluding Postscript.

claim, then either Kant was grossly inconsistent or the reading that imputes to him the inconsistency is mistaken.

This concern, though natural, overlooks an important distinction between ways in which an internal principle can inform the changes in the state of a substance. An inner principle can be *sufficient* or *insufficient* to bring about a change in the state of a subject, and it is only in the former case that a substance's activity in accordance with the principle, and the power that is thereby exercised, can be regarded as *spontaneous*. To see this more clearly, consider the fact that *all* activities are inner principles that ground the inherence of accidents, but, as we saw above, only some of these activities are spontaneous; the others, those that are passive, correspond to exercises of certain capacities that Kant calls receptivities. These activities, too, are inner principles that form part of the nature of the subject, but they are not sufficient, by themselves, to bring about a change in its state. In order to ground an accident, these inner principles must cooperate with an external power, which Kant also calls an external principle of change. This distinction between sufficient and insufficient internal principles, which in turn corresponds to the further distinction between spontaneous and receptive powers, gives us the resources to respond to the objection. We can grant that sensibility has inner principles consistently with denying that it is a spontaneous power as long as we maintain that those principles are insufficient rather than sufficient. Thus, we can grant that the forms of intuition are inner principles of sensibility as long as we deny that these forms are sufficient to ground the inherence of intuitions in a subject.

Now, though this reading seems to do a good job of explaining the role of the forms of intuition in *empirical* intuition – since in that case it is clear that, in addition to the form of sensibility, an external power must affect the senses if an intuition is to inhere in the subject – one might worry that it is ruled out by Kant's doctrine of *pure* intuition. Pure intuition, one might

worry, is exactly a modification of sensibility that does not depend on any external power but solely on the form of intuition inherent to sensibility. The pure intuition of space, for example, might be thought of as an a priori intuition of sensibility whose content is determined solely by the form of outer sense. Thus, it might appear that the internal principles of sensibility *are* sufficient to produce at least some representations, namely, pure intuitions.

The doctrine of pure intuition is a notoriously difficult part of Kant's theory of sensibility, and a full account is beyond my scope here, but in what follows I will show that there is at least one plausible reading of the doctrine on which it is consistent with my view of spontaneity. The objection we are considering presupposes that the representational power that produces pure intuitions is *sensibility*. This assumption, however, requires further argument since Kant distinguishes a second representational faculty that is *also* capable of producing intuitions, namely the imagination (*Einbildungskraft*).[96] If it is the imagination rather than sensibility that produces pure intuitions, nothing in Kant's account of pure intuition will require us to hold that the inner principles of sensibility suffice for a modification of *sensibility*, and since Kant characterizes the imagination as a spontaneous faculty in several places (B151-2; B162n; cf. *DWL* 24: 706), his account of pure intuition will not require me to mislabel a receptive power as spontaneous.

While a full treatment is beyond my scope here, it will be helpful to briefly preview some of the textual considerations that speak in favor of the view that imagination produces pure intuition. First of all, Kant's explicit characterizations of the faculty of imagination strongly

[96] Now, I should note that the claim that the faculty of imagination produces intuition is not uncontroversial. Brian Tracz (2020) has recently argued that imagination produces images rather than intuitions, but, as Tracz acknowledges, this view cuts against the more widespread view that imagination does produce intuitions. For an argument in support of the majority view, see Stephenson (2015a). Since my aim in what follows is simply to outline a plausible reading of the doctrine of pure intuition on which it is compatible with my reading of spontaneity, I do not engage with Tracz's arguments here.

suggest that it is the kind of faculty capable of generating pure intuitions. Consider the following two characterizations of the imagination:

> Power of imagination – also applies to the form of things {without the object being present}(*MD* 28: 673)

> [Sense] is the faculty of intuition in the presence of an object, [imagination] is intuition even *without* the presence of an object. (*APV* 7:153; cf. *APV* 7:176; *MM* 29: 881; B151)

The pure intuition of space is an intuition *without* the presence of an object; rather than being determined by any object, it is simply the representation of the form of intuition. Now, both of these passages suggest that intuitions of the imagination can be had even without the presence of any object, and the first passage specifically ties intuitions of the imagination to the form of things. Taken together, they suggest that the power of imagination is precisely the kind of faculty capable of generating intuitions that do not represent any object and whose content is determined solely by the form of intuition. Further support for this view of the origin of pure intuition comes from the Table of Nothing at the close of the Amphiboly. There, Kant lists 'pure space and time', which he calls 'empty intuitions without objects' (A291/B347–A292/B348), as examples of *entia imaginaria*, which strongly suggests that the pure intuitions of space and time are generated by the imagination.

Now, these considerations are not meant to be conclusive, but I hope to have shown that there is at least one *prima facie* plausible line one could take about pure intuition on which the doctrine is entirely compatible with my reading of spontaneity. Though more work would of course need to be done to substantiate this reading, I hope to have shown that the doctrine of sensibility does not yield a short argument against my view, because it is not at all obvious without further argument that my view entails that sensibility is spontaneous.

Conclusion

Throughout this chapter, I have argued that the deflationist view, insofar as it is committed to a mechanistic conception of the understanding, misses the crucial connection between Kant's doctrine of spontaneity and his theory of causation. I have argued that once we see that connection, it follows that this mechanistic conception is untenable, and the onus is on us to develop an alternative, teleological conception of the faculty of understanding.

Now, one might think that, in stressing the connection between spontaneity and teleology as I have done, my own positive proposal falls straightforwardly into the camp of cognitive agency views, the criticisms of the previous chapter notwithstanding. But this is not necessarily so. We see this when we note that many of the entities that Kant treats as paradigmatically subject to teleological explanation — works of art, clocks, crystal formations, and so on — are *precisely not* beings that possess the kind of freedom that the cognitive agency view identifies with spontaneity. Perhaps there is a version of the cognitive agency view that could build on some of the points made in this chapter, but be that as it may, there is at least logical space for a version of the teleological reading that resists assimilation into either of the existing interpretive options, and in Chapter 5, I argue in support of just such a reading. There, I show that a proper understanding of Kant's distinction between *absolute* and *relative* spontaneity allows us to explain how the understanding could be essentially teleological and yet function unintentionally and automatically, in a way that precludes us from regarding it as subject to the kinds of normative demands that would allow us to ascribe to it any kind of agency.

But before we turn to that issue, I want to turn, in the next chapter, directly to the topic of apperception, since we now have a strong enough grasp of Kant's account of spontaneity to begin the articulation and defense of the formalist proposal. Drawing on the model developed in this

chapter, my aim in Chapter 4 will be to explain in detail how transcendental apperception provides the inner principle that constitutes the human understanding as a spontaneous faculty of thought.

CHAPTER 4

Transcendental Apperception as a Condition of Thinking

In this chapter, I would like to bring together the resources developed in the previous two chapters to articulate and begin to defend the *formalist* model of transcendental self-consciousness. Drawing on my claims about the spontaneity and teleological character of the faculty of understanding, I defend a model that contrasts with standard intentionalist conceptions. The formalist model that I propose rejects the primary assumption that is shared among intentionalist models, namely, that transcendental apperception is a consciousness of some object, for I claim that transcendental self-consciousness is not consciousness *of* anything. Instead, transcendental self-consciousness should be understood as the activity of the subject through which thought first becomes possible. Rather than regarding transcendental self-consciousness as a species of object-directed representation, my reading makes transcendental apperception a condition on the possibility of all conscious object-directed representation (or thinking).[97]

[97] I will say something in Section 1 to clarify this identification of thought with conscious object-directed representation.

The main aim of this chapter is twofold. First, I want to motivate the claim that the systematic role of transcendental apperception within Kant's theory of mind is to explain *how thought is possible*. Second, I want to explain in some detail exactly how it is that Kant thinks that transcendental apperception makes thought possible. On my interpretation, it is in the Transcendental Deduction that Kant undertakes the daunting task of spelling out this connection between apperception and thought, and so this chapter will help us to understand at least one element in the complicated argument of the Deduction. We will see that a proper understanding of the way in which transcendental apperception makes thought possible requires us to investigate in more detail what Kant means when he says that the distinctive unity that characterizes apperception is a *qualitative* unity.[98] Understanding this part of Kant's account, I show, allows us to appreciate both the sense in which the unity of apperception is a merely formal unity, and the sense in which apperception underlies the teleological structure of the understanding.

The chapter is split into two sections. In Section 1, my aim is to identify the contribution that apperception makes to the cognitive faculty and its characteristic representation, thinking. I start by explaining in general terms what Kant takes thought to be; in the course of this investigation, I explain the relationship between thought and *combination* (*Verbindung*), and, on this basis, I identify the role of apperception in making thought possible. Apperception, I argue,

[98]The topic of qualitative unity has been under-studied in Kant scholarship, and where it has received attention, its relevance has tended to be confined to Kant's aesthetics. Melissa Zinkin (2006), whose treatment is the most sustained available, develops her interpretation of qualitative unity primarily in order to make headway with Kant's account of judgments of taste in the third *Critique*, and Rachel Zuckert (2007: 43n, 223n) also appeals to the notion, as well as the related notion of qualitative *perfection*, in her account of purposive form. But to my knowledge there has been no sustained treatment of qualitative unity in the context of the first *Critique*. In Longuenesse's landmark study of the Analytic, for example, the notion is relegated to a footnote (1998: 242n), and in Allison's more recent study of the Deduction, it is only treated in passing (2015, 334-35). Outside of readings of the Deduction, the notion receives passing mention in Newton's (2015) account of empirical concept-formation, but none of these authors accords qualitative unity the centrality I will claim it deserves in an interpretation of Kant's theory of transcendental apperception.

contributes a distinctive formal unity to a representation, through which its constituents are combined. Note that in this section, my aim is only to identify the role of apperception; I do not yet provide a systematic explanation of *how* it performs that role. In Section 2, I begin my systematic account of the role of apperception within the Deduction by turning my attention to Kant's claim in section 15 of the B-Deduction that the unity of apperception is a *qualitative* unity. Understanding what Kant means when he describes the unity of apperception in this way will position us to understand *how* it is that Kant thinks apperception grounds the possibility of thinking, and in particular how it makes possible the kind of unity that we have seen is required for thinking.

Section 1: Transcendental Apperception and Thought

As I have shown in Chapter 1, the vast majority of interpretations of Kant's conception of transcendental self-consciousness share a fundamental commitment to the claim that transcendental consciousness is a consciousness of a particular object. On the view that these interpretations share, self-consciousness is just a peculiar case of consciousness more broadly conceived of as a state of a subject that is of or about something. The starting point is therefore the more 'familiar' phenomenon of object-consciousness, which can be understood straightforwardly in intentionalist terms, that is, in terms of what the conscious state represents. Self-consciousness, on this view, is peculiar just insofar as the object that this consciousness is supposed to be of, or that it is supposed to represent, is not properly speaking independent of it.[99] One basic feature that

[99] Depending on the version of intentionalism that one wishes to endorse, this object might be the subject or the self who is conscious; it might be the activity of consciousness; or it might just be the content of the conscious state of the subject.

all intentionalist conceptions of self-consciousness have in common is thus that they take object-consciousness to be explanatorily (and perhaps metaphysically) prior to self-consciousness since self-consciousness is just a species of object-consciousness in which the object is meant to be in some sense internal to the subject. In chapter 1, I suggested that, for Kant, self-consciousness is both explanatorily and metaphysically prior to object-consciousness, and, therefore, that we should not interpret Kant as accepting an intentionalist view. I also suggested that endorsing an intentionalist model of self-consciousness, regardless of the variety, would introduce an element of reflexivity or self-directedness into transcendental self-consciousness, which, I believe, we have good reason to resist in interpreting Kant's theory of transcendental apperception.

My overarching aim in this chapter is to motivate an alternative model of transcendental apperception, and in the present part my aim is to articulate this alternative model and provide a preliminary textual basis for it. On the model that I would like to propose, transcendental self-consciousness is not reflexive: it is not directed at anything (nor is it about anything) that is in some sense internal to the subject, whether that is the subject itself (or its nature), an activity of the subject, or a state of the subject. Transcendental self-consciousness on my view is not directed at the self at all. One might think that I therefore endorse a view according to which self-consciousness is indistinguishable from consciousness more broadly, understood simply as consciousness of objects. If so, I would be collapsing the distinction between consciousness and self-consciousness entirely. Yet that also is not my view, for I will claim that on Kant's view self-consciousness is not in fact consciousness of any particular object at all, but is rather the form of the consciousness of all objects. On this view, self-consciousness is the formal condition of a certain kind of representation of an object, which Kant calls *thought*.

I proceed as follows. To make the case for this formalist conception, I will start by situating transcendental self-consciousness within the metaphysical framework that I developed in the previous chapter. I show that transcendental apperception is a kind of spontaneous activity or self-activity of the subject, where 'activity' must be understood as the ground of inherence of an accident in a substance. I go on to show that pure apperception grounds the inherence of thoughts in a subject, and that it does so by providing the organizational form or principle that makes a distinctive kind of unity of representation possible, a unity that Kant calls collective unity.

In keeping with the metaphysical framework developed in the last chapter, I will assume that the understanding is a faculty and that its activity expresses a power;[100] and the general power that the understanding manifests is the power of *thinking* (A69/B94). The following passage makes clear what is entailed by the claim that the understanding is a faculty of thinking:

> 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 … 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. (*MM* 29:771)

So, to say that the understanding is the power of thinking is to say that the exercise of the understanding is an activity of the subject, through which certain accidents — thoughts — can inhere in that subject.[101] And, moreover, we know that a subject is capable of such activity just in case it has the relevant faculty — in this case, the faculty of understanding.

---

[100] For the way in which the tradition conceives of the relationship between faculty and power, see Introduction, Section 1.

[101] Note that Kant explicitly refers to the I as a 'subject of inherence' when he says for example that "the subject of inherence is designated only transcendentally through the I that is appended to thoughts" (A355). As we will see presently, the significance of the qualification that this subject is only designated 'transcendentally' is related to Kant's resistance to the rationalist's claim that we have insight into the nature of the substance in which thoughts inhere.

Now, Kant also tells us in this passage that underlying the faculty of thinking is a substance. It is nevertheless important to note that we do not cognize the nature of this substance as it is in itself.[102] In particular, Kant insists that we cannot infer from the claim that the faculty of thinking belongs to a substance that this substance in itself is simple, indestructible, or that it persists through time.[103] Thus, it is not particularly helpful or informative to know that the power expressed by the understanding is the power of a substance, since that fact by itself does not distinguish the substance in question from any other substance, and nor does it tell us anything about the ultimate nature of the subject. We can, however, make headway in understanding Kant's conception of transcendental apperception by focusing more specifically on the activity that makes it possible for thought to inhere in a subject, and through which the power of understanding is therefore manifested.

---

[102] Kant expresses this important point especially clearly in the following passage from his essay *What Real Progress Has Metaphysics Made in Germany since the Time of Leibniz and Wolff?*:

> Of the self in the first sense (the subject of apperception), the logical self as a priori representation, it is absolutely impossible to know anything further as to what sort of being it is, or what its natural constitution may be; it is like the substantial, which remains behind after I have taken away all the accidents that inhere in it, but absolutely cannot be known any further at all, since the accidents were precisely that whereby I was able to know its nature […] the logical self does indeed point to the subject as it is in itself, in pure consciousness, not as receptivity but as pure spontaneity, but beyond that is also incapable of knowing anything of its nature. (*WRP* 20:270-1)

But the point is also emphasized in the *Critique*, when he says that "it is obvious that the subject of inherence is designated only transcendentally through the I that is appended to thoughts, without noting the least property of it, or cognizing or knowing anything about it" (A355).

[103] Kant makes this point especially clearly in the Paralogisms. He claims that while it is correct to ascribe certain predicates to the I, these claims are not substantive synthetic a priori truths about the nature of the substance itself. In fact, he says, these claims are 'completely empty' because they do not apply to anything that can be given to us in intuition:

> Now mere apperception ("I") is substance in concept, simple in concept, etc., and thus all these psychological theorems are indisputably correct. Nevertheless, one by no means thereby cognizes anything about the soul that one really wants to know, for all these predicates are not valid of intuition at all, and therefore cannot have any consequences that could be applied to objects of experience; hence they are completely empty. (A400)

For an especially helpful discussion of Kant's arguments in the Paralogisms, see Ameriks (2000).

Recall from the previous chapter that Kant distinguishes between two kinds of activity: self-activity, or spontaneity, and passivity, or receptivity. This distinction now allows us to get a firmer grasp of the distinction Kant draws between the two cognitive faculties, sensibility and understanding. Kant claims that our sensibility can generate representations only if it is affected by external powers (objects).[104] These sensible representations (intuitions) therefore inhere in the subject through an external power that affects its receptive capacity. Thus, the activity through which accidents inhere in the subject as a result of the exercise of sensibility is passive. We can also infer from the conception of spontaneity that I defended previously that the activity through which accidents (in this case, non-sensible representations) inhere in the subject with an *understanding* is spontaneous: in this case, the representation does not inhere in the subject through the affection of an external power, but solely through an inner principle. Thoughts (the non-sensible representations Kant ascribes to the understanding) therefore belong to the subject (are 'in me') because of the determination of some representation by an inner principle. This much follows straightforwardly from the discussion of Kant's metaphysics in the previous chapter, yet it has important consequences for our discussion of transcendental apperception, for, as I will now argue, 'transcendental apperception' is the name of the activity through which thoughts inhere in the subject.

To begin to see the basis for this claim, we need a clearer account of what Kant takes thought to be. Now, we have seen that thought is the kind of representation that can only belong to a subject that has a faculty of understanding, for it is a kind of representation that requires the spontaneity of the subject. It is, therefore, to be contrasted with other kinds of representation, such

[104] Recall from Chapter 3 that a power is passive or receptive just in case it is not by itself sufficient to ground the inherence of an accident in the substance whose power it is. It requires, in addition, the exercise of an external power (that affects it), and Kant says that it is from their 'union' that 'an effect comes about' (*MM* 29:822).

as perception (*Wahrnehmung*) and images (*Bilder*), but most notably with intuition (*Anschauung*), which is a representation that is generated through sensibility when the senses are affected by an object. Kant claims that the object is 'given to us by means of sensibility' in intuition (A19/B33), but the object *given* in sensibility and represented in intuition is still 'nothing for me' unless I can *think* it, and to do so is to "be able to accompany … my representation" with the I think (B132). Thus, Kant argues in Section 16 of the B-Deduction that "all manifold of intuition has a necessary relation to the I think in the same subject in which the manifold is encountered" (B132), that is, that all of the manifold of intuition, insofar as its object is to be 'something' for the subject, must be capable of being thought.

Now, at the opening of the Transcendental Aesthetic, Kant makes a remark that might be taken to suggest that, conversely, all thought has a necessary relation to intuition, for he says that "all thought … must ultimately be related to intuitions … since there is no other way in which objects can be given to us" (A19/B33). Given Kant's positions elsewhere, however, this passage is best read not as claiming that all thoughts are in fact related to intuitions, but rather that they must be so related *if they are to become cognitions*. In several places, and most prominently in the B-Deduction, Kant makes it very clear that thinking is *not* restricted to what is given in intuition, and has an 'unbounded field' (B166fn); it is only cognitions, a subset of our thoughts, that are directed at the objects given in sensible intuition (cf. B147; A52/B76). However, though not all thought is directed at an object given in intuition, it is nevertheless the case that all thought is directed at *some* object; thought is cognition if the object that the thought is about is given in intuition, but it is mere thought if the object is not so given. In the latter case, Kant claims that "the thought of the object can still have … useful consequences" (B166fn), but it cannot directly contribute to our cognition of nature. I will therefore take thoughts to be conscious object-directed

representation, and I will claim that transcendental apperception is the formal condition of this kind of representation.

We can make progress toward seeing the exact nature of the role of apperception by considering the question how the manifold that is given in intuition comes to be thought. Kant's answer to this question is relatively straightforward: while the manifold "can only be given in the intuition … [it is] thought through *combination* [*Verbindung*] in a consciousness" (B135). Thus, if we can understand what 'combination' is and *how* the manifold is combined in a consciousness, we will have a decent understanding of how thinking is possible.

Fortunately, Kant opens the B-Deduction with an account of combination. There, he claims that "[a]mong all representations **combination** is the only one that is not given through objects but can be executed only by the subject itself, since it is an act of its self-activity [*Selbsttätigkeit*]" (B129).[105] Focusing on the combination that is the result of the 'act of [the subject's] self-activity' allows us to see that the activity in question only contributes one distinctive feature of the resulting representation, namely, its *unity*. Prior to combination, a representation already contains a manifold, and that manifold is already subject to synthesis, but the synthetic manifold is not yet represented as a 'synthetic unity' (B130-31). It is this distinctive synthetic unity that is contributed by the act of spontaneity that underlies combination. We therefore see that whatever else thought is, it is a kind of unity that is the result of an act of the subject that Kant says is spontaneous. In the very next section (Section 16) Kant goes on to identify this act as pure or original apperception (B132).

