The Power *of* Reason

Kant's Empirical Study of the Mind

Christopher Patrick Benzenberg
University of Cambridge
Faculty of Philosophy
St Edmund's College

This thesis is submitted for the degree of *Doctor of Philosophy*.
Supervisor: Prof Angela Breitenbach
August 2024

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Abstract

This thesis is about Kant's account of reason. In the *Critique of Pure Reason*, Kant introduces reason as an infinitely demanding faculty that seeks complete explanations for all observable phenomena. This account of reason is essential to Kant's discussion in the Transcendental Dialectic and prompts the primary question of this thesis: how does Kant justify such an infinitely demanding faculty? How does he think we come to know that we have reason, so understood?

Traditionally, Kant scholars have held that we can grasp our mental faculties either through a priori awareness of their unique activities or through transcendental arguments. Both approaches, however, fail with reason, which presents unique metacritical challenges. We can never be aware of reason's unique activities, which are infinite and so never complete, and reason cannot be established via transcendental argument because it is not necessary for the possibility of experience.

So, how can we know that we have reason? This thesis breaks with tradition by arguing that reason gains self-knowledge in empirical psychology, the study of phenomena in inner sense. Reason, according to Kant, seeks to explain all phenomena, including those of inner sense. To explain inner phenomena, reason hypothesises mental faculties and their laws. Our tendency to ask why-questions, Kant argues, is best explained by hypothesising a faculty that demands complete explanations – i.e., reason.

The thesis has five chapters. The first shows that, for Kant, mental faculties are (also) powers of inner sense. The second argues that the normative demands of these faculties are grounded in constitutive principles or laws. The third finds that the constitutive principle of reason requires us to systematise powers of nature, which, as the fourth chapter explains, we do by hypothesising their respective laws. Finally, the fifth chapter suggests that reason hypothesises its own explanation-seeking law.

Table of Contents

Acknowledgements

This thesis owes much to many. Foremost, I extend my heartfelt thanks to my supervisor, Angela Breitenbach. Her thoughtful feedback, eagle-eyed attention to detail, and endless encouragement have pushed me to grow as both a philosopher and a human being. She granted me the liberty to explore this project to its fullest extent. I couldn't have asked for a better supervisor, and I owe her an immense debt of gratitude for all her support.

I am forever blessed to have met Maya Krishnan, who has spent countless hours with me reading and discussing Kant. Her mind has influenced my thinking more than anything else. I feel honoured to have shared my doctoral journey with her. In the same breath, I also express my deepest gratitude to Nick Currie, who joined Maya and me in the early years of our reading group, and who taught me a great deal about Kant's logic.

Another massive thank you goes to Andrew Chignell. I first met Andrew at a workshop in Berlin in 2016 when I was still an undergraduate, and he has been the most wonderful mentor ever since. During the spring term of 2023, he generously hosted me at Princeton. His warm hospitality was matched only by the infinite supply of wise counsel. The inspiring environment he created gave the project of this thesis a much need boost.

I am also grateful to my other hosts. Alix Cohen welcomed me to the University of Notre Dame in early 2024. She has been exceptionally generous with her feedback on my work. I would also like to thank Sven Bernecker, Thomas Grundmann, Paul Silva Jr., and Sofie Møller for kindly hosting me at the University of Cologne during the academic year 2023/24. Most of this thesis was, in fact, written in my office in Cologne.

Moreover, I wish to thank my former supervisors, Ralf Bader and Tobias Rosefeldt. Ralf continued to be a mentor to me, guiding me along the way. He also went beyond the call of duty and continued to provide razor-sharp feedback on my work. Without Tobias, I would not have found my way to Kant. His seminar on the B Deduction opened my eyes to the awe-inspiring

ambition of Kant's system. For this, I am forever grateful.