[105] It is important to note that the term 'combination' can be read as referring to the activity of combining, or it can be read as referring to the resulting representation. In the passage just quoted, Kant uses it in both senses: at the beginning of the passage, combination is referred to as a unique kind of representation, whereas at the end it is described as an *act* of the subject's self-activity. For our purposes here, I am using the term to refer to the representation that is the result of the activity of combining.

We have seen so far that original apperception is the ground of what Kant calls combination or also synthetic unity, which is exactly the kind of unity that a representation must have in order for it to be the thought. Thus, we know that the self-activity of the subject that Kant calls pure or original apperception (B132) is supposed to ground the inherence of certain accidents – thoughts – in a subject, and we know that it does so by bringing about a certain unity of representation. Now, Kant emphasizes that this act is 'an act of spontaneity' (B132), and it is crucial for my conception of this act as spontaneous that the synthetic unity that it makes possible is produced by an *inner principle*. Putting these points together, original apperception is an act that grounds the inherence of thoughts in the subject by bringing about a unity in the given manifold in accordance with an inner principle.[106]

To see more clearly how this activity might function to bring about the synthetic unity of representation that is required for thinking, we need to start by considering the nature of the unity that Kant claims thought must have and the way in which this unity is reflective of an inner principle. Kant insists that thinking requires a special kind of unity: He says, for example, that "the unity of a thought consisting of many representations is collective" (A353). The reference to *collective* unity in this passage is significant, for, as we will see in Section 2, collective unity is the kind of unity that depends upon a *principle*. And in fact, Kant identifies transcendental apperception itself as the specific principle that is required for the collective unity of a thought: In

[106] For my purposes, it is important only that grounding is here understood as a metaphysical rather than a logical relation, and that we can say that something A is the ground of something B if A is one of the things that makes B possible. As I understand it, this relation is asymmetric and tells us something about the metaphysical dependence of one thing (B) on another (A). For more on grounding in Kant, see especially Stratmann (2018). He draws a helpful distinction between two kinds of metaphysical grounds, material and formal (2018, 5). The latter ground does not determine the *existence* of a thing, but only its organizational form or structure. For reasons that will become obvious, I take transcendental apperception to be a *formal* and not a material ground. For a helpful introduction to the notion of grounding in Kant's Metaphysics lectures, see Stang (2019).

the A-Deduction, he says that "it should not go unnoticed that the mere representation I in relation to all others (*the collective unity of which it makes possible*) is the transcendental consciousness" (A117, my emphasis; note that earlier in this note, Kant identifies transcendental consciousness with original apperception).[107] Thus, it is transcendental apperception that makes possible the distinctive collective unity that is required for thinking. It does so, I suggest, by supplying the organizational principle in virtue of which the manifold contained within a thought constitutes a collective unity.[108]

We are now in a better position to understand the formalist element of my interpretation, for Kant characterizes this organizing principle as a 'formal condition' of thought. To see more clearly the connection between a formal condition and an organizing principle, we can take a moment again to reflect on what thought is. A thought, we have said, is a certain combination of (given) representations. Now, the manner in which those representations are arranged matters for the identity of the thought: a particular thought is the one that it is not only in virtue of the contents of its elements but also in virtue of their specific relations to one another. The thought that all metals are bodies, for example, is not identical to the thought that all bodies are metals, but this difference is not attributable to a difference in the constituents of the two thoughts; instead, it reflects a difference in what we can call their logical form. Once a thought has a given logical form, the order and relations of its parts become necessary for the identity of the thought. Consequently, each thought is a unity, a whole that is 'held' together by the logical relations

[107] Cf. "All empirical consciousness, however, has a necessary relation to transcendental consciousness... namely the consciousness of myself, as original apperception" (A117n).

[108] At this point, it might be helpful to say something about how I am thinking about the elements of the manifold within a thought. As I read him, Kant holds that any representation that is 'in me' is potentially subject to combination and hence a possible element of the manifold within a thought. Thus, these elements could be intuitions, perceptions, images, or concepts. Note, though, that these elements should be thought of as the *inputs* to combination. The result or *output* of combination is a representation whose contents are exclusively conceptual.

between its constituents. When I say that transcendental apperception is the organizing principle or the formal condition of thought, what I mean is that it is responsible for introducing this kind of logical structure into our representations.

To see that this is Kant's view, we can start by noting that he says "[transcendental apperception] is only the *formal condition*, namely the logical unity of every thought' (A398, my italics). This transcendental apperception is designated by the expression 'I' — for example, Kant identifies the I with transcendental apperception at A117, and A400 (also cf. 7:141) —[109] and he claims further that the I "contains a priori a certain form of thinking" (A685/B713). Apperception thus plays its role in making thought possible by supplying the distinctive logical form that characterizes thinking, rather than representing any distinctive kind of object. In the Paralogisms, Kant makes this point explicitly, saying that "[transcendental apperception] 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; for of it alone can I say that through it I think anything" (A346/B404, cf, also B411). In its function as a formal condition of thought, apperception is also a formal condition on cognition, since all cognition contains an act of thinking: "I have in thought merely the unity of consciousness that grounds everything determinate as the mere form of cognition" (B427, cf. also A231/B282). These passages make it clear that Kant conceives of transcendental apperception as a form of thought, and therefore of cognition, and that, as a condition of thinking, it itself does not represent anything but only makes a certain kind of representation possible. For a representation to be 'accompanied by' the I think, then, is not for

[109] He sometimes also uses the I to indicate the transcendental subject in general (cf, for example A123, B408, where Kant refers to 'the I of apperception'), but it is clear in those contexts that he uses the I to refer to the subject of apperception, and in that sense the subject is not conceived of as distinct of this activity or consciousness.

that representation to be *represented* in any given way (e.g., as *mine*), but rather for it to be combined in such a way that it inheres, as a thought, in a subject.

But even if we grant *that* transcendental apperception makes the collective unity of thought possible, we still do not know *how* it does so. All we know so far is that apperception makes possible the collective unity of a thought, and thereby functions as a principle of thinking, but we have yet to isolate the feature of apperception in virtue of which it is able to function as such an organizational principle. In the next part, I address this issue directly, and in so doing, I will turn my attention to what Kant calls the *unity* of apperception. We will see that the collective unity characteristic of a thought presupposes a prior unity that pertains to apperception as such, that is, to the act itself, and this unity Kant calls 'qualitative', or 'formal' unity. It will be important in what follows to keep track of the difference between these two unities. First, there is a unity that pertains to the act of apperception itself (cf. B130), which I will go on to argue is a kind of formal or qualitative unity. Kant calls this unity simply the 'unity of apperception' *simpliciter*. Second, there is the unity that pertains, not to the act of apperception itself but to the representations that are combined through this act, a unity that all conscious object-directed representations must exhibit. Kant calls this unity the *synthetic unity of apperception*. Finally, within the domain of this second kind of unity, Kant identifies the unity that pertains to *pure* representations that are combined through the act of apperception as the *original* synthetic unity of apperception.

In the next section, through a careful investigation of the qualitative unity that Kant ascribes to transcendental apperception, I will argue that pure apperception functions *teleologically* to bring about the unity of the pure manifold of given sensible representation (intuition), or what Kant calls the original synthetic unity of apperception, which is what makes it possible for the subject to *think* the objects of sensible intuition (appearances). We will see that qualitative unity

functions as a principle of organization, which brings about the required connections in the manifold given in intuition in accordance with a goal or an aim of intelligibility. Kant describes this unity as a *thematic* unity, which can best be understood by analogy with the unity that determines the organization of a work of art (like a play or a story), and which functions to establish connections within the material that render it intelligible. In the next section, we will spell out this unity and its teleological dimension in detail.

Section 2: The Qualitative Unity of Apperception

In Section 1, I outlined a conception of transcendental apperception according to which it is the activity of a subject that grounds the inherence of certain accidents in it, and I indicated that it plays this grounding role in virtue of having a certain kind of unity, which Kant designates as 'formal' and 'qualitative'. This unity is an inner principle that determines representations in accordance with a certain form — the form of thinking. Thus, the unity of transcendental apperception, as an inner principle that is a condition of thinking, constitutes a principle of intelligibility for finite minds like ours. In the rest of this chapter, I want to elaborate this final claim by considering in more detail how it is that transcendental apperception, so understood, makes it possible for us to think objects of sensible intuition. We need to understand better both i) what it means to say that apperception makes thinking possible, and ii) *how* exactly apperception makes thinking possible.

My discussion in this part is split into three sub-sections. In the first sub-section, I will provide a brief overview of Kant's project in the Deduction. Here, I am particularly interested in the role that apperception plays in making thought, as a condition of the cognition of objects, possible. In my treatment of the Deduction, I endorse a proposal by Watkins and Willaschek (2017)

according to which there are two conditions on the narrow conception of cognition that is Kant's primary focus in the Transcendental Analytic: a givenness-condition and a thought-condition. According to this conception, a subject can have cognition of an object just in case the object is given to the subject in some way (for our purposes the relevant way is through sensibility, but there might also be other ways in which objects can be given), and the object given is thought by the subject. I use this framework as a lens through which to view the Deduction, arguing that one of Kant's central aims in the Deduction is to show that and how the thought-condition on cognition can be met.[110] In the second sub-section, I build on this reading of Kant's project to argue that transcendental apperception plays an indispensable role in his account of how the thought-condition is met. I will show that, on Kant's view, it is because transcendental apperception makes possible a distinctive kind of unity, which he calls the *collective* unity of representations, that those representations can be said to inhere in the thinking subject. In the third section, I go on to explain *how* transcendental apperception functions to bring about the kind of unity that Kant claims is necessary for thinking. I argue that transcendental apperception functions teleologically, generating the connection of a manifold of representations for a *purpose* or *end*; and the unity thereby established in the manifold, which I call *teleological connection*, is what first makes it possible for the subject to *think* that manifold. Transcendental apperception is therefore an activity that grounds the relation of certain given representations (intuitions) to the subject as an intellect: It is in virtue of the activity of apperception that a sensibly given manifold can stand to the subject in a relation of *thought*. We can appreciate the teleological dimension of Kant's account by

[110] In thinking of Kant's project in the Deduction in these terms, as concerned primarily with the thought-condition on cognition rather than the givenness-condition, my reading is in broad agreement with Allais' (2015). According to Allais, '[Kant's] concern with 'relation to an object' in the Deduction is not about what it takes for us to have experience of perceptual particulars (to be presented with objects), but about the conditions of referential thought' (2015, 259).

focusing on the fact that Kant characterizes the unity of apperception as a *qualitative* or *formal* unity. As we will see, this unity must be understood as a kind of 'thematic unity' of thinking, which can be understood by analogy to a work of art.

2.1 Brief Overview of Kant's Project in the Deduction

Let me start by saying something briefly about how I understand Kant's project in the Deduction, with the following disclaimer: My aim here is not to propose a comprehensive account of the Deduction and nor is it to suggest that my description of his project captures the sole or even the primary concern of this important section. I simply want to explain the context in which I take Kant's theory of apperception to unfold in a way that highlights certain important concerns that he has, which are directly relevant to our understanding of the role of apperception. I therefore focus on aspects of the Analytic that are specifically helpful for articulating my account of apperception, and I exclude other important aspects that are not directly relevant to this account, including, among other things Kant's account of objective representation or cognition, the extent to which he is or is not responding to Humean skepticism, and his account of the imagination and synthesis.[111] It is not essential for my account of apperception that we accept all of the details of the project of the Deduction *exactly* as I propose here, either, but it nevertheless helps to have this framework in place for orientation.

With this disclaimer in place, on my understanding, one of Kant's central aims is to explain how it is possible that the very objects that affect our senses and which are represented sensibly in

[111] For a recent comprehensive commentary see Allison (2015). On the question of Humean skepticism and the nature of the argument in the Deduction, see especially Ameriks (1978), Strawson (1966), Dyck (2011), Kohl (2018), Hatfield (2003), Gomes (2010), Schulting (2019). For more comprehensive accounts of synthesis and the role of the imagination see Kitcher (1990) and Longuenesse (1998).

intuition can also meet the conditions of *thought* and can therefore be represented *intellectually*, which is to say, through the understanding. If sensibly given objects can meet that condition, which we will call the 'thought condition' following Watkins and Willascheck's account of cognition, then we have *cognition* of those objects, and accordingly, the concepts through which we think them, namely the categories, will have 'objective validity' (at least with respect to appearances).[112] One of the questions that Kant takes up in this important section is thus how sensibly given objects can meet the thought condition on cognition (supposing that they meet the givenness condition).

Now, I will also suppose that the givenness condition on cognition does not pose that much of a problem in the standard case, for on Kant's view, one way in which objects can be given is through affecting the senses, and those objects are represented sensibly as spatio-temporal appearances in intuition.[113] The sensible representations of objects therefore satisfy the a priori conditions on sensible representation simply in virtue of the fact that they are generated by affecting our sensibility. Accordingly, when Kant motivates the problem of the Deduction, namely how the subjective conditions of thinking could be valid with respect to the objects of intuition, he says that 'a difficulty is revealed here that we did not encounter in the field of sensibility' (A89/B122). Given that the objects of sensible intuition are not generated by the understanding itself, it is not so obvious why those same objects can be thought; that is to say, it is not so obvious why the representations of those objects can and must meet the a priori conditions on intellectual representation. Kant's aim in the Deduction is then to show that they can meet those conditions.

[112] As support for this reading, consider the way in which Kant states the preliminary result of the A-Deduction: "[The categories] are therefore… fundamental concepts for thinking objects in general for the appearances, and they therefore have a priori objective validity, which was just what we wanted to know" (A111). In this passage, Kant quite explicitly infers that the categories have objective validity from the fact that they are conditions for "thinking objects in general for the appearances".

[113] For my purposes in this chapter, it does not matter whether there are other ways in which objects might be given and if there are what they might be. For an extensive treatment of the notion of *givenness* cf. Watkins and Willaschek (2017).

He does so by showing that the conditions on the intellectual representation of objects, the categories, are products of the understanding that apply to appearances (the objects of sensibility).

My focus in what follows will be on Kant's account of how thought about sensible objects is possible, and not on his argument for showing that the categories apply to appearances. On my view, the first part of the B-Deduction does not yet show that the categories apply to appearances, but instead explains the origin of a certain kind of representation that makes the categories as the most general concepts of an object possible in the first place, and therefore makes it possible to think the objects given in intuition.[114] I will argue that transcendental apperception plays an indispensable role in the context of this argument and that the way in which it makes thought possible is through a special kind of unity that functions as a principle for connecting the manifold of sensible representation in an intelligible way.

2.2 Collective Unity as a Condition of Thought

In this section, I start by arguing for two claims: first, that the unity that Kant claims is required for thinking is *collective*, and second, that transcendental apperception is what makes it possible for representations to have this kind of unity. Once we have a better sense of the nature

[114] It is worth noting that on my reading, Kant is making claims about the origins of the categories within the transcendental deduction. This type of reading cuts against a more traditional narrative about the division of labor between the metaphysical and transcendental deductions, according to which Kant's explanation of the origin of the categories is completed within the metaphysical deduction, whereas his concern in the transcendental deduction is purely justificatory. This traditional reading has recently been challenged by Kohl (2018), and my own reading of Kant's project puts me into alignment with Kohl. I cannot fully substantiate this reading of the Deduction here, but as *prima facie* evidence for thinking that Kant has questions of origin in mind within the transcendental deduction, note that when Kant is laying out the principles of the deduction, he asserts, in opposition to Locke's attempted physiological derivation of the categories, that 'an entirely different birth certificate' for the categories must be supplied if we are to secure their objective validity (A86/B119), which suggests that Kant will in part be providing such an account in the transcendental deduction.

of the unity that is required for thinking, we will be in a better position to explain how and why Kant takes transcendental apperception to be an activity that makes this unity possible.

It is unfortunate that Kant does not, to my knowledge, offer a detailed discussion of collective unity in published works; it appears to be a notion with which he assumes his readers will be familiar. Within the first *Critique*, Kant only mentions collective unity once in the Analytic, in a footnote in the A-deduction (A117n); otherwise, mention of it is confined to the Dialectic, where he mentions it once in the A-edition Paralogisms (A353), once in his discussion of the ideal (A583/B611), and once in the Appendix to the Dialectic (A644/B672). Both published and unpublished references to the notion of collective unity make it clear that what distinguishes this type of unity from any other (and, most prominently, from what Kant calls distributive unity — see A644/B672), is the fact that this unity is meant to be *whole* or *complete*. We can get a better grasp of this distinctive feature of collective unity by noting that Kant often contrasts the completeness of a unity that is collective with what he calls an aggregate, and claims that the 'multitude' of a collective unity comprises a *system*.[115]

In published work, Kant often appeals to this contrast between a *systematic* connection of a manifold and an aggregate, in which elements are grouped together arbitrarily.[116] In the first *Critique*, the contrast is an important theme in the Analytic, and again in the Architectonic (A832/B860). In the Analytic, the distinction is especially prominent in Kant's discussion of the

[115] We see this contrast particularly clearly in the following passage from the *Anthropology*. Kant explains that to take the human race together, collectively, as a *whole* species is to regard it not as composed of a mere aggregate of individuals, but rather to conceive of humanity in a systematic way. In taking the human race together as a whole, he says, one is "taking its [human] species as a whole, that is, collectively (*universorum*), not all of the individuals (*singulorum*), where the multitude *does not yield a system but only an aggregate gathered together*" (*APV* 7:328, my emphasis).

[116] It is important to note that, in calling an aggregate 'arbitrary', Kant is not claiming that the manifold is generated utterly at random, without any constraints governing what is and is not included. The point is primarily that an aggregate is not generated in accordance with a specific kind of principle, namely, one that determines the conditions under which the collection is *complete*.

completeness of the table of categories (cf. A65/B90; A67/B92; A81/B107). What is interesting for our purposes, however, is the fact that Kant explains the distinction between a system and an aggregate by appealing to the different ways in which the collections of items that comprise these groupings are generated. Whereas systems are generated in accordance with a 'principle', aggregates are arbitrarily generated, which is to say that they are not generated in accordance with the relevant kind of principle. In the Prolegomena, Kant takes the fundamental distinction between a system and an aggregate to rest on the possibility of deriving the items that belong together in a system from a principle:

> 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* [...] 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 *comprehending* [*ein Begreifen*], and only now does he have *a system*. (*Prol* 4:322, my emphasis)

The principle of a system determines the conditions of its completeness; because a system is generated in accordance with such a principle, we can measure the completeness of the system — whether any integral components are missing or foreign elements are present — by appeal to its governing principle (cf. A646/B674). But aggregates, since they are arbitrarily generated, are not generated in accordance with a principle, which means that we have no comparable means of determining the completeness of an aggregate.

In the *Metaphysik Vigilantius*, Kant claims that the possibility of thinking requires that representations can be taken together collectively in one act of consciousness, and he contrasts this with a situation in which the representations are only an aggregate whose elements are not connected through a *principle*:

> [I]f one assumes that the thinking matter <*material cogitan*> is an aggregate of substances, then the representation attributed to it would also have to contain an aggregate of representations, namely its parts, which would be separated from one another, whereby from one part of this, one part would be conceived, by the other another would be thought, and a unity would not be attained thereby. But in order for it still to be brought to a *whole*, it remains always necessary to assume a unifying subject which again connects all of these parts under itself and with each other. ...: i.e., it connects the representation *to a simple principle* ... (*MV* 29:1025-6, my emphasis)

Again, we see that what makes possible the wholeness or completeness that distinguishes a collective unity from a mere aggregate, is that there is a principle of connection among the elements of a collective unity (in this case the representations which together comprise a thought), which makes it possible to comprehend all of the elements together as one.

Now that we have a clearer sense of what collective unity is, I want to turn to Kant's discussion of collective unity in the Paralogisms chapter, since it is there that he claims it to be a requirement for thinking. While I cannot offer a complete interpretation of Kant's argument in this important section, it is nevertheless worthwhile to get an overview of the issues at stake in the second paralogism, since they are directly relevant to the discussion of apperception and collective unity. What these sections make clear is that pure or transcendental apperception, as the form of thought, makes possible all thoughts, as *collective* unities.

Very briefly, then, Kant wants to reject a claim made by the rationalist psychologist, namely, that "many representations have to be contained in the absolute unity of the thinking subject in order to constitute a thought" (A353). The rationalist argument for this claim proceeds by *reductio*. Assume for *reductio* that a composite substance can think. The rationalist then provides an account of what it would be for a composite substance to act. Since a composite substance is made up of an aggregate of substances, the action of such a substance is distributed among the individual substances that comprise the aggregate. This distribution means that the

action (of the composite) is really an aggregate of the actions of the individual substances that make up the composite.[117] It is, therefore, strictly speaking a mistake to speak of *the* action of a composite substance; the distribution is such that the individual actions do not comprise *one* action, but merely a concurrence of many actions. With this point about the action of a composite secured, we turn to an account of the nature of *thought*. The rationalist points out that thought is constituted by a *collective unity* of representations. The rationalist claims, therefore, that the representations that compose the thought must be *taken together* as a whole in order for them to be comprehended together in thought. The representations must, therefore, be connected together in a single act of comprehension to form one meaningful whole, and for that to be possible they must be represented together by a single subject.