There are many others who have helped me over the years through gestures big and small. I owe a debt of gratitude to Alex Englert, Alexander Hutterer, Andrew Cooper, Andrew Stephenson, Anil Gomes, Connor Tannas, Davide Dalla Rosa, Des Hogan, Gabriele Gava, Haley Brennan, Hatice Kaya, Ian Proops, Jacub Techert, Jess Leech, Johannes Haag, Johannes Wagner, Juan Garcia-Berdoy Carvajal, Jürgen Stolzenberg, Karin Nisenbaum, Karl Schafer, Katharina Kraus, Lawrence Pasternack (†), Lena Ghareh Baghery, Lewis Wang, Mahdi Ranaee, Marvin Backes, Michael Oberst, Myriam Stihl, Nick Stang, Nina Dmitrieva, Pirachula Chulanon, Rima Hussein, Samantha Matherne, Sen Bhuvanendra, Yahata Sakura, and Yazmeen Martens.

I have presented parts of this thesis at various workshops and conferences, including the International Kant Congress in Oslo, the UKKS Annual Conferences in London and Dublin, the Moral Sciences Club in Cambridge, the PPPR Colloquium in Princeton, and the Brown Bag Seminar in Cologne, along with further workshops in Germany and Switzerland. I am grateful to the audiences for their stimulating discussions and valuable feedback.

Writing this thesis would have also not been possible without the generous funding of the *Studienstiftung*. In particular, I wish to thank Thomas Ludwig, Claudia Mund, and Christine Schade, who have been exceptionally helpful and flexible in all matters financial. I also wish to thank the Faculty of Philosophy at Cambridge for the funding during my final year. The financial support allowed me to focus all energy on my research.

As I reflect on my doctoral journey, I am particularly appreciative of the non-philosophers in my life, who must have regarded my academic pursuits as particularly alien. My parents, offering unconditional support, were with me every step of the way. Likewise, my wife, Yasemin, stood by my side, even during the most arduous of times. Her love and patience gave me the fortitude to complete this project.

Note on Citation

References to the *Critique of Pure Reason* are to the standard A and B pagination of the first and second edition. Quotations are based on Jens Timmermann's German edition (1998, Hamburg: Felix Meiner). All other citations are to the volume and page of *Kants gesammelte Schriften*, edited by the Royal Prussian, later German, Academy of Sciences (1900–, Berlin: Georg Reimer, de Gruyter). Reflections are additionally cited by their R-number. Emphasis is original, unless noted otherwise. All translations are my own.

'[Kants] ächte und letzte Philosophie [...], [kann]
nichts seyn, als [...] geistige Chemie.'

— Arthur Schopenhauer, *Nachlass*

1

Introduction

This thesis is about Kant's theory of reason. In the following, I first clarify the specific sense of 'reason' under consideration and present the primary question of my thesis, referred to as the 'Epistemic Question' (§ 1.1). Next, I map the solution space by examining four potential answers to the Epistemic Question, all of which I find unconvincing (§ 1.2). I then propose my own positive answer, which draws on Kant's theory of empirical psychology (§ 1.3). Subsequently, I introduce four major problems to my answer, which I will address throughout the thesis (§ 1.4). Finally, I provide an overview of the structure of this thesis (§ 1.5).

1.1 The Epistemic Question

Throughout this thesis, I am interested in theoretical reason in the narrow sense as Kant discusses it in the Transcendental Dialectic of the *Critique of Pure Reason*. Unlike practical reason, which is concerned with '*what ought to be the case*', theoretical reason is concerned with '*what is the case*' (A633/B661; see also 9:86). And unlike theoretical reason in the broad sense, which covers the 'entire higher faculty of cognition', including the understanding (A835/B863), theoretical reason in the narrow sense excludes the understanding.[1]

[1] Note that Kant sometimes uses the understanding in a broad sense as well to denote the entire higher faculty of cognition. Indeed, he initially claims that the two stems of human cognition are sensibility and understanding (A15/B29), but later states that reason (as distinct from sensibility) constitutes one of the stems (A835/B863). Thus, taken in their broad senses,

Theoretical reason, so understood, can be defined by its characteristic activity. Whereas the understanding synthesises concepts into judgements, theoretical reason takes these judgements and synthesises them into syllogisms or mediate inferences. Kant accordingly characterises reason as 'the faculty of mediate inference' (A299/B355; see also 2:59).[2] Moreover, insofar as syllogisms derive cognition from principles, Kant also characterises theoretical reason as '*the faculty of principles*', in contrast to the understanding, which is the 'faculty of rules' (A299/B356).