This claim about the nature of thought is clearly on display at the start of the rationalist argument, which Kant reconstructs as follows:

> For suppose that the composite were thinking; then every part of it would be a part of the thought, but the parts would first contain the whole thought *only when taken together*. Now this would be contradictory. (A352, my emphasis)

It is important to be clear about what constitutes the contradiction here. To understand it, one needs to bear in mind the first claim rehearsed above, that the action of a composite is merely a concurrence of many individual actions. In the argument, the supposition is that such a substance thinks. Each substance in the aggregate is meant to contain one of the parts that, taken together, are supposed to constitute the whole thought. The contradiction arises once we insist that these parts must be *taken together*, for that requires another action, which is not an action of any one of the individual substances, but an action that can encompass all of the individual actions. Or, put

[117] Here is Kant: "Every **composite** substance is an aggregate of many, and the action of a composite, or of what inheres in it as such a composite, is an aggregate of many actions or accidents, which is distributed among the multitude of substances" (A351).

another way, thought does not arise through the mere concurrence of all the actions (representations) of the individual substances; something is required that comprehends all the actions (or representations) as one. The argument continues:

> For because the representations that are divided among different beings (e.g., the individual words of a verse) never constitute a whole thought (a verse), the thought can never inhere in a composite as such. (A352)

The problem seems to be that the individual representations only constitute a thought if they are taken together as a whole, but they cannot be taken together as a whole since then the supposition that they are distributed among the different beings (substances) is given up.

What is more important than the exact nature of the argument that the rationalist proposes is just how much of it Kant in fact agrees with. Importantly, he acknowledges that thought is a collective unity; thus, for a set of representations to comprise a thought, they must be taken together as a whole. He says explicitly that 'the unity of a thought consisting of many representations is collective' (A353). Nevertheless, he goes on to resist the conclusion of the rationalist argument that this means that the subject of thought must be a *simple* rather than a composite *substance* (A353), for Kant denies that we can draw any conclusions at all about the nature of the subject itself on the basis of the fact that thought is a collective unity. He agrees that for thought to be possible it must be possible to take together and comprehend the manifold of representations that compose it, but he denies that this requirement can only be satisfied when the substance in which the thought inheres is simple. Instead, he thinks we can only make a claim about the *formal* requirements on thought, and that, he says, is only the claim that "the subjective **I** cannot be divided or distributed, and this **I** [is] presuppose[d] in all thinking" (A354). Now, two claims follow from this: First, the 'subjective I' is not to be identified with the underlying substance in which thoughts inhere. Second, the subjective I is itself a formal condition of thinking.

On my account of transcendental apperception, we have the resources to account for both of these claims. As we have already seen in Section 1, Kant identifies the I not with the substance in which thoughts inhere, but rather with the activity through which thoughts inhere in a substance (about whose nature we are ignorant). What is important for the moment is to see that Kant in fact *does* take this I (the self-activity of the subject) to be a condition of the collective unity of representations in thought. He says that "it should not go unnoticed that the mere representation **I** in relation to all others (*the collective unity of which it makes possible*) is the transcendental consciousness" (A117fn; my italics). He claims, moreover, that "consciousness is the one single thing that makes all representations into thoughts", and that "it grounds all thoughts" (A350).[118] Finally, he says that "it [transcendental apperception] is only the formal condition, namely the logical unity of every thought" (A398). Thus, instead of claiming that the nature of the thinking substance itself is simple, he claims that it is this activity that is simple, and that it has a special kind of *formal* or *qualitative* unity, which is exactly *not* the kind of unity that should be applied to *objects*, but is a kind of unity that Kant claims is "a logical requirement for every cognition" and, therefore, must be understood in a formal rather than in a material sense (B113). In the next section, I explain exactly how it is that this qualitative unity makes the unity of thought possible.

2.3 Qualitative Unity and Combination

In this section, I would like to turn to the opening section of the B-Deduction, and consider how the activity that Kant calls 'pure apperception' is meant to facilitate the production of a kind of unity in the manifold given through sensibility. Here, Kant calls the unity that is generated the

[118]Note that the context of this passage, in which Kant is describing the 'constant logical subject of thinking', it is clear that he is talking about transcendental apperception rather than empirical consciousness.

original synthetic unity of apperception and he says that it is "the condition of all thinking" (B138). I will argue now that pure apperception functions *teleologically* to produce this original combination.

In section 15 of the B-Deduction, Kant explains that the understanding is responsible for generating a specific kind of representation that he calls *combination* [*Verbindung*]. Here it is important to note that Kant takes combination to be *the only* representation that requires the spontaneous self-activity of the subject and which, therefore, cannot be given through sensibility (B131). Kant tells us that combination requires the representation of a particular kind of unity, which is not itself a representation of combination, but is a unity which, by being added to a manifold, *results in* the representation of combination. Moreover, the unity in question cannot be understood in terms of the category of unity, since the category already requires combination. He tells us, therefore, that

> we must seek this unity (as qualitative, §12) someplace higher, namely in that which itself contains the ground of the unity of different concepts in judgments, and of the possibility of the understanding, even in its logical use. (B131)

Now, we have just seen that the 'the unity of different concepts in judgments' is a *collective* unity, and that the subjective I, or transcendental apperception, constitutes the ground of this unity. Thus, qualitative unity, as a unity that grounds the collective unity of concepts in thought, is the unity of apperception. I will show, in what follows, that this qualitative unity must be understood in teleological terms. Specifically, we must understand this unity as generating the connection of a manifold of representations for a *purpose* or *end*. I call a connection established in accordance with such a principle, *teleological connection*.

Before we proceed to a more detailed discussion of qualitative unity, it is useful to briefly note how Kant conceives of purposes. Kant draws numerous fine-grained distinctions between

different kinds of purposiveness, but here it will suffice to focus on Kant's general definition of a purpose or end in § 10 of the *Critique of the Power of Judgment*, where he defines it as follows:

> An end is the object of a concept insofar as the latter is regarded as the cause of the former (the real ground of its possibility); and the causality of a **concept** with regard to its **object** is purposiveness (*forma finalis*). Thus where not merely the cognition of an object but the object itself (its form or its existence) as an effect is thought of as possible only through a concept of the latter, there one thinks of an end. (*CPJ* 5:220)

Now, the paradigmatic case of an object that is caused by a concept thereof is a human artifact: an object is made or produced *intentionally*, and thus is produced in accordance with a concept or a design in the mind of the agent. A common example of something that Kant takes to be a purpose is a work of art, which is made in accordance with a concept in the mind of the artist, which concept governs the construction of the art-object. Note that Kant distinguishes two ways in which an object may be thought as an end: either when its 'existence' or when its 'form' is possible only through a concept thereof. In the present case, I take it that, since the materials themselves are not produced by the artist, the concept is the cause of the *form* rather than *existence* of the object; say, the cause of the particular arrangement of different colored paints on the canvas. I will also assume that the causality of the concept in the production of the object is expressed by the activity of the artist, such that we can call the activity of producing the object in question purposive. Now, although transcendental apperception, unlike the artist in this example, does not produce any representation intentionally, it is still the case that it produces representations in accordance with a certain form, which is the inner principle of the understanding.

My discussion of qualitative unity will focus on the passage in §12 of the Analytic, to which Kant refers back at B131. There, Kant defends the completeness of his table of categories against the charge that there are three concepts — unity, truth and perfection, as expressed in the scholastic principle that 'every being is one, true and good' — which should have the status of

pure concepts of the understanding and should, therefore, be incorporated into his table. By drawing a distinction between the material and the formal use of these concepts, he argues that, whereas the categories have *material* application, that is, apply to *things*, the concepts in question are only legitimately applied formally as *logical requirements* for the cognition of things, and hence do not belong on the table of categories.

In this context, he explains that in its formal use, the concept of unity, which he calls '*qualitative* unity,' or which he also sometimes refers to simply as '*formal* unity' (cf. *MV* 29:989 and *R*5663, 18:322), cannot be ascribed to *things*, since that would be a material use of the concept, but must be understood as pertaining to the *cognition* of things. He describes the unity as follows:

> In every cognition of an object there is, namely, unity of the concept, which one can call qualitative unity insofar as by that only the unity of the comprehension [*Zusammenfassen*] of the manifold of cognition is thought, as, say the unity of the theme in a play, a speech, or a fable. (B114)

From this passage it follows that the unity in question is expressed by any concept through which an object is cognized. The unity must be understood as pertaining to the *form* of the cognition insofar as the manifold in it is taken together or comprehended. As examples of this kind of 'taking together', Kant points to 'the theme of a play, a speech, or a fable'.

These examples suggest two things: First, that the taking together is made possible through a kind of *intelligible* connection of the elements of the manifold; second, that the intelligible connection is established in accordance with a principle (an organizational form). I take the first suggestion to be the result of the fact that the theme of a play or a story constitutes a kind of intelligibility, since it is in virtue of the theme that the 'manifold' of the action of the play, or the words and sentences of the speech or story, *make sense*. Thus, qualitative unity (the taking together of the manifold) should be understood as the result of establishing intelligible connections between the elements of the manifold such that the manifold constitutes the content of a thought (a concept).

It is important to notice that the theme of a play or a story is the *principle of connection* between the elements of the 'manifold' that it contains. It is the theme that explains why the sentences or words or ideas in a story are presented and connected in one way rather than another. In this way, the theme is a principle of organization, since without it there is no connection between the elements, and thus, no play. In the same way, qualitative unity is a principle of connection between the elements of the manifold of sensibility; without such a principle, there is no thought or concept.

As to the second point, reflecting on the examples Kant provides suggests that he is committed to the idea that this intelligible connection must be understood *teleologically*; that is, the connections must be understood as presupposing an idea of a whole. In the *Metaphysik Vigilantius*, where Kant uses the same example of the unity of a play to illustrate the nature of qualitative, or as he calls it there, formal unity, he uses explicitly teleological language in explaining how the connection between the elements of the manifold is established, saying that

> a play has plurality of objects [a manifold], but [the play has] unity in the representation [of this plurality], for without this *it would be connected for no end.* (*MV* 29:989)

In the case of a play (or a speech or fable), the idea that constitutes the principle in accordance with which (and for the sake of which) the manifold of materials is arranged is not difficult to ascertain. As Kant makes clear in the *Critique of the power of Judgment*, it is the theme that is the principle of connection, and the theme is the concept or idea in the mind of the author in accordance with which she creates the play. Once we notice that Kant is talking about works of art, and is understanding the unity of the representation, which constitutes that work, in terms of a kind principle of intelligibility, it is clear that the phenomenon that he is explaining must be understood on the model of a purpose of the sort I described above.

Works of art, like all artifacts, are ends or purposes. They are produced in accordance with a concept (in this case the theme), which concept is also the principle of the unity of the manifold. Kant says that such works are "*comprehended* [*zusammengefasst*] under a concept or an idea that must determine *a priori* everything that is contained in it" (*CPJ* 5:373). Like the theme of a play or a story, the principle that Kant has in mind functions to connect various elements (sentences, words, scenes, etc.) in a meaningful way, or in a way that makes sense, and also in such a way that the resulting product is unified and connected in accordance with an aim (whatever is to be expressed or communicated). Correspondingly, the unity of apperception (qualitative unity) is what connects the elements of a thought in such a way that the resulting representation is intelligible. It is like the theme of a story, which, insofar as we can conceive of that theme as constituting the purpose of the story, functions teleologically to connect the elements in the story in such a way as to express that theme. Transcendental apperception connects the elements of our intuition in such a way as to express the unity of apperception, by giving the manifold of intuition the form of thinking.

Note, too, that the theme, as a principle of the connection of the manifold, cannot really be abstracted from the play, story, song, etc. In one sense, of course, it precedes the work, in the sense that it is the 'purpose' and the reason for it, and in that sense the work is created so as to fulfil that purpose. But in another sense, the theme only really exists within the work as a whole; it exists through the connections that it makes possible. In that sense, too, we might say that the theme is *formal*: it shapes the content of a work for a purpose or end and does not incorporate any content 'of its own', independently of the content of the art work. On my reading, transcendental apperception, as the thematic unity of thought, is formal in the same way that the theme of a work of art is formal. It does not in itself represent anything, but it contributes something essential to

the possibility of a certain kind of representation, namely the necessary connections between the representational elements in a thought that make it a synthetic *unity*.

Let me summarize the argument so far. I have aimed to show that the 'taking together', or comprehension, in which the collective unity of a thought consists is made possible through an intelligible connection between the elements of a manifold of given representation. I subsequently argued that Kant's use of the example of the theme of a play suggests that we should understand this intelligible connection in teleological terms; that is, we should understand it as a connection established for an end or purpose. Not only does Kant explicitly use teleological language in his explanation of the connection of the manifold of a play, it is also suggested by the fact that plays (and speeches and fables) are works of art. Kant takes works of art to be purposes insofar as they are created by ordering materials together in accordance with a concept or an idea of the whole in the mind of the author or artist.

Now, in the case of a play, it is quite easy to point to the principle or the concept in accordance with which the intelligible connection is effected in the manifold, since it is simply the theme as it is conceived by the author of the play. In the case of the principle that would bring about the appropriate kind of connection within the manifold of sensibility in order for it to be thought, the matter is not quite as straightforward as pointing to a concept or an idea in the mind of a particular individual. Instead, Kant tells us that such a principle is to be found "in that which contains the ground … of the possibility of understanding" (B131). This ground is pure apperception; thus, we must understand teleological connection as an achievement of a fundamental spontaneous act of the subject, which Kant calls pure apperception. In the next chapter, I will argue that we should understand pure apperception as an act that brings about its

effect (the intelligible connection of the manifold of intuition) in accordance with the *form* of the whole of cognition, and, as we will see, this form is given by an idea of reason.

We can now turn to the representation that Kant claims is made possible through transcendental apperception. In Section 16 of the B-Deduction, Kant says that the 'I think' is the product of pure apperception. This 'I think' is what Kant also refers to as the *original synthetic unity of apperception* or *original combination*. As we saw above, all combination requires qualitative unity, and so this original combination, which Kant calls the 'I think', consists in the taking together of the a priori manifold of given representations, namely, pure intuition. Since pure apperception is the activity that generates the original synthetic unity, it is responsible for establishing intelligible connections within the manifold of pure given intuition such that it can be thought. Kant's aim in § 16 is, after all, to explain how the representations given "prior to all thinking", intuition, can have "a necessary relation to the I think" (B132). I take it that a representation that has a necessary relation to the 'I think' constitutes a thought. Now, given that Kant conceives of the unity required for thought as qualitative, I am suggesting that the possibility of "bringing the manifold of intuition to the unity of apperception" requires that the manifold is 'taken together' by establishing intelligible connections between the elements of the intuition. I will say more in the next chapter about what those intelligible connections are. For now, I am suggesting that the original synthetic unity of apperception, which Kant identifies with the understanding (B134n), should be understood as the product of an activity that brings about its effect in accordance with an inner principle, which is, as we have seen, the form of thinking.

Conclusion

It might be helpful to summarize the main points of the interpretive framework I have sought to motivate in this chapter. In sharp contrast to intentionalist proposals, which understand transcendental apperception as a kind of second-order awareness, either of the self, its activity, or of the representations produced through its activity, on my formalist proposal, transcendental apperception should not be understood as a state of representational awareness at all. Instead, I have proposed that we view Kant's theory of apperception in terms of his metaphysics of substance and causation. Transcendental apperception, I have argued, is the *act* through which thoughts inhere in a substance. The way in which transcendental apperception grounds the inherence of thoughts, I have argued, is by supplying the formal principle that makes possible the distinctive collective unity that pertains to the manifold of all thought as such. This principle is the qualitative unity of apperception, which is a principle of intelligibility that governs the taking together of the manifold of intuition in such a way that the resultant representation is combined in the manner required for thought.

In our discussion of the qualitative unity that characterizes the act of apperception, I pointed out that to function as an organizational principle, apperception must introduce connections into a manifold in accordance with an idea, which must give the form of the whole to which each of the components of the complex representation belongs. In the last and final chapter of this dissertation, I elaborate on the role that this idea plays in regulating the activity of the understanding in the generation of experience. I will argue that reason provides the understanding with a goal, which is given by reason's idea of the whole of cognition. It is the form of this idea of the whole of cognition that determines the nature of the understanding itself, and on which the formal unity of experience as a systematic representation of nature depends.

CHAPTER 5

Relative Spontaneity and the Role of Reason in Cognition

So far, I have shown that, given Kant's characterization of the understanding as a spontaneous faculty of cognition, we should conceive of its functioning in *teleological* rather than in *mechanistic* terms. I have also argued that we can provide an account of the teleological functioning of the understanding by appealing to pure apperception, or transcendental self-consciousness, as an organizing *principle* that makes possible a certain kind of unity among the representational elements given in sensibility. We saw that this unity is required for the understanding to generate the kind of representation that belongs to the subject as a thinker or as an intellect, and that it is the constitutive end or aim of the understanding to bring it about.

Now, it is important to clarify, as I will do in this chapter, what is involved in our conception of the understanding as *essentially* teleological. In particular, I want this chapter to address two concerns that one might have about the teleological conception of the understanding that I have been developing. First, one might be concerned that this way of explaining the function of the understanding depends upon exactly the kind of normativist framework that I set out to reject in Chapter 2. Second, one might be concerned that this way of talking about the

understanding blurs the distinction that Kant draws between two of the higher faculties of cognition, understanding and reason, since it is typically *reason* and not the understanding that is conceived of as the intrinsically teleological faculty. We will be in a position to allay both of these concerns if we investigate more closely how Kant conceives of the relationship between the understanding and reason.

To explain the way in which I understand the relationship between these faculties, I will draw on an important distinction that Kant develops in his metaphysics lectures, between 'relative' and 'absolute' spontaneity. I argue that *reason* plays an important role in determining the formal unity of experience, which is the principle of organization that is inherent to the understanding and is expressed through the categories of the understanding. As the principle of organization of our sensible representations, this form allows the understanding to generate a systematic representation of nature, which Kant calls experience. While this form is inherent in the understanding, I will argue that it is ultimately determined by reason, which represents an idea of the whole of cognition, through which the goal of the understanding is given and without which the formal unity of the cognition of the understanding would not be possible. We will see that this distinction between the ways in which the two faculties are related to the form of experience allows us to explain the sense in which the spontaneity of the understanding is merely relative, in contrast to the absolute spontaneity that characterizes reason.

The chapter proceeds as follows. In Section 1, I draw on Kant's distinction between relative and absolute spontaneity to show that, while the understanding is not absolutely spontaneous, it is nevertheless essentially teleological – that is, it has an end that is essential to its nature. I show that the standard conception of this distinction is flawed and has generated a false dichotomy between mechanistic and normativist conceptions of the understanding. My account of the understanding

as relatively spontaneous is an alternative to both of these conceptions. The interpretation of relative spontaneity that I develop has the further consequence that the understanding depends upon an external power to provide the end for which it acts, and in Section 2, I argue that this power is reason. Section 2.1 provides a textual basis for this claim: I show that the goal of the understanding is given by an idea of reason; and section 2.2 provides a more systematic argument. I argue that reason generates an idea of the whole of the cognition of the understanding and that without the representation of such a whole, it would not be possible to represent necessary relations between the manifold given in sensibility, for the representation of necessary relations requires an a priori representation of the form of a whole, without which the relations between the parts cannot be determined a priori. I go on to show that empirical cognition, and experience more broadly, depends on the possibility of generating and representing such necessary relations in the manifold.[119] Finally, in Section 3, I consider an objection to my position, which holds that, in granting reason a role, albeit an indirect one, in the generation of experience itself, my view cannot explain the sense in which the legitimate employment of reason's ideas in relation to experience must be merely 'regulative', and not constitutive. I show that this objection rests on a misconstrual of Kant's distinction between regulative and constitutive principles, and I argue for an alternative reading of the distinction on which my view is comfortably positioned to respect the regulative status of reason's role in experience.

[119] It is important to note that the term "cognition" is not univocal in Kant's work. Watkins and Willaschek (2017) distinguish between broad and narrow conceptions of cognition. The broad conception includes characterizations of cognition in generic terms, such as the one provided in the *Stufenleiter* passage of cognition as an 'objective perception' (A320/B377). They note that Kant calls the more specific account of cognition in terms of the contributions of distinct cognitive faculties "cognition in the proper sense" (A78/B103). My focus throughout this chapter will be on the relatively stable narrow conception of cognition that is also the one central to Kant's project in the Critique. I also want to note that the term 'experience' [*Erfahrung*] is not univocal; I distinguish more clearly between particular and universal experience in section 2.2 below.

Section 1: The Spontaneity of the Understanding

In this first section, I clarify the way in which the understanding, as a sub-faculty of the more complex higher faculty of cognition, is spontaneous. The higher faculty of cognition is differentiated into three specific faculties, understanding, the power of judgment, and reason, and my primary aim here will be to give an account of the kind of spontaneity proper to understanding and reason, in particular.[120] I will do so by appealing to Kant's distinction between relative and absolute spontaneity. I provide an interpretation of this distinction that appeals to the teleological structure of spontaneous powers laid out in Chapter 3. On the basis of that structure, we see that both the understanding and reason are essentially and intrinsically purposive faculties, but that only reason and not the understanding can set ends for itself. As I will continue to argue in the next section, it is reason that provides the understanding with its essential end, and in so doing, it plays a complex and important role in the theoretical domain. I start in this section by showing that whereas reason is an absolutely spontaneous faculty, the understanding is only relatively spontaneous and that this means that, although the understanding functions in accordance with an inner principle, it still depends on an external power to provide the goal that 'directs' this inner principle. I also show that the fact that the understanding does not set its own ends does not compromise its status as an essentially teleological faculty.

[120] My interest in this chapter is specifically in investigating the nature of the relationship between reason and understanding (in the narrow sense) within the faculty of cognition as a whole. I do not take up the question of the power of judgment and its relation to reason and understanding respectively. On my reading, the understanding is the faculty for *generating* concepts, but the power of judgment is the faculty responsible for applying them. Kant says that it "is the faculty of subsuming under rules" (A132/B171), and the rules are furnished by the understanding. The exact details are not important to me here, for my main concern in this chapter is not to give a complete account of the relationship between the higher faculties of cognition but only to explain more clearly the way in which the understanding, and not just reason, is teleological.

On Kant's view, the higher cognitive faculty as a whole is spontaneous. This spontaneity is exactly what makes it the *higher* faculty, which is to be distinguished from the *lower* cognitive capacity, sensibility, which Kant claims is passive or receptive and therefore not spontaneous (A51/B75). Kant claims that the foundation of the higher cognitive faculty – of both the understanding and reason – is pure apperception. Thus, the activity that we identified in the previous chapter as the organizing principle of the understanding, is also the activity that makes reason possible:

> The I is the foundation of the capacity for understanding *and reason*, and the entire power of cognition [...] In the little word 'I' one finds… the consciousness of self-activity; for the I is not an external thing. (*AC* 25:10; my emphasis)

Since both of these faculties depend on pure apperception, they can be considered spontaneous because their exercise depends on the self-activity that we have been analyzing in the previous two chapters.

Kant, however, draws an important distinction between *absolute* and *relative* spontaneity, and so a question remains about how to situate each faculty with respect to this further distinction. I will show that whereas reason is purely or absolutely spontaneous, the understanding has only *relative* spontaneity. Now, the claim that the understanding is relatively and not absolutely spontaneous is not new, and there is ongoing debate in the literature about whether this label is correctly applied to the understanding.[121] I will argue, in keeping with the dominant trend in the literature, that the understanding *is* relatively spontaneous, but I propose a new interpretation of the distinction, one that rejects the widely accepted view that relative spontaneity is, as Sellars (1970) claims, 'set in motion' by 'foreign causes' (1970, 20). I develop this interpretation within

[121] Cf. Allison (1990); Ellis (2017); Kitcher (1990, 2011); Kohl (2020); Land (2006); McLear (2020a); Pippin (1987); Sellars (1970); Sgarbi (2012).

the framework of causal activity that I articulated in Chapter 3, by introducing a teleological dimension into our understanding of this distinction. As I will go on to show, the nature of the difference between absolute and relative spontaneity is crucial for a proper understanding of the relationship between understanding and reason.

Kant explains the distinction between absolute and relative spontaneity in the following passage:

> Spontaneity <*spontaneitas*> is either absolute or without qualification <*absoluta vel simpliciter talis*>, or qualified in some respect <*secundum quid talis*>. - Spontaneity in some respect <*spontaneitas secundum quid*> is when something acts spontaneously under a condition… 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. But the spontaneity is not without qualification <*simpliciter talis*> because here the inner principle <*principium*> was determined by an external principle <*principium externum*>. The internal principle <*principium internum*> with the watch is the spring, with the turnspit the weight, but the external principle <*principium externum*> is the artist who determines the internal principle <*principium internum*>. The spontaneity which is without qualification <*spontaneitas simpliciter talis*> is an absolute spontaneity. (*ML1* 28:267)

Allison (1990) reads this passage as supporting the standard conception of relative spontaneity as requiring a 'foreign cause' to be set in motion. He says that Kant here characterizes relative spontaneity as "the kind of spontaneity attributable to a body once set in motion or a machine such as a watch or turnspit. In short, it is the 'spontaneity' of an agent whose internal principle of motion must be activated by an external cause" (1990, 60). On this kind of view, the understanding qualifies as relatively but not absolutely spontaneous because Kant claims that the understanding requires sensibility to be 'set in motion', so to speak. Allison's reading is the standard view of

relative spontaneity; what is called into question in the literature is not whether this account of relative spontaneity is correct but only whether it is aptly applied to the understanding.[122]

But this conception of what makes the understanding a relatively spontaneous power relies on a flawed reading of the above passage. Allison claims that the external principle, or the 'foreign cause', is the cause that sets the mechanism (the internal principle) into motion. In the case of the watch, that would be the winding up of the spring, for that is what *activates* the watch and sets its mechanisms in motion. But Kant doesn't say that. He says that the external principle is *the artist*, the rational intelligence responsible for designing the watch. What this 'artist' determines is not the motions of the mechanism but the *inner principle of the watch itself*. The way in which the inner principle is determined is through the concept or idea of the artist, not through the winding of a spring, and ultimately, the watch, *qua* artifact, has an inner principle (a spring) that makes it suitable for the purpose for which it was designed (to tell time). If we miss this teleological dimension of the characterization of relative spontaneity, we are liable to think that the understanding is spontaneous because it requires sensibility to 'set it in motion', whereas, as I will show, the understanding is relatively spontaneous because its end (which it has constitutively and essentially) is not one that it sets for itself, but is determined for it by an external power.

Kant's characterization of relative spontaneity becomes clearer when we consider the way in which the above passage continues. He is concerned, in this context, to establish whether the soul is relatively spontaneous, in which case, like the watch or the turnspit, there must be some

[122] As a proponent of the view that the understanding *is* relatively spontaneous, Hanna (2009) argues, for example, that "human a priori cognition is only relatively spontaneous because it requires sensory inputs via empirical intuition" (2009, 109). Sgarbi (2012) claims, in a similar vein, that to the extent that the cognition of the understanding requires "the concurrence of receptivity" the spontaneity in question here is "always a relative spontaneity" (2012, 48)

external power or cause that determines its inner constitution and on which it therefore depends. He says that if we consider the soul to be relatively spontaneous,

> it [the soul] is a being derived from another <*ens ab alio*>, [and] then it appears to be quite probable that it is also determined by this cause in all its thoughts and actions, thus has only spontaneity in some respect <*spontaneitatem secundum quid*>; that *it indeed acts freely according to the inner principle, but is determined by a cause*. (*ML1* 28:268; my emphasis)

This passage makes it clear that if the soul is to be considered absolutely spontaneous, the inner principle according to which it acts cannot be determined by another external power; if its inner principle *is* determined by another power, it can still count as spontaneous, but only relatively so. Thus, Kant seems to be suggesting that it is possible for us to conceive of a being whose activity is wholly determined by an inner principle, that is to say, whose changes of state do not require the exercise of an external power, but whose *inner principle* (note: not its state) is nevertheless in some way dependent on an external power. We must, consequently, make room for two ways in which the powers of a substance can depend on an external principle. In the first case, which I have already discussed in chapter 3 in characterizing passivity, the external principle determines the *state* of a substance by affecting the (receptive) power of the substance. In the second case, which is relevant to explaining relative spontaneity, the external principle determines the *nature* or the *essence*, that is to say, the *internal* principle of the powers of a substance. What we need to understand more clearly is, therefore, (1) what it means to say of an inner principle that it is self-determined or independent of any external power, (2) what it means to say that an inner principle is determined by an external power, and (3) how such a determination of an inner principle by an external power does not collapse the distinction between relative spontaneity and passivity.

Let us now return to Kant's characterization of the distinction in the passage from *ML1* above. He explains that absolute spontaneity is 'unconditioned' or 'without qualification' because

the absolutely spontaneous subject does not merely act in accordance with an inner principle, but it is also the case that the inner principle itself is not determined by an external power, and is therefore determined by the subject itself. By contrast, we can call a power relatively spontaneous, or as Kant puts it here, 'spontaneous in some respect' or 'under a condition', in case the subject does not determine the inner principle itself; instead, in this case, the inner principle is determined by an external power, but the substance still acts in accordance with this inner principle, and to that extent the activity is not determined through an external power, hence qualifying as spontaneous.

To understand this distinction, it is crucial that we have a clear conception of what an inner principle is. As we saw in Chapter 3, we must take an activity to be the ground of the change of the state of a substance. We can call a principle of such a change *internal* just in case the exercise of the power, or its actualization, is achieved without reliance on any power outside of itself and is sufficient by itself for the change of the state of the substance. Insofar as passive substances require the exercise of an external power to determine a change in their state, we can call the principle of such changes in the state of a substance *external*. What makes Kant's conception of relative spontaneity particularly tricky is that it initially seems as though these two kinds of principles, internal and external ones, are incompatible: if a change in the state of a substance is determined by an inner principle, that seems simply to rule out that an external principle is required to explain the change. To make sense of this conception, we therefore need to see why Kant takes the two principles to be compatible in relatively spontaneous substances.

Since, as we have seen already, spontaneous beings are best understood in teleological terms, it is helpful in this context to conceive of inner principles as purposive; that is to say, these principles determine changes in the state of a substance for a goal or purpose. Thus, the inner

principle of the watch – the spring – determines changes in the state of the watch (its moving hands) for the purpose of telling the time. In non-living beings like artifacts, the inner principle is whatever mechanism is meant to enable the artifact to perform its function. We saw that in the case of living beings, these internal principles are the representations according to which the being acts. These representations determine the activity of the living being for the sake of its survival and proliferation.

If we conceive of inner principles in these teleological terms, then we can draw the distinction between absolute and relative spontaneity as follows: An absolutely spontaneous being, insofar as it determines its own inner principle, sets its *own* ends, and also acts in accordance with those ends; a relatively spontaneous being, insofar as its inner principle is determined externally, does not *set* its own ends, but nevertheless has or is given an end through an external power, and acts in accordance with its inner principle as it is determined by those ends. Kant can say that a relatively spontaneous being, a watch, for example, acts in accordance with an inner principle, and, therefore, also for an end, but it is not an end that the watch determines for itself. Instead, an external principle, which Kant identifies as the artist, constructs this object in such a way as to act in accordance with an end (namely, telling time). The artist does so by providing the object with an internal principle (a spring), which allows it to achieve that function. Thus, to say that an inner principle is self-determined means that the end in accordance with which the being acts is determined by the being itself, whereas to say that an inner principle is determined by an external power means that the end in accordance with which such a being acts was not determined by the being itself, but given to, or 'implanted' in it, by an external cause (in the case of the watch, by its creator).

Kant conceives of the spontaneity of non-human animals on the model of relative spontaneity:

> With animals...there is an instinct of sensibility whereby they need no reason, but rather which an external being *placed in them* for acting, or for working according to instinct. (*MVM* 28:450, my emphasis)

Like in the case of the watch or the turnspit, an external being 'places' the inner principle according to which a non-human animal acts, an instinct, in the animal. Unlike the watch or the turnspit, however, that external principle cannot be the finite intellect of an artist but is a special act of creation that we cannot adequately grasp. We nevertheless think of such a special act of creation on the model of an intentionally acting intelligent cause (cf. *CPJ* 5:434).

We now have a clear way of distinguishing relative spontaneity from passivity. A being is wholly passive if none of its powers are sufficient by itself to bring about a change in its state. Thus, wholly passive beings (material bodies) do not act in accordance with inner principles at all: as we saw in Chapter 3, every change of a material body requires the exercise of an external power. Thus, there is no change of state of a passive substance that is the result of an inner principle. By contrast, the change of the state of a relatively spontaneous being is the result of an inner principle, even though it is also the case that the inner principle itself was determined by an external power.[123]

It is important to emphasize, though, that while the inner principle of a relatively spontaneous being is determined by an external power, the end for which it acts is nevertheless essential to it. On the model I am proposing, whether a being is relatively or absolutely spontaneous, the inner principle that constitutes its essential nature is one that determines the changes of its state for the sake of an end or purpose. Therefore, that end is part of its essential

[123] As we saw in Chapter 3, material bodies only admit of mechanical explanation – they do not act for ends. In fact, in the third Critique Kant insists that teleology is not appropriate to explain any events in material nature (*CPJ* 5:417).

nature just as the end of telling time is essential to something's being a watch. The only difference between an absolute and relatively spontaneous power is that an absolutely spontaneous power has the ability to set its own end; once the end of the power is set (whether from within the power itself, as with an absolute spontaneity, or externally, as with a relative spontaneity), there is nothing further to distinguish the two kinds of power.

As we will see, Kant takes the understanding (like the watch or the turnspit) to be relatively spontaneous but takes reason to be absolutely spontaneous. My aim in what follows will be twofold. First, I will explain why this conception of relative spontaneity allows me to claim that the understanding is essentially and constitutively teleological while nevertheless avoiding the normativist implications of the cognitive agency view. Secondly, I will go on to trace out the implications of conceiving of the two faculties in these terms for a proper understanding of the relationship between them.

The question of whether something is absolutely spontaneous concerns whether it is the sort of thing that sets ends for itself. I will take it as uncontroversial that Kant conceives of reason in these terms. For my purposes, it is more important to consider the status of the understanding. If, like reason, the understanding can set ends for itself, then it too is an absolutely spontaneous power. But if the understanding cannot set ends for itself, but, as I will show, depends on an external power to determine its inner principle, then the understanding is *relatively* spontaneous, that is, it acts in accordance with an end that it does not set for itself. It is nevertheless *essentially* teleological, because, like the watch or the non-human animal, having an end is constitutive of its nature.

Preliminarily, it is helpful to note that Kant characterizes the human being as "the sole being on earth who has reason, and thus the capacity to set voluntary ends for himself" (*CPJ*

5:431). Kant therefore clearly takes reason to be a faculty to set ends and not merely a faculty to act for an end. The more pressing question is whether Kant thinks that reason is the only such faculty. Certainly, it seems clear that Kant thinks *only* human beings can set ends for themselves and that is because *only* human beings have reason. Moreover, I take it to be the consensus in the literature that Kant takes reason alone to be a capacity to set ends for oneself. It is therefore safe to assume that the understanding does not set ends for itself; but in that case, given that the understanding is after all characterized as a spontaneous and not a receptive faculty, it must have an inner principle for its activity, and since that inner principle must be determined by an external power, we need an account of that external power and the way in which the understanding depends on it.

Now, as I mentioned above, several commentators have noted that Kant is committed to the view that the understanding is relatively spontaneous. According to some, notably Sellars (1970) and Neiman (1994), this entails that the understanding functions automatically and is not responsive to the kinds of normative constraints that the cognitive agency view describes. On this view, the understanding functions, for lack of a better term, mechanistically. According to Sellars, for example,

> the spontaneity of which we are conscious is, though not a *sheer* passivity, nevertheless a passivity in that the inner development is set in motion by a foreign cause and follows a routine. In the awareness of noumenal activities of synthesis we would encounter simply another example of a cause the causality of which is caused. If this were all that the spontaneity of the noumenal self amounted to, then although it would not be part of the phenomenal nature of outer and inner sense, it would be *like* an object in nature and might be called a noumenal mechanism. (1970, 24)

In Chapter 3, we have already seen that Neiman (1994) expresses a similar conception of the understanding as '*routine*, *automatic*, and *mechanical*'. It is worth noting that Sellars is suggesting an analogy here with a mechanism of nature, and he therefore, presumably, does not mean that the

understanding is mechanical in a strict sense. It is less clear whether Neiman's conception is metaphorical because she seems to explicitly to want to rule out that the understanding can act for ends (see, for example, 1994, 62). Regardless of whether or not this claim is metaphorical, it is misleading to conceive of the function of the understanding in these terms if it in fact functions teleologically.

We can diagnose the tendency to talk about the understanding as mechanistic by noting that our starting point is a false dilemma: These commentators seem to assume that the activities of the understanding are *either* subject to certain norms – and therefore something for which we are responsible and over which we have control – *or* they are mechanistic. On their view, relative spontaneity is meant to signify a certain kind of automaticity, which is correctly identified as a lack of the kind of control required for responsibility, but this is then taken to imply that the understanding cannot function purposively. As we have already seen, however, something can be counted as mechanistic just in case the changes of its state are determined by an external principle rather than an internal principle. This certainly implies that its functioning will be automatic in the sense that its state is not the result of free choice, but the fact that it functions automatically does not by itself imply that it is mechanistic in the sense that the changes of its state are determined by an external cause. On the standard view of relative spontaneity, this automaticity is simply identified with mechanism, and as a result, mechanism and normativism appear to be exclusive and exhaustive options in characterizing how a power functions.

I am now in a position to explain why my claim that the understanding is purposive does not rule out that its operations are automatic and not under the control of the agent. Simply put, the fact that the understanding acts for a purpose does not imply that the understanding must be responsive to norms in the way that the cognitive agency view suggests. I take the claim that the

understanding's function is automatic to incorporate two distinct but related claims. The first is that the understanding does not freely choose to act; in that sense, it acts automatically. The second is that a certain psychological component is missing: the understanding acts 'blindly', or is unaware of its activity as being for any purpose. But as we will see, both of these conditions can be satisfied consistently with the understanding being purposive.

First, the fact that the understanding's activity is automatic does not entail that its states must be explained by citing an external cause. It is therefore open to me to insist that however 'routine', 'automatic', necessary and non-voluntary we take the activity of the understanding to be, it is, nevertheless, purposive. To see why, I want to briefly consider an example of an object that Kant takes to be purposive, but whose functioning can clearly be described as non-volitional, necessitated and automatic. As it turns out, those are the very objects that Kant describes as relatively spontaneous:

> This [relative] 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. But the spontaneity is not without qualification <*simpliciter talis*> because here the inner principle <*principium*> was determined by an external principle <*principium externum*>. (*ML1* 28:267)

As we have already seen, Kant continues this passage by pointing out that the external principle that determines the inner principle of, for example, the watch, is the designer or artist that made it. As an artifact, it satisfies Kant's criterion of purposiveness; hence, it cannot be explained in purely mechanistic terms. To understand what the watch is and how it came about we have to appeal to a rational cause. Its continuing operation, however, is in one sense independent of the artist's design: once it is produced it simply operates in accordance with its nature, that is, in accordance with the principles of the design, which principles allow it to achieve its end of telling time, but without any direct interference from the artist. If we know enough about the principles that constitute the

nature of a watch, then we can understand it to be, as it were, automatic. The watch obviously has no choice about how it functions and is not responsive to any norms, and in that sense its operations are necessary (that is, given its nature). The watch, therefore, is purposive, but it is also automatic and its operations are not under its control. Now, if the understanding, too, is relatively spontaneous, it is in that respect like the watch. An explanation of the nature or form of the understanding might require an appeal to an external power, but the changes of its state cannot be explained by appeal to external causes. The activity of the understanding is therefore automatic. Thus, the operation of the understanding and the application of the concepts of the understanding are, we might say, automatic and, in fact, necessitated by its very nature.

As I suggested above, the characterization of the operation of the understanding as automatic also incorporates the idea that its procedures are unreflective and blind; that would be to say that the understanding is not *aware* of its functioning as being end-directed or purposive. On this account, the understanding simply produces concepts without being aware of its own activity as being for the sake of any particular end. In support of this reading, note that Kant claims that the understanding does not act *intentionally* (*CPJ* 5:187). But this, again, does not rule out that the understanding is purposive, since it is a mistake to think that something acts purposively only if it is aware of itself as acting for an end. We see this clearly by noting that very few of the things that Kant characterizes as purposive set ends for themselves or are aware of the ends that they have. Not even all organisms fit this characterization. There are many examples of things that Kant would consider to be purposive that lack this kind of awareness. Some organisms might be aware of the fact that they act for ends – humans certainly are. But many others are not; a bee, for example, is not: we can consider its activity automatic and unreflective in much the same way that

the activity of a watch is.[124] Certainly it has no conception of itself as acting for an end, but from the perspective of the biologist not only is its *activity* purposive, but we cannot begin to understand its *generation* in mechanistic terms, either. On my view, we should conceive of the understanding in analogous terms. From the point of view of transcendental philosophy, both the function and 'generation' of the understanding must be understood in teleological terms. Nevertheless, the understanding *itself* is unaware of its activity as being for a purpose: it does not need to conceive of itself as a faculty for rendering the sensible intelligible, neither does it need to set that end for itself, nor does it have to be aware of itself as acting in accordance with that end.

The perspective afforded by a proper understanding of the distinction between absolute and relative spontaneity thus gives us a way of reconciling automaticity and purposiveness in our account of the understanding, opening an interpretive middle ground between mechanistic and agential accounts. As a *relatively* spontaneous faculty whose end is not self-generated, the understanding neither chooses to act in accordance with its own end nor is conscious of so doing. But as a relatively *spontaneous* faculty whose states are determined by an inner principle, the understanding eludes mechanistic explanation in just the same manner as organisms and works of art.

In this section, I have offered an interpretation of Kant's distinction between relative and absolute spontaneity, according to which a power is relatively rather than absolutely spontaneous just in case it does not determine its own ends. The understanding, as a relatively spontaneous power, must therefore depend on an external power to determine the ends in accordance with which it acts. In the following section, I will argue that we must look to *reason* as the external power that provides a goal for the understanding's activity.

[124] That is not to say that we do not in addition need to say something about what sets natural ends apart from artifacts. For discussion of how to draw this distinction, see Ginsborg (2015, 225 ff.).

Section 2: The Relationship between Reason and Understanding

In Section 1, I argued that the understanding is relatively spontaneous, and that this means that, while the understanding functions in accordance with an inner principle that is determined by an end, this end is not one that the understanding has set for itself. Thus, while the understanding has an end for which it acts, it depends on an external power to provide that end. Accordingly, we must identify the external power that determines the understanding by providing that end. In this section, which is split into two parts, I argue, first, that we must look to *reason* as the power that provides the understanding with a goal, and, second, that this goal is necessary in order for the understanding to generate cognition of objects. Reason provides a unifying goal for the understanding's activity, and without this unifying goal the understanding cannot establish the necessary connections in the manifold of sensibility that are required for experience. My argument in Section 2.1 is primarily textual. A more systematic argument for the claim that the understanding depends on reason, set forth in Section 2.2, appeals to the systematicity of the cognition of the understanding and of the nature of experience, without which, I will show, cognition (in the strict sense) would not be possible for us.

It is worth briefly noting that reason plays distinct roles in Kant's practical and theoretical philosophy. Kant takes reason and its ideas, specifically the idea of freedom, to be *constitutive* with respect to our faculty of desire, but reason does not play a constitutive role in the theoretical domain, that is, with respect to our faculty of cognition. To say that reason does not play a constitutive role is to say that its ideas do not apply (immediately) to the objects within the domain that we are considering – in the case of the faculty of cognition, that is the domain of nature. Kant nevertheless thinks that reason functions in the theoretical domain to form ideas such as the idea

of freedom, and that in this domain it has a *regulative* function. In its regulative function, reason admits of an 'immanent' or 'indigenous' use, in which its ideas are not directed 'straightaway to a supposed object corresponding to them', but are rather directed "only to the use of the understanding in general" (A643/B671).[125] But what is it for the ideas of reason to be 'directed at' the use of the understanding? I will propose that we understand the regulative function of reason in the theoretical domain as providing the resources to explain the status of the understanding as *relatively* spontaneous. My fundamental contention will be that reason is 'directed at' the understanding insofar as its spontaneously generated ideas, in the theoretical domain, provide the end or goal of the inner principle according to which the understanding acts (Section 2.1) and upon which all of our cognition of objects depends (Section 2.2).

2.1 Reason and the Goal of the Understanding

In what follows, I provide some textual justification for the claim that the understanding depends on reason to provide it with a goal in a way that fits with our conception of the understanding as only relatively spontaneous in the sense described in Section 1. As we have seen, a spontaneity is relative to the extent that an activity determines the subject in accordance with an inner principle, but where that inner principle itself depends on an external power for its end. My claim in this section is that reason is the external power that determines the inner principle of the understanding by providing it with an end. I will provide textual support for the following three claims: first, that Kant claims that reason determines the understanding; second, that Kant claims that reason does not determine the states of the understanding, but determines the faculty itself (its

[125] I will discuss the distinction between regulative and constitutive functions of reason in more detail in Section 3 below.

inner principle or nature); and third, that reason determines the understanding by providing it with a goal.

Before assessing the textual basis for these claims, it is important to make explicit how I understand Kant's conception of 'determination' (*Bestimmung*) in this context. While it is not possible to provide a full account of Kant's use of this term in his theoretical philosophy, I think it is important to clarify some importantly distinct conceptions of the term 'determination' (and its cognates) that are present in the text, and to specify the way in which I understand it to function in the context of Kant's metaphysics of causation.[126] First, then, it is important to note that when Kant explicitly defines or explains determination, he does so in semantic or epistemic terms. So, for example, both in the first *Critique* and also in the *Metaphysik Mongrovius* (and other lectures and notes), Kant claims that "[t]o determine means to attribute to a thing one of two contradictorily opposed predicates" (*MM* 29:818), and again that "[d]etermination is a predicate of a thing by which the opposite is excluded" (*MM* 29:820). Now, this characterization is unproblematically applied to the first of two senses of determination that I think is operative in Kant's work, which we might call semantic or epistemic determination. It is by attributing certain predicates to an object that we 'determine' the concept of that object. Kant is very explicit about the fact that this kind of determination is not merely analytic; we also, he claims, determine concepts *synthetically*, by adding or attributing predicates that are not already contained in the concept of the object that we are considering. Whether the determination is analytic or synthetic, however, it remains semantic (or epistemic), for it concerns only the predicates through which the concept is clarified (in the analytic case) or extended (in the synthetic case). This further distinction between analytic

[126] I am not interested in providing an exhaustive list of the different senses of determination in Kant here. For helpful and more comprehensive guides, see Ameriks (2018) and Stang (2016, 36 ff.).

and synthetic determination is not directly relevant to my argument in this chapter, and therefore, despite its importance, I set it aside in what follows.

Now, in addition to this *semantic* conception of determination as the determination of a *concept*, Kant also employs the term in a distinct, *metaphysical* sense, on which what is determined is not a concept but rather the power of a substance or of an *object*. Instead of characterizing determination in terms of the predicates that belong to the concept of an object, this second conception of determination characterizes it in terms of the properties or accidents of an object. Kant sometimes introduces this more metaphysical language explicitly in the context of explaining the concept of determination. He says, for example that "[i]nsofar as a thing is determined positively, accidents <*accidentia*> inhere in it; insofar as it is negatively determined, they do not inhere in it" (*MM* 29:770). Passages such as this suggest that determination is a matter of bringing about a certain way in which something exists, that is, bringing about the inherence of certain accidents in a substance. Now, Kant claims that the existence of an accident is brought about by the acts of a substance: "Substance acts, insofar as it contains not merely the ground of the accidents, but rather also determines the existence of the accidents" (*MM* 29:822), and he claims that "[t]he acting substance <*substantia agens*> determines the power of the substance being acted upon <*substantiae patientis*> in order to produce this accident" (*MM* 29:824). Thus, the metaphysical conception of determination involves the actual *production* of accidents, and accordingly determination in the metaphysical sense means that something is made (through some power) to be a certain way. Below, we will see that when Kant says that reason determines the understanding or its use, he has in mind this metaphysical conception of determination.

In Kant's descriptions of the regulative function of reason, one of the ideas that most frequently reappears is that reason 'determines' the understanding and its use. It is worth reviewing some instances of this idea. First, in the following passage from the *Prolegomena*, Kant says that

> certain principles of reason are put forward that determine the order of nature a priori, or rather *determine the understanding a priori*, which is supposed to search for the laws of this order by means of experience. (*Prol* 4:364; my emphasis)

When Kant says that 'certain principles' of reason 'determine' the understanding here, I want to suggest that the language of 'determination' is not accidental. We must conceive of reason as a power that determines another power, the understanding, in the metaphysical sense that I have just outlined. Moreover, it is clear that it is only if reason determines the understanding through these 'principles' that it is possible for the understanding to function in its empirical use to search for laws of an order of nature. In the *Critique*, we see this same combination of commitments. There, Kant also explicitly claims that reason determines the understanding in accordance with its ideas:

> We call [the higher] faculties understanding and reason; chiefly the latter is distinguished quite properly and preeminently from all empirically conditioned powers, since it considers its objects merely according to ideas and in accordance with them *determines the understanding*, which then makes an empirical use of its own concepts (even the pure ones). (A547/ B575; my emphasis)

There are two important things to note about Kant's characterization of the relationship between reason and understanding in this passage. Again, we should take the claim that reason 'determines' the understanding as conforming to Kant's metaphysical account of the determination of one power by another. Now, since the understanding is not a passive power, it cannot be the case that reason *directly* determines the particular representational states that the understanding generates; rather, and this is the second important thing to note, reason instead determines an inner principle according to which the understanding can generate its representations. That is, it is only the

understanding as so determined that *then* makes an empirical use of its *own* concepts in experience and thereby functions in accordance with its own inner principles.

Now, I argued in Section 1 that there are two ways in which an external power can determine (in this metaphysical sense) the way in which something is (or exists), and this corresponds to the distinction between merely passive and relatively spontaneous substances. I argued that a being is merely passive if all of the ways in which it exists (all of its states) are determined directly by external powers. A being is relatively spontaneous, however, if the way in which it exists (its states) is determined by its own nature or inner principle, but that inner principle itself is determined by an external power. In the latter case, the external power does not directly bring it about that a certain accident is realized or inheres in the substance; instead, the external power brings it about that the power itself behaves or acts in a certain way: it settles how the inner principle of the power of the substance functions to bring about changes in its own states by providing it with an end. Since the understanding is relatively spontaneous, rather than passive, I suggest that it must be the second of these modes of determination Kant has in mind when he holds that reason determines the understanding.

I take it, then, that there are strong textual grounds for the general claim that reason's ideas determine the understanding, as well as my specific interpretation of that claim to mean that reason determines the inner principle of the understanding. What is not yet clear is that this determination takes place through reason providing a *goal* for the understanding.

One piece of evidence that strongly suggests that reason determines the understanding *by* setting its goal is to be found in the second Critique, where Kant says that

> [t]o every faculty of the mind one can attribute an *interest,* that is, a principle that contains the condition under which alone its exercise is promoted. Reason, as the faculty of principles, determines the interest of all the powers of the mind but itself determines its own (*CPr* 5:119-20)

We saw in Chapter 3 that the notion of an 'interest' is a teleological notion directly related to (if not interchangeable with) the concept of a goal, and here Kant is claiming that it is *reason* that 'determines' the interest of all other powers of the mind.[127] Notice, too, that this passage also aligns with our account of the distinction between relative and absolute spontaneity, because Kant claims that reason is unique among faculties of the mind in determining its *own* interest.

In the Appendix to the Dialectic, we also see Kant clearly expressing this view in several places. In his description of the regulative employment of reason, Kant tells us that "reason has as its object only the understanding and its *purposive* application" (A644/B672; my emphasis) and claims that "reason … unites the manifold of concepts through ideas by positing a certain collective unity as *the goal* of the understanding's actions" (ibid.; my emphasis). He goes on to say further that the ideas of reason "have an excellent and indispensably necessary regulative use, namely that of directing the understanding to a certain *goal*" (A644/B672; my emphasis). Now, I am claiming that this goal is necessary for the understanding and, specifically, that it is necessary for the understanding to generate experience, and to establish that claim I will need to make a further systematic argument. Making such an argument is the task of the next section.

2.2 Systematicity and the Unity of Experience

In this section, I will argue that the goal that reason sets for the understanding is given by an idea of reason. Kant introduces this idea in the Appendix to the Dialectic, where he explains

[127] Now, one might worry that this passage forces me to concede that sensibility, too, has an interest, and therefore acts in accordance with a goal. But, to avoid that consequence, I would like to point out that Kant's most common characterization of sensibility is not as a faculty (*Vermögen*), but as a receptivity, and occasionally as capacity (*Fähigkeit*). In keeping with the terminology that he inherits from Baumgarten and others, I therefore think that Kant takes 'faculty' to refer only to the possibility of spontaneous acts (and powers), which suggests that when Kant makes a claim about 'all faculties of the mind' in the passage, he does not include sensibility within the scope of that claim.

this idea and its role in cognition at length. He claims that it is an idea of the whole of cognition, and that this idea makes possible the 'coherent use' of the understanding (A651/B679), and even the understanding itself (A647/B675; A657/B685). Kant says that 'this idea postulates complete unity of the understanding's cognition', and that without such an idea there can be no lawful and necessary relations between appearances (A645/B673). The idea itself functions as a *goal*, or a '*projected*' (A647/B674) collective unity, which reason 'posits' for the 'purposive application' of the understanding, that is, 'as the goal of the understanding's actions' (A644/B672). In this final section, I explain what role this idea plays in making understanding and the ultimate product of the understanding, experience, possible. I will argue that this idea is essential to the understanding because it makes possible the formal unity on which the cognition of the understanding depends.

In explaining the role of reason's idea of a whole of cognition, it is helpful to briefly say something about the aim of reason in the theoretical domain, and, more specifically, about the role of this aim in relation to the systematicity of our cognition. Reason's aim, which is generated by its own nature and which it therefore also sets for itself, is to find complete unity: "The law of reason to seek unity is necessary", Kant claims, "since without it we would have no reason" (A651/B679). In this search for unity, reason tries to complete the series of conditions by identifying the unconditioned, which would provide the ultimate explanatory and metaphysical ground of the series of conditioned objects.[128] In pursuing this aim, reason generates its ideas. Kant discusses several of these ideas at great length in the Dialectic, and these are the traditional ideas or concepts of metaphysics, ideas of God, the world whole, and the soul. Since reason aims for completeness in every domain of inquiry, it also generates an idea of the 'whole of cognition' of the understanding, or, of the whole of possible experience (A645/B673). It is this idea, of what

[128] For especially helpful discussions see Stratmann (2021); Watkins (2018); and Willaschek (2018).

Kant calls a *collective* unity of empirical representations, that it sets as the goal for the understanding (A644/B672).

While there is little agreement in the literature on what role the idea of systematic unity is meant to play in Kant's account of the understanding, there is at least some *prima facie* textual evidence that suggests that Kant takes this idea to make an essential contribution to the understanding's role in generating experience. He says, for example, that the idea that reason sets as a goal for the understanding is 'the **touchstone of truth** for its rules' (A647/B675), and that without it we have 'no coherent use of the understanding, and, lacking that, no sufficient mark of empirical truth' (A651/B679). Moreover, he repeats the claim that we have an understanding in the proper sense only if we presuppose some of the principles derived from the idea of systematic unity (A657/B685) and that without it 'no understanding at all would obtain' (A653/B681).

In addition to these textual grounds, I will make a systematic case for thinking that reason's idea of a systematic whole of cognition plays an essential role in the generation of experience, for I argue below that the empirical cognitions in which our actual experience of the world consists depend on an idea of systematic unity. Empirical cognition, I will argue, requires necessary connections to be established in the given manifold, without which there can be no integration of perceptions into a whole of connected representation, which is what our actual experience requires. As we will see, Kant claims that mere "perceptions by themselves are encountered dispersed and separate in the mind" (A121) and that "[i]f every individual representation were … isolated and separated from [the other], then there would never be cognition, which is a whole of compared and connected representation" (A97). In what follows, I will show that such a whole of connected representation is not possible without an idea of the form of the whole of cognition, which is an idea that reason provides for the understanding.

Now, this account of the role of reason's idea of a whole of cognition cuts against a common conception of the role of reason in the generation of experience. Many commentators hold, against the view developed here, that reason's idea of a whole of cognition does not play an important role in our empirical cognition, and that it is therefore not required for us in order to have empirical cognition or actual experience that the understanding functions in accordance with the goal of bringing about such a whole.[129] They argue instead that the idea of a systematic whole of cognition becomes relevant only when we want to transition from particular cognitions to *scientific* cognition. In what follows, I will argue against this view by motivating the claim that even particular cognitions depend on necessary connections. In fact, I will show that the understanding itself, as a faculty of cognition, and the concepts that the understanding generates from its own nature, depend on this idea, for without it the formal unity of experience itself would not be possible. This formal unity is not something that reason can generate after the understanding applies its concepts, for the concepts themselves must express that unity and thereby generate our cognition of objects.

[129] Paul Guyer gives forceful expression to this view in the following passage:

> Unfortunately, he does not explain how or why systematicity is required in order to have an empirical criterion of truth. On the contrary, most of what he says in the first Critique suggests that the understanding can succeed in subsuming empirical intuitions under empirical concepts without reference to any constraint of systematicity, and that the discovery of systematicity satisfies only an additional interest of reason rather than the fundamental interest of the understanding in the unity of experience itself. If this is so, then the law of reason which not only prescribes the search for systematicity but also postulates its existence in nature, although transcendental in some sense, would not be a necessary condition of the possibility of the unity of experience itself. (Guyer 1990, 28)

See also Buchdahl (1967); Grier (2001); Moore (2020); Pickering (2011); and Rush (2000). For an alternative style of reading much closer to my own, according to which the understanding needs to presuppose the systematicity of nature just in order to generate experience, see especially Anderson (2015), Geiger (2003), and Ginsborg (2017). It is important to emphasize, though, that my proposal is still substantively different from the readings of this second set of authors. See footnote 136 for discussion of some of the important differences.

Let me start by considering in a bit more detail the standard view of systematicity and the role that it plays in our theoretical cognition of the world. I will call this view the *faculty independence* view, since it claims not only that the understanding and reason are *distinct* faculties, but also that they are *independent* faculties in the sense that they each make independent contributions to theoretical cognition. The advocate of this view insists on a strict division of cognitive labor between the two faculties by emphasizing distinctive features of their respective natures. Susan Neiman (1994) makes a strong and compelling case for this view in her monograph, *The Unity of Reason*, and for the purpose of my discussion here I will use her account as a model. I will consider, first, what the division of cognitive labor is, according to this view, and second, I will explain how the view traces the division back to the nature of the two faculties.[130]

On the faculty independence view, understanding provides us with a product – what Neiman calls 'low-level knowledge' or cognition, but which other commentators simply refer to as particular experiences – which it is able to generate without any help from or contribution of reason. Without relying on reason, the understanding simply applies, in a mechanical way, that is, without an end or purpose, a set of a priori concepts, the categories, to the manifold of sensory representation afforded by intuition, and thereby generates particular experiences, that is, discrete representations of particular objects within a causal structure. Reason has no role whatsoever in the application of these concepts to the sensory manifold. Instead, reason's task is considerably more sophisticated: it takes the experience afforded by the understanding and makes it *scientific*. That is to say, it provides a kind of systematicity to the empirical cognitions of the understanding by relating them to each other in a hierarchical structure of genus and species, and it does so by establishing inferential relations between the cognitions generated by the

[130] It is worth noting, though, that one might insist on the strict division of cognitive labor even if one conceives of the distinction between the faculties in a slightly different way.

understanding. Reason, therefore, does not influence what the understanding does while it generates cognitions, and it does not provide the understanding with any kind of goal or end that is necessary for the generation of experience itself. On this account, the understanding is not just a *distinct* faculty; its fundamental operation is also *independent* of anything that reason does.

On Neiman's view, the contribution of the understanding is constrained by the fact that it lacks autonomy or freedom, owing to which it is therefore restricted to a mere 'recording' of how things are. She takes this lack of autonomy to imply that the understanding functions mechanistically.[131] Given this function, the understanding can provide us only with minimal or 'low-level' cognitions that are not systematically related to form an integrated whole of scientific cognition. Reason, however, insofar as it is an autonomous faculty, is not limited to merely recording how things are. It is capable, through its autonomous activity, of systematizing the minimal cognitions that the understanding supplies, and it therefore provides us with *bona fide* scientific cognition of nature.

Now, I have already suggested that this picture of how the understanding works is incorrect – that, in fact, we must conceive of the understanding as functioning teleologically and that we can make sense of such functioning in a way that does not render the understanding's activities voluntary. What I want to focus on in this chapter, though, is the division of cognitive labor that Neiman takes to be implied by the distinct natures of understanding and reason. I will suggest that, given how I have characterized the nature of the understanding above, we need not insist on a strict division of cognitive labor and that, in fact, the low-level empirical cognitions or particular

[131] She says, for example, that "[t]he understanding is incapable of anything other than this meager performance because it lacks autonomy, and its mechanical nature is inseparable from the abstractness of its results" (Neiman 1994, 59).

experiences that the understanding produces are not possible without reason's idea of systematicity.

In recent literature, it has become customary to distinguish carefully between the different stages or 'grades' of representation and cognition within Kant's philosophy of mind.[132] This more fine-grained way of distinguishing between the different representations suggests that we should distinguish not only between intuition (*Anschauung*) and experience (*Erfahrung*), but also between intuition and perception (*Wahrnehmung*), and between perception and experience. But we should be yet more careful and note that 'experience' is not obviously a univocal term. Kant sometimes uses it to denote what we might call *particular* experience, which can in turn be distinguished from what Kant on occasion calls *universal* experience.[133] Whereas universal experience is basically identical with a complete scientific comprehension of nature, particular experience, in contrast, is just particular perceptual episodes to which the concept of an object applies. Now, the commentator who wants to defend the faculty independence view will want to exploit this distinction and recommend that when Kant makes remarks that suggest that experience requires reason, he has in mind universal experience only, and is thus not going beyond the familiar claim that reason is required for scientific theory construction. Still, they will maintain that, although particular experiences qualify as objective representations (unlike perceptions), they are not *systematically* related, and they will accordingly maintain that no contribution by reason is required to generate particular experiences. In what follows, I argue that this strategy is not viable: without systematicity, I argue, no necessary relations between representations, and consequently, no objective representations and thus no particular experiences are possible.

[132] See especially Tolley (2016) and Beizaei (2017).
[133] For discussion of this distinction, see especially Stang (2018) and Stephenson (2015b).

To start toward seeing this point, it would be helpful to understand Kant's distinction between perception and particular experiences. Kant tells us that perceptions consist in the empirical consciousness of an appearance. The appearance is given to us in intuition, and is then taken up into consciousness by an act or a synthesis that he calls the synthesis of apprehension:

> First of all, I remark that by the synthesis of apprehension I understand the composition of the manifold in an empirical intuition, through which perception, i.e., empirical consciousness of it (as appearance), becomes possible. (B160)

It is important to note that Kant takes the connections between the manifold of perception, which are generated in empirical consciousness by the synthesis of apprehension, to be arbitrary and contingent; that is to say, the manifold in it just happens to be presented together in the way it is, but it didn't have to be presented that way (rather than some other way). Thus, perceptions are not initially combined or 'taken together' in consciousness, but they exist unrelated and independently of each other within the mind:

> The first thing that is given to us is appearance, which, if it is combined with consciousness, is called perception … But since appearance contains a manifold, *thus different perceptions by themselves are encountered dispersed and separate in the mind*, a combination of them, which they cannot have in sense itself, is therefore necessary. (A121, my emphasis)

Note that this does not mean that perceptions are not *conscious* representations; Kant's claim is simply that there is no determinate relation between the different *consciousnesses* in which perceptions consist. It does mean, however, that these perceptions are not yet empirical cognitions. What is clear at this point is that a further act of the mind, a *combination* or a 'taking together' of the manifold, is required in order for there to be the right kind of connection between the perceptions such that they cease to be dispersed and disconnected in the mind, and acquire the status of cognition.

Now, one might think that the imagination is capable of generating these kinds of connections through a further act of synthesis: the synthesis of reproduction. Of course, the imagination does generate connections between perceptions, but the imagination by itself is incapable of yielding the right kinds of connection, because the connections that are required to generate cognition must be necessary. In the synthesis of reproduction, however, the imagination generates connection through the laws of association, and as Hume taught, those laws are merely *subjective* and do not express any necessary connections between the objects. Thus, Kant tells us that while the imagination is capable of 'associating' perceptions, which does constitute a *kind* of connection, this connection is nevertheless not of the right kind, for it does not suffice to establish the *necessary* connection between the elements of the sensory manifold to count as objective:

> For even though we have a faculty for associating perceptions, it would still remain entirely undetermined and contingent whether they were also associable; and in case they were not a multitude of perceptions and even an entire sensibility would be possible *in which much empirical consciousness would be encountered in my mind, but separated, and without belonging to one consciousness of myself*, which, however, is impossible. (A122; emphasis mine)

This passage makes clear that all that the imagination by itself is capable of securing is a kind of empirical consciousness of appearances and a contingent association of them. In order to generate necessary connections, the association cannot be merely subjective, and therefore these connections must be made necessary by the object to be represented.

In the *Prolegomena*, Kant explicitly cautions the reader not to confuse the empirical combination of perceptions (which rests on the laws of association) with experience. Experience, he claims, requires an a priori unity of the understanding, and he claims that this unity *precedes* any empirical combination of perception:

> For the reader who is stuck in the long habit of taking experience to be a mere empirical combining of perceptions – and who therefore has never even considered that it extends much further than these reach, that is, that it gives to empirical

> judgments universal validity and to do so requires *a pure unity of the understanding that precedes a priori* – I cannot adduce more here, these being prolegomena, except only to recommend: to heed well this distinction of experience from a mere aggregate of perceptions, and to judge the mode of proof from this standpoint. (4:330; my emphasis)

It is notable that in this passage Kant claims that this empirical combination of perception (in the imagination) can only generate an aggregate of perceptions. As we saw in the previous chapter, Kant typically draws a contrast between an aggregate, in which things are grouped together arbitrarily, and a *systematic* connection of a manifold, in which things are *necessarily* grouped together and determined a priori in accordance with an idea of a whole. At this point, then, we can note that the transition from perception to particular experience requires a transition from merely contingent and arbitrary relations between appearances to necessary ones.

The claim that I want to motivate now is that in order for the relations between the representations generated in sensibility (intuition) and made conscious through apprehension (perception) to become systematic, an idea of the form of the whole of cognition is required, and without such an idea there can be no necessary relations between the representations that are supposed to constitute our experience. But, as I will go on to show, experience constitutively requires such relations.

To understand the role of reason and systematicity in cognition, we need a clearer understanding of what it means to say that elements in a manifold are necessarily rather than contingently or arbitrarily connected. In general, we take something to be necessary if it couldn't be otherwise, or if it must be the way that it is. By contrast, something is contingent if it could be otherwise or if it could be in some way different from the way that it is in fact. In the context of connections between elements in a manifold of representations, we can call these connections necessary just in case the elements couldn't be arranged (or represented) any other way; that is, in

case they must have the particular form that they do. If connections between elements are contingent, that means, by contrast, that they only happen to be arranged (or represented) one way but that they could be arranged some other way.

In the latter case, in which the connections between representations are contingent, Kant calls the manifold represented together an *aggregate*, but in the former case, he calls it a *system*. Kant gives the following explanation of what he calls a system:

> I understand by a system, however, 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 one another is determined a priori. … The unity of the end, to which all parts are related and in the idea of which they are related to each other, allows the absence of any part to be noticed in our knowledge of the rest, and there can be no contingent addition … that does not have its boundaries determined *a priori*. (A832/B860)

In this same passage, he claims that reason is the faculty responsible for generating the ideas on which such systems depend. Now, we see that by an *idea*, here, Kant means 'a concept of the form of a whole'. Kant also claims that this idea "contains the end and the form of the whole that is congruent with it' (ibid.). This idea determines i) the domain of a manifold, ii) the position of the parts in relation to each other, and Kant further claims iii) that the position of the parts is determined *a priori* (and, we can infer, is therefore *necessary*), and iv) that the absence of any part can be noticed (again, that each part is necessary and essential given its position in the whole); finally, Kant claims v) that there can be no contingent addition to the cognition of objects in the domain determined by the idea. A system, then, requires an idea that provides its a priori structure (the form of the whole), because without such an idea there would be no necessary connections in the manifold (the relations of the parts would not be fixed and necessary) and no *a priori* boundary that determines the domain and in accordance with which the manifold representations can be regarded as complete. This kind of representation has formal or qualitative unity, because it is the

representation of a whole in which the connections between the parts are necessary. And as we have seen, it is only if the manifold is ordered in accordance with an idea of the form of the whole that these necessary connections are possible.

Now, the goal that reason sets for the understanding cannot be generated by the understanding itself, because only reason aims to represent totalities – the complete set of conditions of a given condition or the unconditioned itself.[134] Reason is capable of generating ideas of such totalities, and these ideas are representations of collective unities (of objects or of cognitions). The understanding is less ambitious, however, and it is capable only of representing what Kant calls distributive unities (A643-4/B671-2). Kant contrasts distributive unity with collective unity insofar as only items that constitute a collective unity are a complete or whole set of items and therefore constitute a totality.[135] Distributive unities lack such a wholeness or completeness, because there is no principle that one can appeal to that guarantees that no further items can be added to the set of items, that is, no idea that determines the domain and the relation of the items within the domain a priori.[136]

Now, the goal that reason provides for the understanding is given by an idea of the collective unity of all of the representations that it could possibly generate, thus, an idea of a whole of possible experience or cognition. Kant describes this goal as "a certain unity of which the

[134] Cf. esp. Stratmann (2021).
[135] Cf. also my discussion of collective unity in Chapter 4.
[136] Now, we have seen that perception is characterized by the fact that representations occur in empirical consciousness that are disconnected and unrelated. Kant, in later work, explicitly characterizes the unity that perceptions have as distributive:

> [t]here exists only one experience; and, if one is to speak of experiences, this signifies only *the distributive unity of manifold perception not the collective unity of the object itself in its thoroughgoing determination*. (*OP* 22:549)

In this passage from the *Opus Postumum*, Kant draws a clear contrast between perception and experience in terms of the kinds of unities that they comprise.

understanding has no concept", which, he says, "comprehend[s] all of the actions of the understanding … into a complete whole" (A326/B383); and he says that the ideas of reason more broadly "determine the use of the understanding according to principles in the whole of an entire experience" (ibid.).

It is important to note that the idea of reason that functions as a goal for the understanding has both a matter and a form. The matter of the whole of possible experience cannot be given a priori, since it depends on the affection of sensibility. As a result, the idea that we have of this whole is formal rather than material. That means that the significance of this idea for the understanding and for the experience that the understanding generates is also only formal. It is, namely, to provide the formal unity of the representations that the understanding generates, without which there could be no necessary connections represented in the manifold of cognition, and therefore, as Kant says, "no coherent use of the understanding". The material aspect of the idea, which would be a representation of science in its final and completed form, is not relevant for the understanding for the purpose of generating experience, but relevant for reason in its empirical use, which aims to develop our everyday experience into a scientific theory of the world.[137]

It is uncontroversial that reason's idea of systematicity plays a central role in this second enterprise, but to see that it is equally important to the more basic cognitive task of generating

[137] I can't say more about this distinction here. What is important for my purpose is only that the role of this idea in relation to the understanding is formal and not material. This sets my account of the way in which the idea of systematicity is relevant for experience apart from an account like Geiger's (2003). Geiger argues that "to have one empirical concept is to have the regulative idea of a hierarchical system of all empirical concepts. The idea of a systematic whole of empirical concepts is a condition of the very meaningfulness of any empirical concept…" (274). I take it that Anderson's explanation of the necessity of the idea of systematicity is kindred to Geiger's, for he also draws on a holistic account of empirical conceptual content to make his case (see especially Anderson 2015, 356 – 366). I reject this kind of view, because on the view that I develop it is not the meaningfulness of the concepts that depend on the idea of systematicity. The meaningfulness of a concept – its content – depends on something that is given in sensibility. It is rather the necessary relations between the representations given in sensibility and connected in the imagination that requires the idea of the *form* of the whole of cognition, which is represented a priori through the categories.

experience, we need to clarify the relationship between particular and universal experience. In the following passage from the A-Deduction, Kant says that

> [t]here is only **one** experience, in which *all perceptions are represented as in thoroughgoing and lawlike connection*, ... If one speaks of different experiences, they are only so many perceptions insofar as they belong to one and the same universal experience. The thoroughgoing and synthetic unity of perceptions is precisely what constitutes the form of experience, and it is nothing other than the synthetic unity of the appearances in accordance with concepts. (A110, my emphasis)

This passage strongly suggests that an idea of collective unity – which, we have seen, may only be generated by the faculty of reason – is necessary for the understanding to generate experience out of given perceptions. Now, my opponent might point out that Kant is clearly talking about *universal* experience in this passage, and so, even if we grant that strict systematicity is required for the necessary connections that govern the one all-encompassing universal experience, that does not show that systematicity is required for *mere* empirical cognition or the particular experiences that make up our actual experience. To argue that it is in fact so required, I will show that the transition from perception to particular experience can only take place by determining perceptions as standing in necessary relations to each other as parts of a system. That is why Kant says that 'different experiences [particular experiences] are only so many perceptions insofar as they belong to one and the same universal experience.' A perception can be a particular experience if and only if it fits into the form of the whole of experience or cognition.

As I have already argued, the transition from perception to cognition depends on producing a certain kind of formal unity in our representations. It is because of this unity that we can refer our representations to something beyond them, for Kant says that this unity is the reason why we "speak of an object corresponding to and therefore also distinct from the cognition" (A104). Thus,

the concept of an object is exactly a representation of this unity, which Kant tells us expresses a certain kind of necessity:

> We find, however, that our thought of the relation of all cognition to its object carries something of necessity with it, since namely the latter is regarded as that which is opposed to our cognitions being determined at pleasure or arbitrarily rather than being determined *a priori*, since insofar as they are to relate to an object our cognitions must also necessarily agree with each other in relation to it, i.e., they must have that unity that constitutes the concept of an object. (A104-5)

What we see in this passage is that whatever unity "constitutes the concept of an object", it is one that can be determined only in relation to our other representations, for it is only if all of these representations "agree with each other" that any individual one of them can express that unity. Thus, the transition from perception to cognition can take place only by transitioning from the dispersed empirical consciousness in perception to a 'unity of consciousness'. As we have already seen in Chapter 4, this unity of consciousness is achieved by pure apperception, which brings about a certain kind of unity in the dispersed representations, one without which the thought of the object itself would not be possible. Accordingly, Kant says that without the unity that apperception provides,

> it would be possible for a swarm of appearances to fill up our soul *without experience ever being able to arise from it*. But *in that case all relation of cognition to objects would also disappear, since the appearances would lack connection in accordance with universal and necessary laws*, and would thus be intuition without thought, but never cognition, and would therefore be as good as nothing for us. (A111, my emphasis)

This transcendental unity is introduced through the act of apperception and is represented by the understanding, formally, through the categories. As fundamental concepts of an object in general, the categories express the necessary relations in the manifold of representation that constitute the representation of the object, both internally (the mathematical categories) and in relation to other objects (the dynamical categories). But, as I now go on to argue, the categories express this formal

unity only insofar as they reflect the form of the whole of cognition, or the form of universal experience in accordance with an idea of reason. We are in a position to see this point once we appreciate the connection between Kant's remarks, in the Appendix to the Dialectic, about reason's role in systematizing the cognition of the understanding, and his claims, at the outset of the Analytic, about the systematic unity of the cognition of the understanding through the categories.

Kant explicitly claims in the Appendix that reason's idea of a whole of cognition is required for our cognition to transition from being a "contingent aggregate" into "a system interconnected in accordance with necessary laws". I use numbers to distinguish the several important claims Kant makes in the following passage:

> [1] If we survey the cognitions of our understanding in their entire range, then we find that what reason quite uniquely prescribes and seeks to bring about concerning it is the **systematic** in cognition, i.e., its interconnection based on one principle. [2] This unity of reason always presupposes an idea, namely that of the form of the whole of cognition, [3] 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. [4] Accordingly, this idea postulates complete unity of the understanding's cognition, [5] through which this cognition comes to be not merely a contingent aggregate but a system interconnected in accordance with necessary laws. [6] One cannot properly say that this idea is the concept of an object, but only that of the thoroughgoing unity of these concepts, insofar as the idea serves the understanding as a rule. (A645/ B673)

Kant makes a number of claims in this passage, and since all of them are important for my purposes, it is worth separating them out and discussing each of them in turn. The first claim, [1], is one that we are already familiar with: Kant claims that reason's role in cognition is to systematize our representations. The second claim, [2], we also saw explained in Kant's general characterization of a system. It is the claim that systematic unity requires an idea of reason, and Kant here identifies that idea as the form of the whole of cognition. The third claim that Kant makes, [3], which is again in line with his general characterization of a system in the Architectonic,

is that this idea must precede the parts, and he says here that this idea precedes the particular cognitions that are the parts of the whole of cognition, and is required in order to determine the position of each part in relation to the others. Kant then makes a claim about this idea in relation to the understanding. He says, first, [4], that this idea postulates, that is to say, it assumes, the complete unity of the cognition of the understanding, and he then makes the claim that I have been working towards in this section, which is, [5], that this idea is required in order to transition from a contingent aggregate of representations to a system in which representations are connected in accordance with necessary laws. Now, we have already seen that experience requires lawful necessary relations, and so by pointing to the role of reason's idea of a whole of cognition in generating lawful relations between representations, Kant is plausibly regarded as pointing to a role for the idea of reason in generating experience. And the final claim in this passage gives us some insight into exactly what that role is. Kant claims, [6], that reason's idea of a whole of cognition, rather than itself constituting a concept of an object, represents the 'thoroughgoing unity' of the concepts of an object. Now, since Kant defines the categories as "*concepts of an object in general*, by means of which its intuition is regarded as determined" (A95/B128, my emphasis), he should be understood as referring to the categories when he talks of concepts of an object here. Putting these points together, Kant can be read as claiming that reason's idea of a whole of cognition imparts systematic unity to the representations of the understanding, thus facilitating the transition from perception to experience, *by* representing the thoroughgoing unity of the categories. Accordingly, Kant claims here that it is by representing thoroughgoing unity among the concepts of an object that reason serves as a 'rule' for the understanding.

With this passage fresh in our minds, the parallels with Kant's claims, early in the Analytic, concerning the unity of the understanding and the categories are impossible to ignore. Kant claims

there that the categories

> spring pure and unmixed from the understanding, as absolute unity, and must therefore be connected among themselves in accordance with a concept or idea. Such a connection, however, provides a rule by means of which the place of each pure concept of the understanding and the completeness of all of them together can be determined a priori, which would otherwise depend upon whim or chance. (A67/B92)

In this passage, Kant claims, first, that there is an idea that precedes the determinate parts of the a priori cognition of the understanding, which renders the parts of cognition and the relations between them necessary. He also claims that the idea serves as a rule for the purpose of determining the structure of cognition a priori. If we now note that Kant also explicitly claims in the opening section of the Analytic that the cognitions of the understanding "constitute[s] a system and that "the sum total of its cognition" is therefore "grasped and determined under one idea" (A65/B90), and that he identifies this idea as "an idea of the whole of the *a priori* cognition of the understanding" (A64/B89), then we see that all six of the claims that we identified in the previous passage from the Dialectic already appear as fundamental assumptions of Kant's project in the Analytic.

On my view, the categories are formal representations of all of the possible necessary connections that could exist within the given manifold of sensibility. The manifold given in sensibility is the matter of our experience, but the formal connections on which the unity of that experience depends are represented by the pure concepts of the understanding. The representation of this formal structure in turn requires an idea of reason, the idea of a whole of cognition, without which there can be no a priori determination of the relations that need to obtain in a manifold to represent an object. Kant thus claims at the outset of the Analytic that

> the sum total of [the understanding's] cognition will constitute a system *that is to be grasped and determined under one idea*, the completeness and articulation of which system can at the same time yield a touchstone of the correctness and

genuineness of all the pieces of cognition fitting into it." (A64-5/B89-90)

It is by being subject to the categories that it is possible for our sensible representations to become experience, which, as we have seen, is a representation of the necessary connections in the manifold of sensibility. Accordingly, Kant says in the *Prolegomena* that experience is directly subject to the legislation of the understanding, through the categories (and compare A126). But in being so subject, experience is thereby subject to reason, for without an idea of a whole of cognition, as a goal that directs the use of the understanding in generating experience, there would be no formal representation of the systematic and necessary connection of the manifold of sensibility that is required for the possibility of experience in the first place. Thus, Kant claims that experience is *indirectly* subject to reason, for, he says, "a thoroughgoing unity in the use of this understanding, for the sake of a unified possible experience (*in a system*), can belong to the understanding only in relation to reason, hence experience, too, [is] indirectly subject to the legislation of reason …" (*Prol* 4:364).

I conclude that reason plays a much more fundamental role in cognition than is generally acknowledged in Kant scholarship. While it is true that experience is directly subjected to the legislation of the understanding through the categories, I have argued that this very legislative function of the understanding presupposes an idea of the form of the whole of cognition. This idea can be generated only by reason, and without it the representations of the understanding cannot be necessarily related in the way that our cognition of objects requires. The categories represent the necessary relations that must obtain within the manifold of sensibility in order for the understanding to represent an object; but the categories can only represent necessary relations between representations, of the kind that pertain to a system rather than an aggregate, if there is some form of the whole of cognition that is represented a priori and given through an idea of

reason. In providing the understanding with this idea, reason sets a goal for the understanding to bring the sensible manifold to the formal unity that would complete our cognition of nature had our understanding not been limited by our sensible nature.

Section 3: Regulative and Constitutive Uses of Reason

On the view I have developed, the faculty of reason plays an ineliminable role not just in the sophisticated cognitive task of generating a scientific theory but in the much more elementary cognitive task of generating experience. In denying the faculty independence view, I have argued that reason's idea of systematicity determines the inner principle of the understanding, without which the understanding simply would not be the faculty that it is. As we saw in the previous section, in granting reason's idea of systematicity this kind of indirect role in the generation of experience, the view developed in this chapter is in broad agreement with the work of Anderson (2015), Geiger (2003), and Ginsborg (2017), all of whom also claim that the idea of systematicity plays some necessary role in the generation of experience.[138] Now, one worry that has been raised for readings of this kind is that they fail to uphold Kant's distinction between the regulative and constitutive use of reason's ideas.[139] For readers who confine the immanent use of reason to the generation of scientific theories, it is straightforward to explain how the ideas of reason merely regulate the use of the understanding without playing a constitutive role, but if reason also plays a role, albeit an indirect one, in the generation of experience, it can be harder to see on what basis we could claim that its role is nevertheless merely regulative and not constitutive. But it would be

[138] As I emphasized in the previous footnote, though, there remain important differences between my view and the kind of view developed by Anderson and Geiger.
[139] See Moore (2020) for a for a helpful recent articulation of this objection; cf. Guyer (1990).

a disastrous consequence for my reading if it entailed that reason's role in experience is constitutive, and so I would like to close by explaining in a little more detail why I am confident that my reading does not have this implication.

Now, part of what creates the appearance that a view such as mine cannot respect the regulative-constitutive distinction is the widespread impression that only constitutive uses of a principle could be necessary conditions on experience.[140] If *that* account of what makes a use of an idea constitutive is correct, then my account certainly entails that reason's role in experience is constitutive, since I hold that reason is a condition without which there could be no understanding, and, *a fortiori*, no experience as the characteristic representation of understanding. But this exclusive alignment of constitutive uses of ideas and necessary conditions on experience is not at all obligatory. When we look at Kant's explanation of the distinction between regulative and constitutive uses, we see that what settles the regulative or constitutive status of a principle is not whether it plays a role in making experience possible, but rather the nature of its relation to objects: either the principle relates to the object *immediately*, in which case it is constitutive, or it relates to the object *mediately,* via its application to some intermediate cognitive faculty, in which case it is regulative. The contrast Kant draws between uses of ideas is *not* between one that is, and one that is not, necessary for experience to obtain, but between a use in which "one directs them *straightway* [*geradezu*] to a supposed object corresponding to them…[and one that directs them] only to the use of the understanding in general regarding objects with which it has to do…" (A643/B671).

The discussion of Section 2.1 allows us to explain this contrast in more detail, and in terms that are germane to my reading of the spontaneity of the understanding. We saw there that Kant

[140] Moore makes this assumption explicit and endorses it, claiming that a use of an idea is constitutive if and only if it is a necessary condition of experience.

consistently claims that, in its regulative use, reason 'determines' the faculty of understanding, and, drawing on Kant's account of metaphysical determination, together with his claim that the understanding is itself a relatively spontaneous faculty, I argued for a specific interpretation of that claim. It is unclear what it would mean for reason to *semantically* determine the understanding, and, given that understanding is a relatively spontaneous rather than passive power, it cannot be that reason metaphysically determines (i.e., brings about) the particular states of the understanding. Instead, I suggested that we must take Kant to mean that in its regulative use, reason metaphysically determines the faculty of understanding by setting an end for it that determines its inner principle. With this account of the regulative use of ideas in mind, consider the way in which Kant contrasts this kind of use with a further (and illegitimate) *constitutive* use of ideas:

> Thus the transcendental ideas too will presumably have their good and consequently immanent use, even though, if their significance is misunderstood and they are taken for concepts of real things, they can be transcendent in their application and for that very reason deceptive. For in regard to the whole of possible experience, it is not the idea itself but its use that is either extravagant (transcendent) or indigenous (immanent), according to whether one directs them *straightway* (*geradezu*) to a supposed object corresponding to them, or only to the use of the understanding in general regarding objects with which it has to do… (A643/B671)

Whereas in their regulative use, ideas of reason are directed "only to the use of the understanding in general", in their constitutive use, they are directed "straightaway to a supposed object corresponding to them", and, as a result of such a use, they are falsely taken to be "concepts of real things" (A643/B671). Once again, our discussion of the notion of determination is helpful here. In directing the ideas of reason 'straightaway' to a supposed object, I suggest, we are treating the ideas as devices for semantically determining the concept of an object in general, which is why we then misconstrue them as 'concepts of real things'. To sum up, then, the contrast between the two kinds of use is one between i) a use of the ideas of reason to metaphysically determine the faculty

of understanding (regulative employment), and ii) a use of the ideas of reason to semantically determine the concept of an object (constitutive employment).

Once we understand the distinction between the two uses of ideas in this way, nothing obliges us to deny that the regulative use of an idea could be necessary for the generation of experience. Instead, the distinction constrains the kind of account we can give of the *way* in which reason's ideas contribute to the generation of experience. We cannot say that the ideas directly supply the content through which we think objects in experience, for that would be to misconstrue them as 'concepts of real things' and thus to envision an illegitimate constitutive use. But it is entirely consistent with this to hold, as my account does, that reason's idea of systematicity metaphysically determines the faculty of understanding itself by determining its inner principle.

Conclusion

In this chapter, I argued, first, that the understanding is a relatively spontaneous power. Against the standard conception of a relatively spontaneous power as requiring 'activation' from an external source, I argued, in section 1, that a power is relatively spontaneous if the end or goal in accordance with which its inner principle functions is not determined by it, but is set for it by an external power. I showed that this does not compromise the understanding's status an essentially teleological power, because the end for which the understanding acts remains essential to its nature as a faculty for making the sensible manifold intelligible. Since the understanding is relatively spontaneous, however, it does depend on an external power to provide it with an end. In the second section of this chapter, I argued that, and attempted to explain how, it is the faculty of reason that provides the understanding with its end. In Section 2.1, I make a textual case for the claim that reason is the external power that determines the understanding by providing the end for which it

acts. Section 2.2 is then devoted to developing a systematic argument for the claim that the understanding depends on reason for its end, and that the end is given through an idea of reason – the idea of the whole of cognition. In this section, I argued that the transition from perception to particular experience requires the integration of perceptions into a systematic whole of representation, which is also a transition from the 'dispersed consciousness of perception' to the 'unified consciousness' of pure apperception. The formal unity of this consciousness is expressed in experience by the fundamental concepts of an object (the categories), and for these concepts to adequately express such a unity they must be generated in accordance with an idea of the form of the whole of cognition. We thus see that reason's role in the generation of empirical cognition, though indirect, is nevertheless far more fundamental than has traditionally been countenanced by Kant's commentators. As against the faculty independence view, which only grants reason a role in the generation of *scientific* cognition, I have argued that reason's idea of a whole of cognition in fact precedes and makes possible even the particular experiences upon which scientific theories may subsequently be built. Finally, in Section 3, I explained how this alternative to the faculty independence view can nevertheless respect Kant's insistence that the use of reason in experience is merely regulative. Once we get clear on the regulative-constitutive distinction and see that nothing in Kant's way of drawing that distinction entails that only constitutive uses of ideas can be conditions on the possibility of experience, we can explain how a regulative use of reason's idea could nevertheless play an essential role in the generation of experience.

Conclusion

Postscript on Practical Self-Consciousness

In this dissertation, I have argued that we can make sense of Kant's conception of the spontaneity of the understanding and of the role that transcendental apperception plays in Kant's account of how the understanding functions, by recognizing that on Kant's view this cognitive faculty functions teleologically. On this basis, I argued that Kant's conception of transcendental self-consciousness is *not* to be understood in representational terms. Transcendental self-consciousness is not a consciousness of a particular object (the self) or some aspect of the self, and nor is it a consciousness of some particular object. In fact, we shouldn't think that transcendental self-consciousness *as such* is an act that represents anything. Instead, I argued that it is an act of the subject that grounds the possibility of certain kinds of representations (thoughts) in a subject by bringing about a distinctive formal unity in the representations given in sensibility.

In developing this account, I combine the commitments of two views that have long been at loggerheads with each other. These views, which I called the epistemic and psychological readings of Kant's account of spontaneity and apperception, diverge along three lines: Reflexivity, Causality, and Normativity. I have sided with the psychological reading on the question of Reflexivity and Normativity, but I suggested that the epistemic reader is correct to insist that when

Kant characterizes the understanding as spontaneous, he means to indicate that the understanding functions outside of the causal framework that characterizes material nature. The resulting view is admittedly closer to the psychological than to the epistemic reading, but I have shown that such a reading is only plausible if we can respect this consequence of Kant's commitment to the spontaneity of the understanding.

In removing the understanding from the causal framework that governs material nature, it is tempting to think that Kant must conceive of the understanding as 'free' in the way that would be required for us to hold the subject responsible for her thoughts and judgments, but my reading resists this temptation. I have denied that any strong normativist consequences follow merely from the fact that the operations of the understanding are not explicable within the mechanistic causal terms appropriate to material nature. Instead, I have argued that the understanding must be conceived of as operating teleologically even if it does so unintentionally. Perhaps teleology brings with it a *kind* of normative dimension, insofar as we can assess beings that act for an end in terms of how well they achieve those ends, but this kind of normativity is not the kind of normativity that we need in order to make sense of the transcendental subject as an epistemic or cognitive *agent*, for it is entirely consistent with a sub-conscious and unintentional end-directedness.

Throughout the dissertation, my strategy has been to approach Kant's theory of apperception from within the framework of his faculty psychology and its underlying metaphysics of causal powers. In closing, I would like to say something about how this approach can helpfully be extended to Kant's practical philosophy. I want to suggest, in particular, that Kant's conception of transcendental self-consciousness has an analogue in his practical philosophy. Although he only briefly alludes to practical self-consciousness (*CPr* 5:29), I believe that we can make progress towards understanding Kant's account of practical cognition, and the fundamental role of

transcendental freedom in his moral philosophy, by pursuing the parallel that I want to briefly outline here. My aim here is not, of course, to provide a full account of practical cognition – that is a project for future work – but I do want to briefly highlight certain parallels between theoretical and practical self-consciousness, which make me optimistic that it is possible to fruitfully extend the framework of the project that I have developed here to Kant's account of practical reason.

To see these parallels, let me briefly rehearse my account of the role that transcendental apperception plays in making theoretical cognition possible. I argued that pure apperception makes an important contribution to theoretical cognition insofar as it makes thought possible. As we have seen, on Kant's account, there are two conditions on cognition (in the strict sense), which, following Watkins and Willascheck (2017), we can call the thought-condition and the givenness-condition. For our purposes here, the givenness-condition is not important, but let me just note that in a standard case of theoretical cognition, the givenness condition is met if an object is given to us in sensible intuition. My argument in Chapter 4 focused on the thought condition: I argued there that transcendental apperception plays a role in satisfying this second condition on cognition. I showed that it does so by contributing a certain *form* to our representations, which I identified as *logical* form, and I suggested that it is in virtue of having such a form that representations have the kind of unity that is required for them to be thought. I suggested, moreover, that transcendental self-consciousness contributes this form insofar as it functions as an organizing principle that is part of the nature of the thinking subject. Now, I want to claim that Kant makes an analogous claim about the role of practical self-consciousness in practical cognition.

To see the basis for the analogy, it would be useful to briefly discuss practical cognition and the way in which Kant distinguishes it from theoretical cognition. Kant claims that in theoretical cognition one represents *what is*, and in practical cognition one represents *what ought*

to be (A633/B661). The former kind of cognition, Kant claims, "merely determin[es] the object and its concept (which must be given from elsewhere)" (Bx), whereas the latter cognition aims to "mak[e] its object actual" (Bx). Practical cognition, therefore, aims to bring about an effect, which effect is an action (in the everyday sense) that the subject performs. Accordingly, practical cognitions have a kind of causality that theoretical cognitions lack. Kant claims that in its practical use,

> [r]eason is concerned with the determining grounds of the will, which is a faculty either for producing objects corresponding to representations or of determining itself to effect such objects … that is, of determining its causality. (*CPr* 5:15)

Theoretical cognitions do not have this kind of causality because they do not determine the subject to *do* anything; they simply determine an object conceptually, and thereby represent how things are:

> [the human understanding's] entire capacity consists in thinking, i.e., in the action of bringing the synthesis of the manifold that is given to it in intuition from elsewhere to the unity of apperception, which therefore cognizes nothing at all by itself but only combines and orders the material for cognition, the intuition, which must be given to it through the object. (B145, also cf. *CPr* 5:101)

Practical cognitions, by contrast, have a kind of non-natural causality, which we will see shortly Kant identifies as freedom. These cognitions are states of a subject's will, they are volitional, and, as such, they determine how a subject acts to produce objects and effects in the world.

Despite this fundamental distinction between practical and theoretical cognition, there are, I think, important parallels between them. Like theoretical cognition, practical cognition has a matter and a form. We saw that the form of theoretical cognition is just the logical arrangement of given representations that qualify it as a thought. This form, moreover, is determined in accordance with the nature of the subject as a thinking being, and it therefore depends on the fundamental "act of the subject's self-activity" (B130). We also saw that this act is transcendental self-

consciousness. Now, the form of practical cognition is not any kind of logical arrangement of representations, but a certain status that qualifies what is represented as to be done as *law*. Moreover, in the same way that the form of theoretical cognition is determined in accordance with the nature of the subject as a thinking being, the form of practical cognition as the form of *law* or lawfulness is determined in accordance with the nature of the subject as a moral or practically rational being. This form is contributed to practical cognition through a fundamental self-activity of reason that Kant identifies as practical self-consciousness. In both cases, therefore, the nature of the faculty is expressed in a fundamental act of self-consciousness that imparts to the exercises of the faculty a characteristic form. Let me briefly provide some textual motivation for this view.

Kant claims that an appreciation of what we ought to do can be *objective* only if the representation of it has the form of a law, which, he says "must contain the *very same determining ground* of the will in all cases for all rational beings" (*CPr* 5:25). I have noted already that practical cognition has a matter and a form. We call its matter a maxim, which Kant tells us is a subjective principle for action (cf. *CPr* 5:19). To have practical cognition, we must therefore represent our maxims in accordance with this form, but, and this is crucial for Kant, it must be the form of lawfulness itself, and not the matter which is constituted by the object represented, that is the determining ground of the will. He insists that if this maxim is what determined the will, then the principle of action is not a *law* because it is only "the subjective condition of receptivity" (*CPr* 5:21). Consequently, it must be because it is lawful that the representation moves me to action, and not because of the relation of the object to my feeling of pleasure or displeasure. We see, then, that the possibility of practical cognition depends on the possibility of representing what is to be done in accordance with this form of universal law.

It is, of course, reason that is meant to contribute this form to our representations. Indeed, Kant famously claims that it is in our consciousness of this form of universal moral law that reason "announces itself as originally lawgiving" (*CPr* 5:31). What is especially important for our purposes, though, is not simply *that* reason imparts universal law to practical cognitions but more specifically *how* it does so. For Kant claims that the law-giving through which reason contributes the form of law to practical cognitions "is merely the self-consciousness of pure practical reason" (*CPr* 5: 30). The formalist proposal I have developed in this dissertation makes perfect sense of this connection between universal law-giving and the self-consciousness of practical reason. Transcendental apperception, we have seen, is an act of self-activity that imparts a characteristic logical form to theoretical cognitions, thus providing representations with the kind of unity that is required for the object given in intuition to be thought through the understanding. In just the same way, I want to claim now, the self-consciousness of pure practical reason is a spontaneous act that imparts the distinctive lawful form in virtue of which a representation constitutes a practical cognition.

In fact, this act (of law-giving) underlies the faculty of practical reason in the same way that pure apperception, as the act of combining representations, underlies and makes possible the faculty of understanding. Kant says that if there were "no *merely formal* laws of the will sufficient to determine it, then neither could *any higher faculty of desire* be admitted" (*CPr* 5:22). On the view Kant is articulating, the higher faculty of desire that could not be admitted in such a scenario is just practical reason. Thus, he claims that unless we can discover a formal law that is sufficient to determine the will of a subject, reason is not practical (*CPr* 5:25). Kant conceives of the practicality of reason as a causality that constitutes the faculty of a subject to produce objects in accordance with its representations (*CPr* 5:15), and it is only in virtue of reason's spontaneous act

of lawgiving that such a causality exists. It is a causality that is 'given within reason' as its own self-activity, through which the subject can properly be considered an *agent*, or a subject that has a *will* (*G* 4:452).

To crystalize the parallels, we can say this. The understanding is the faculty of thinking, and practical reason is the faculty of willing. In both cases, the status of the faculty as such is underwritten by a spontaneous act of self-consciousness. This act is an inner principle that expresses the nature of the subject. In the case of the understanding, the act functions to impart to its representations the logical form of thinking, thereby constituting it as a faculty of thinking and cognizing. In the case of practical reason or the will, self-consciousness functions to impart to its representations the form of law, thereby constituting it as a practical faculty. My claim is therefore that in the same way that the fundamental spontaneous act of transcendental self-consciousness provides the form of theoretical cognition, the absolutely spontaneous act of reason provides the form of practical cognition, and in both cases this act constitutes the faculty in question as the very faculty that it is.

Now, Kant does not only claim that reason's giving of universal law is identical to the self-consciousness of pure practical reason; he *also* claims that the self-consciousness of pure practical reason... "[is] identical with the positive concept of freedom" (*CPr* 5:30). Thus, Kant also claims that if we can show that reason is practical, we will thereby have proven the objective reality of the concept of freedom. As well as clarifying the parallels between theoretical and practical cognition, the formalist model also helps us understand the connection between the self-consciousness of practical reason and the reality of freedom in the positive sense. Kant identifies the positive concept of freedom with the law-giving act of reason:

> [T]he sole principle of morality consists in independence of all matter of the law (namely, from the desired object) and at the same in the determination of choice

> through the mere form of giving universal law … That independence, however, is freedom in the negative sense, whereas the lawgiving of its own on the part of pure and, as such, practical reason is freedom in the positive sense. Thus, the moral law expresses nothing other than the autonomy of pure practical reason, that is, freedom, and this is itself the formal condition of all maxims, under which alone they can accord with the supreme practical law. (*CPr* 5:33; cf. 5:29)

The lawgiving that Kant claims constitutes freedom in the positive sense is itself a fundamental self-activity of reason. He says that transcendental freedom "is absolute spontaneity, and is self-activity from an *inner principle*" (*ML1* 28:267). Since Kant identifies this law-giving with transcendental freedom, we can say that the activity of reason is itself an act of transcendental freedom. Kant says this activity is

> … an incontestable and indeed an objective principle of causality … a principle in which reason does not call upon something else as the determining ground with respect to its causality but already itself contains this determining ground … and in which it is therefore as pure reason itself practical. (*CPr* 5:105; cf 4:457)

It is this very act that expresses or *gives* the moral law, which is the form of "intellectual causality, that is, freedom" (*CPr* 5:73).

Reason, we know, is a spontaneous faculty, and, as we saw in Chapter 3, that means that it has a sufficient inner principle for determining a change in the state of a substance. In Chapter 5, we reviewed the way in which reason is spontaneous within the domain of theoretical cognition, namely, by providing an end for the operations of the understanding; but if reason is additionally to be a *practical* faculty, it must also be the case that we can discover in it an inner principle that is sufficient, independently of any external conditions, to move the subject to *action*. Now, we have seen that this inner principle is the self-consciousness of pure practical reason, and, in line with the formalist proposal, we argued that this self-consciousness imparts the lawful form to practical cognitions. Kant thus concludes from the status of reason as practical that "there are practical laws" (*CPr* 5:15). Reason is self-legislative insofar as its exercise determines a causally

efficacious faculty, the will, on the basis of which the subject can think herself part of an intelligible order, (*CPr* 5:106) as a *causa noumenon*.

It is thus my hope that the formalist model will be of use beyond Kant's theoretical philosophy, and that it can provide us with resources to gain a deeper understanding not only of how reason functions as a practical faculty, but also of the unity of reason across the theoretical and practical domains, and, in turn, the unity of the critical system itself.

REFERENCES

PRIMARY

Kant, Immanuel. 1991-. *The Cambridge Edition of the Works of Immanuel Kant*. Ed. P. Guyer and A. Wood. Cambridge: Cambridge University Press.

Kant, I. 1902-. *Kants Gesammelte Schriften*. Ed. Königlich Preussische Akademie der Wissenschaften, vols 1–29. Berlin: de Gruyter.

SECONDARY

Allais, Lucy. 2015. *Manifest Reality: Kant's Idealism and his Realism*. Oxford: Oxford University Press.

Allison, Henry. 1990. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Allison, Henry. 1996. *Idealism and Freedom: Essays on Kant's Theoretical and Practical philosophy.* Cambridge: Cambridge University Press.

Allison, Henry. 2004. *Kant's Transcendental Idealism: An Interpretation and Defense.* London: Yale University Press.

Allison, Henry. 2015. *Kant's Transcendental Deduction: An Analytical-Historical Commentary*. Oxford: Oxford University Press.

Alston, William. 1988. "The Deontological Conception of Epistemic Justification". In *Philosophical Perspectives, Vol. 2, Epistemology*: 257 – 299.

Ameriks, Karl. 1978. Kant's Transcendental Deduction as a Regressive Argument. *Kant-Studien* 69 (1-4):273-287.

Ameriks, Karl. 1983. Kant and Guyer on Apperception. *Archiv für Geschichte der Philosophie* 65 (2):174-186.

Ameriks, Karl. 1995. "From Kant to Frank: The Ineliminable Subject" in Ameriks, Karl & Sturma, Dieter (eds.) *The Modern Subject: Conceptions of the Self in Classical German Philosophy*. State University of New York Press.

Ameriks, Karl. 1997. Kant and the self: a retrospective in David Klemm and Zöller (ed.), *Figuring the Self*. Suny Press. 55—72.

Ameriks, Karl. 2000. *Kant's Theory of Mind: New Edition.* Oxford: Oxford University Press.

Ameriks, Karl. 2012. *Kant and the Fate of Autonomy: Problems in the Appropriation of the Critical Philosophy*. Cambridge: Cambridge University Press.

Ameriks, Karl. 2018. "On the many sense of 'Self-Determination'". In *Kant on Freedom and Spontaneity*, edited by Kate Moran. Cambridge: Cambridge University Press.

Aquila, Richard. 1989. *Matter in Mind: A Study of Kant's Transcendental Deduction*. Bloomington: Indiana University Press.

Audi, Robert. 2001. "Doxastic Voluntarism and the Ethics of Belief". In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford: Oxford University Press.

Banham, Gary. 2008. "Apperception and Spontaneity". In *Kant Studies Online*.

Baumgarten, A. 2013. *Metaphysics: A Critical Translation with Kant's Elucidations, Selected Notes, and Related Materials*. Edited and translated by Fugate, Courtney and Hymers, John. Bloomsbury, London.

Beizaei, Banafsheh. 2017. “Kant on Perception, Experience and Judgements Thereof”. In *Kantian Review* 22(3): 347 - 371.

Bell, David. 1987. “The art of judgment” in *Mind*, New Series, Vol.96, No. 382, pp. 221– 244.

Bird, Graham. 2006. *The Revolutionary Kant: A Commentary on the* Critique of Pure Reason. La Salle, Il.: Open Court.

Boyle, Matthew. 2011. “’Making up Your Mind’ and the Activity of Reason”. In *Philosopher’s Imprint* 11.

Boyle, Matthew. 2015. “Die Spontaneitaet des Vesrtsandes Bei Kant Und Einigen Neokantianern” in the *Deutsche Zeitschrift fuer Philosophie*, 63 (4): 35–55.

Boyle, Matthew. 2020. “Kant on Logic and the Laws of the Understanding.” In *The Logical Alien*, edited by Sofia Miguens, Cambridge, MA: Harvard University Press, 117 – 144.

Brandom, Robert. 2002. *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*. Cambridge: Harvard University Press.

Brook, Andrew. 1994. *Kant and the Mind*. Cambridge: Cambridge University Press.

Buchdahl, Gerd. 1967. “The Relation Between ‘Understanding’ and ‘Reason’ in the Architectonic of Kant’s Philosophy”. In *Proceedings of the Aristotelian Society* 67(1): 209 – 226.

Callanan, John. 2021. “The Comparison of Animals”. In Allais, Lucy and Callanan, John (eds.) *Kant and Animals*. Oxford: Oxford University Press.

Carl, Wolfgang. 1997. “Apperception and Spontaneity”. In *International Journal of Philosophical Studies* 5(2): 147 – 163.

Cassam, Quassim. 2010. “Judging, Believing and Thinking”. In *Philosophical Issues*, 20: 80-95.

Chignell, Andrew. 2007a “Kant’s concepts of Justification”. In *Nous* 41(1): 33 – 63.

Chignell, Andrew. 2007b “Belief in Kant”. In *The Philosophical Review* 116(3): 323 – 360.

Cohen, Alix. 2013. “Kant on Doxastic Voluntarism and its implications for Epistemic Responsibility”. In *Kant Yearbook*: 33 – 50.

Cramer, Konrad. 1974. “‘Erlebnis’: Thesen zu Hegels Theorie des Selbstbewußtseins mit Rücksicht auf die Aporien eines Grundbegriffs nachhegelscher Philosophie”. In H. G. Gadamer (ed.), Stuttgarter Hegel-Tage 1970, Beiheft 11, Bonn: Hegel-Studien.

Dörflinger, Berndt. 1995. “The Underlying Teleology of the First Critique” in the *Proceedings of the Eight International Kant Congress*, pp. 813 – 826.

Dyck, Corey. 2011. “Kant’s Transcendental Deduction and the Ghosts of Descartes and Hume”. In *British Journal for the History of Philosophy* 19(3): 473 – 496.

Dyck, Corey. 2016. “Spontaneity *before* the Critical Turn: The Spontaneity of the Mind in Crusius, the Pre-Critical Kant, and Tetens” in the *Journal of the History of Philosophy*, Vol. 54, No. 4, pp. 625 – 648.

Ellis, Addison. 2017. “The Case for Absolute Spontaneity in Kant’s *Critique of Pure Reason*”. In *Con-Textos Kantianos* (6): 138 – 164.

Engstrom, Stephen. 2006. “Understanding and Sensibility”. In *Inquiry* 49(1): 2 – 25. DOI: 10.1080/00201740500497225.

Frank, Manfred. 2004. “Fragments of a History of the Theory of Self-Consciousness from Kant to Kierkegaard”. In *Critical Horizons: A Journal of Philosophy and Social Theory*, Vol. 5, No. 1 (translation of Frank, Manfred. 1991. “Fragmente einer Geschichte der Selbstbewußtseins-Theorie

von Kant bis Sartre", in M. Frank (ed.), *Selbstbewubtseinstheorien von Fichte bis Sartre*. Frankfurt am Main: Suhrkamp.

Frank, Manfred. 1991. *Selbstbewubtsein und Selbsterkenntnis*. Stuttgart: Reclam.

Frank, Manfred. 2007. "Non-objectal Subjectivity" in *Journal of Conscious Studies*, Vol. 14, No. 5-6, pp. 152 – 73.

Fugate, Courney. 2014. *The Teleology of Reason: A Study of the Structure of Kant's Critical Philosophy*. De Gruyter.

Ginsborg, Hannah. 2015. *The Normativity of Nature*. Oxford: Oxford University Press.

Ginsborg, Hannah. 2017. "Why Must We Presuppose the Systematicity of Nature?". In *Kant and the Laws of Nature*. Edited by Massimi, Michela and Breitenbach, A. Cambridge: Cambridge University Press.

Golob, Sacha. 2020. "What Do Animals See? Intentionality, Objects, and Kantian Nonconceptualism". In Allais, Lucy and Callanan, John (eds.) *Kant and Animals*. Oxford: Oxford University Press.

Gomes, Anil. 2010. "Is Kant's Transcendental Deduction of the Categories Fit for Purpose?" In *Kantian Review* 15(2): 118 – 137.

Grier, Michelle. 2001. *Kant's Doctrine of Transcendental Illusion*. Cambridge: Cambridge University Press.

Guyer, Paul. 1987. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.

Guyer, Paul. 1990. "Reason and Reflective Judgment: Kant on the Significance of Systematicity". In *Nous*, 24, 17-43.

Guyer, Paul (2003) “Kant on the Systematicity of Nature: Two Puzzles”. In *History of Philosophy Quarterly*, 20, 277-95.

Hanna, Robert. 2004 (revised 2017). “Kant’s Theory of Judgment”. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. URL = https://plato.stanford.edu/entries/kant-judgment/

Hanna, Robert. 2009. “Freedom, Teleology and Rational Causation”. In *Kant Yearbook* 1(1): 99 - 142.

Hanna, Robert. 2011. “Kant’s Non-conceptualism, Rogue Objects, and The Gap in the B-Deduction. In *International Journal of Philosophical Studies* 19(3): 399 – 415.

Hatfield, Gary. 2003. “What Were Kant’s Aims in the Deduction?”. In *Philosophical Topics* 31(1/2): 165 – 198.

Henrich, Dieter. 1969. “The Proof-Structure of Kant’s Transcendental Deduction” in *Review of Metaphysics* 22 (4): 640-659.

Henrich, Dieter. 1982. "Fichte's Original Insight" in *Contemporary German Philosophy* 1, 15–52 translated by David R. Lachterman (translation of Henrich, Dieter (1966), "*Fichtes ursprüngliche Einsicht*", in: *Subjektivität und Metaphysik. Festschrift für Wolfgang Cramer* edited by D. Henrich und H. Wagner, Frankfurt/M., pp. 188–232).

Hieronymi, Pamela. 2008. “Responsibility for Believing”. In *Synthese* 161(3): 357–73.

Indregard, Jonas. 2014. *Kant’s Doctrine of Self-Affection*. Ph.D. thesis in the Department of Philosophy at the University of Oslo.

Kern, Andrea. 2006. “Spontaneity and Receptivity in Kant’s Theory of Knowledge”. Translated by Daniel Smyth. In *Philosophical Topics* 32(1/2): 145 – 162.

Kitcher, Patricia. 1990. *Kant’s Transcendental Psychology*. Oxford: Oxford University Press.

Kitcher, Patricia. 2011. *Kant's Thinker*. Oxford: Oxford University Press.

Kitcher, Patricia. 2017. "Kant on the Faculty of Apperception". In *British Journal of the History of Philosophy* 25(3): 589 – 616.

Kohl, Markus. 2015. "Kant in Freedom of Empirical Thought". In *Journal of the History of Philosophy* 53(2): 301 – 326.

Kohl, Markus. 2018. "The Metaphysical Deduction and the Shadow of Humean Skepticism. In *Kant-Studien* 109 (3): 367 – 394.

Kohl, Markus. 2020. "Kant on Cognizing Oneself as a Spontaneous Cognizer. In *Canadian Journal of Philosophy* 150 (3): 395 – 412.

Kraus, Katharina and Freitag, Wolfgang. 2020. "An Expressivist Interpretation of Kant's 'I Think'". In *Nous*.

Kraus, Katharina. 2020. *Kant on Self-Knowledge and Self Formation: The Nature of Inner Experience*. Cambridge: Cambridge University Press.

Land, Thomas. 2006. "Kant's Spontaneity Thesis". In *Philosophical Topics* 34(1/2): 189 – 220.

Land, Thomas. 2021. "Epistemic Agency and the Self-Knowledge of Reason: On the Contemporary Relevance of Kant's method of Faculty Analysis". In *Synthese* 198 (Suppl 13): 3137 - 3154.

Leech, Jessica. 2017. "The Normativity of Kant's Logical Laws". In *History of Philosophy Quarterly* 34(4): 351 – 371.

Leibniz, Gotfried. 1951. *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origin of Evil*. Edited by Farrer, Austin and translated by Huggard, E.M. London: Routledge & Kegan Paul Ltd.

Leon, Mark. 2002. "Responsible Believers". In *The Monist* 83(3): 421 – 435.

Longuenesse, Beatrice. 1998. *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the* Critique of Pure Reason. Translated by Wolfe, C. T. Princeton: Princeton University Press.

Longuenesse, Beatrice. 2017. *I, Me, Mine: Back to Kant and Back Again*. Oxford: Oxford University Press.

Lu-Adler, Hauping. 2017. "Kant and the Normativity of Logic". In *European Journal of Philosophy* 25(2): 207 – 230.

MacFarlane, John. 2002. "Frege, Kant, and the Logic in Logicism". In *The Philosophical Review*, 111(1): 25 – 65.

McDowell, John. 1994. *Mind and World*. Cambridge, MA: Harvard University Press.

McDowell, John. 1998 "Having the World in View: Sellars, Kant, and Intentionality". In *Journal of Philosophy*, Vol. 95, No. 9: 431- 492.

McHugh, Conor. 2013. "Epistemic Responsibility and Doxastic Agency". In *Philosophical Issues* 23: 132 – 157.

McLear, Colin. 2011. "Kant on Animal Consciousness". In *Philosophers' Imprint* 11(15): 1 – 16.

McLear, Colin. 2016. "Kant on Perceptual Content". In *Mind* 125 (497): 95-144.

McLear, Colin. 2020a. “On the Transcendental Freedom of the Intellect”. In *Ergo An Open Access Journal of Philosophy* 7(2): 35 – 104.

McLear, Colin. 2020b. “Animals and Objectivity”. In Allais, Lucy and Callanan, John (eds.) *Kant and Animals*. Oxford: Oxford University Press.

Merritt, Melissa. 2009. “Reflection, Enlightenment, and the Significance of Spontaneity in Kant”. In *British Journal of the History of Philosophy* 17(5): 981 – 1010.

Merritt, Melissa. 2015. “Varieties of Reflection in Kant’s Logic”. In *British Journal for the History of Philosophy* 23(3): 478 – 501.

Merritt, Melissa. 2018. *Kant on Reflection and Virtue*. Cambridge: Cambridge University Press.

Moore, Thomas. 2020. *Kant and the Problem of the Regulative*. Unpublished doctoral dissertation, Brown University.

Nagel, Thomas. 1974. “What is it Like to be a Bat?”. In *The Philosophical Review* 83(4): 435 – 450.

Neiman, Susan. 1994. *The Unity of Reason*. Oxford: Oxford University Press.

Newton, Alexandra. 2015. “Kant on the Logical Origin of Concepts”. In *European Journal of Philosophy* 23(3): 456 – 484.

Newton, Alexandra. 2017. “Kant on Animal and Human Pleasure”. In *Canadian Journal of Philosophy* 47(4): 518 – 540.

Newton, Alexandra. 2019. “Kant and the transparency of the mind”. In *Canadian Journal of Philosophy* 49(7): 890 – 915.

Nunez, Tyke. 2021. “Kant on Plants: Self-Activity, Representations, and the Analogy with Life”. In *Philosophers’ Imprint* 21(11).

O’Neill, Onora. 1989. *Constructions of Reason: Explorations of Kant’s Practical Philosophy*. Cambridge: Cambridge University Press.

Oberst, Michael. 2019. “Kant and Crusius on Causal Chains” in *Journal of the History of Philosophy* 57(1): 107 – 128.

Oberst, Michael. Unpublished MS. “Kant on Essence and Nature”.

Pereboom, D (2006) “Kant’s Metaphysical and Transcendental Deductions” in Bird, G (ed) *A Companion to Kant*. Oxford: Blackwell Publishing, pp. 161 – 2.

Pickering, Mark. 2011. “The Idea of the Systematic Unity of Nature as a Transcendental Illusion”. In *Kantian Review* 16(3): 429 – 448.

Pippin, Robert. 1987. “Kant on the Spontaneity of Mind”. In *Canadian Journal of Philosophy* 12(2): 449 – 475.

Pollok, Konstantin. 2017. *Kant’s Theory of Normativity: Exploring the Space of Reasons*. Cambridge: Cambridge University Press.

Pothast, Ulrich. 1971. *Über einige Fragen der Selbstbeziehung*. Frankfurt am Main: Vittorio Klostermann.

Putnam, Hilary. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Rush, Fred. 2000. “Reason and Regulation in Kant”. In *Review of Metaphysics* 53(4): 837 – 862.

Schafer, Karl. 2021. “A System of Rational Faculties: Additive or Transformative?”. In *European Journal of Philosophy* 29(4): 918 - 936.

Schafer, Karl. Forthcoming. “Kant’s conception of Cognition and our Knowledge of Things-in Themselves”. In Schafer, K. and Stang, N. (eds.) *The Sensible and Intelligible Worlds: New Essays on Kant’s Metaphysics and Epistemology*. Oxford: Oxford University Press.

Schulting, Dennis. 2013. *Kant’s Deduction and Apperception: Explaining the Categories*. New York: Palgrave MacMillan.

Schulting, Dennis. 2019. *Kant’s Deduction From Apperception: An Essay on the Transcendental Deduction of the Categories*. Berlin: De Gruyter.

Sellars, Wilfred. 1971. “’This I or He or It (The thing) which thinks…’”. In *Proceedings and Addresses of the American Philosophical Association* 44: 5 – 21.

Setiya, Kieran. 2013. “Epistemic Agency: Some Doubts”. In *Philosophical Issues* 23: 179 – 198.

Sgarbi, Marko. 2012. *Kant on Spontaneity*. London: Bloomsbury Publishing.

Siewert, Charles. 2017. "Consciousness and Intentionality”. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. URL = <https://plato.stanford.edu/archives/spr2017/entries/consciousness-intentionality/>.

Smit, Houston. 2000. “Kant on Marks and the Immediacy of Intuition”. In *The Philosophical Review*, 109(2): 235 – 266.

Smit, Houston. 2009. “Kant on Apriority and the Spontaneity of Cognition” in Newlands, S and Jorgensen, L.M. (eds.) *Metaphysics and the Good: Themes From the Philosophy of Robert Merrihew Adams*. Oxford: Oxford University Press.

Sosa, Ernest. 2013. “Epistemic Agency”. In *The Journal of Philosophy* 110(11): 585 – 605.

Stang, Nicholas. 2016. *Kant's Modal Metaphysics*. Oxford: Oxford University Press.

Stang, Nicholas. 2018. "Hermann Cohen and Kant's Concept of Experience. In Christian Damböck (ed.), *Philosophie und Wissenschaft bei Hermann Cohen*. Cham, Switzerland: Springer.

Stang, Nicholas. 2019. "A Guide to Ground in Kant's Lectures on Metaphysics". In *Kant's Lectures on Metaphysics: A Critical Guide*. Edited by Courtney Fugate. Cambridge: Cambridge University Press.

Stephenson, Andrew. 2015a. "Kant on the Object-Dependence of Intuition and Hallucination". In *Philosophical Quarterly* 65(260): 486 – 508.

Stephenson, Andrew. 2015b. "Kant, the Paradox of Knowability, and the Meaning of 'Experience'". In *Philosophers' Imprint* 15(27): 1 - 19.

Stratmann, Joseph. 2018. "Kant Grounding, and Things in Themselves". In *Philosophers' Imprint* 18(7): 1 - 21.

Stratmann, J. 2021. "The End of Explanation: Kant on the Unconditioned". In *Philosophy and Phenomenological Research*. https://doi.org/10.1111/phpr.12828

Strawson, Peter. 1966. *The Bounds of Sense: An Essay on Kant's* Critique of Pure Reason, London: Methuen.

Tolley, Clinton. 2006. "Kant on the Nature of Logical Laws". In *Philosophical Topics* 34(1/2): 371 – 407.

Tolley, Clinton. 2012. "The Generality of Kant's Transcendental Logic". In *Journal of the History of Philosophy* 50(3): 417 – 446.

Tolley, Clinton. 2013. “The Non-Conceptuality of the content of intuitions: A New Approach”. In *Kantian Review* 19(1): 107 – 136.

Tolley, Clinton. 2016. “Between ‘perception’ and understanding, from Leibniz to Kant”. In *Estudos Kantianos* 4(2): 71 – 98.

Tolley, Clinton. 2020. “Kant on the place of cognition in the progression of our representations. In *Synthese* 197, 3215 - 3244. https://doi.org/10.1007/s11229-017-1625-3.

Tolley, Clinton. 2021. “The Metaphysics of Powers in Kant and Hegel” in Jurati, J (ed.) *Powers: A History*. Oxford: Oxford University Press.

Tracz, Brian. 2020. “Imagination and the Distinction between Image and Intuition”. In *Ergo: An Open Access Journal of Philosophy* 6(38):1087 – 1120.

Valaris, Markos. 2013. “Spontaneity and Cognitive Agency: A Kantian Approach to the Contemporary Debate”. In *Kant Yearbook* 5(1): 107 – 126.

Vitz, Rico. 2008. “Doxastic Voluntarism”. In *Internet Encyclopedia of Philosophy*.

Watkins, Eric and Willaschek, Marcus. 2017. “Kant’s Account of Cognition”. In *Journal of the History of Philosophy* 55(1):83 – 112.

Watkins, Eric and Willaschek, Marcus. 2020. “Kant on Cognition and Knowledge”. In *Synthese* 197: 3195 – 3213.

Watkins, Eric. 2005. *Kant and the Metaphysics of Causality*. New York: Cambridge University Press.

Watkins, Eric. 2018. “Kant on Real Conditions.” In Violetta L. Waibel, Margit Ruffing, & David Wagner (Eds.), *Natur und Freiheit. Akten des XII Internationalen Kant-Kongresses*. Berlin and Boston: De Gruyter.

Watkins, Eric. 2019. “What Real Progress has Metaphysics Made since the Time of Kant? Kant and the Metaphysics of Grounding. In *Synthese* 198(Suppl 13): 3213 - 3229.

Waxman, Wayne. 2015. *Kant’s Anatomy of the Intelligent Mind*. Oxford: Oxford University Press.

Willaschek, Marcus. 2018. *Kant on the Sources of Metaphysics: The Dialectic of Pure Reason*. Cambridge: Cambridge University Press.

Wolff, Michael. 1995. *Die Vollständigkeit der kantischen Urteilstafel*. Frankfurt: Klostermann.

Wunderlich, Falk. 2011. *Kant und die Bewußtseinstheorien des 18. Jahrhunderts*. Berlin: de Gruyter.

Zahavi, Dan. 2007. “The Heidelberg School and the Limits of Reflection. In *Consciousness: From Perception to Reflection in The History of Philosophy* edited by Sara Heinamaa, Vili Latheenmaki, and Pauliina Remes. Dordrecht: Springer.

Zinkin, Melissa. 2006. “The Unity of a Theme: The Subject of Judgments of Taste. In *British Journal of the History of Philosophy* 14(3): 469 – 488.

Zuckert, Rachel. 2007. *Kant on Beauty and Biology: An Interpretation of the* Critique of Judgment. Cambridge: Cambridge University Press.