But theoretical reason is not just the faculty of mediate inferences and principles, it is also the faculty that seeks to find 'the *totality of conditions* for a given conditioned' (A322/B379). That's Kant's way of saying that theoretical reason doesn't just draw some inferences here and there, but that it demands complete explanations for all experienced phenomena. On Kant's account, then, theoretical reason turns out to demand totalities of conditions, and in this sense, it is an infinitely demanding faculty, which I will denote with capital-R 'Reason' (except in quotes). It is Reason that interests me.

Importantly, Reason's quest for totalities is not incidental or arbitrary. Quite the contrary. Kant claims that the totality of conditions is the 'final [end], in which all rational efforts must at last unite' (A463/B491). In fact, the fundamental principle of Reason – i.e., the principle that defines Reasons activity, and which I will simply call the 'Principle of Reason' – demands that we seek to find the totality of conditions (A307-8/B364-5). I will say more about the Principle of Reason in Section 3.3, especially how it relates to the 'logical maxim' and the 'supreme principle of pure reason' (ibid.).

This demanding account of Reason does substantial philosophical work for Kant. For one, it underpins his 'Rational Sources Account' (Willaschek 2018: 3) of the errors of speculative metaphysics. Kant argues that we tend to constitutively assert the existence of a soul, world, and God (understood as

theoretical reason and understanding should be identified. See also Gomes (2017: 7) and Willaschek (2018: 22n).

[2] Or as Kant puts it in Book One of the Dialectic: 'Reason [...] is the faculty of inferring, i.e., of judging mediately' (A330/B386). On this point, see also Kitcher (1990: 217), Heßbrüggen-Walter (2004: ch. 7), and Kraus (2020: 174).

total objects) because we mistake Reason's interest in totalities for objective grounds. However, Kant's account of Reason also underpins his positive story about the regulative use of the ideas. It is because we seek complete explanations in both psychology and physics that we can regulatively assume the existence of a soul, the world, and God (A669-704/B697-732).

This brings me to the main question of my thesis, which I call the Epistemic Question: *how can we, according to Kant, come to know that we have Reason?* Another way of asking the same question is: on what basis can Kant assert the existence of a faculty that is governed by the Principle of Reason, that is, a faculty that seeks to find totalities of conditions? In addition to the general importance of metacritical questions to Kant's project, this question matters because, as we have seen, Reason does substantial work for Kant, but also because, as we will shortly see, answers to the Epistemic Question are uniquely hard to come by.

But before I turn to potential answers to the Epistemic Question, let us contrast it with two other questions. First, there is the 'Material Question', which asks: *what are Reason's exact demands?* Or asked differently, what is the exact content of the Principle of Reason? Second, there is the 'Normative Question': *why should we abide by Reason's demands?* That is, why should we adhere to the Principle of Reason? While I will touch on both the Normative Question (Chapter 3), and the Material Question (Chapters 4 and 5), I will do so only insofar as it helps to answer the Epistemic Question, which is my primary interest.

1.2 Exploring Solution Space

I now consider four potential answers to the Epistemic Question: (i) that it is obvious that we have Reason; (ii) that Kant is historically justified in assuming Reason; (iii) that Reason can be proven analytically; or (iv) that it can be established on grounds that are synthetic a priori. I will argue that these answers are unconvincing, in part because Reason poses particular challenges that don't arise for other faculties. I don't claim, however, that

these answers are irredeemably false – indeed, some version of answer (iv) may be made to work –, but only that they are faced with problems serious enough to motivate the search for an alternative.

First answer first. Isn't it obvious that we have Reason? Well, I don't think so. It may be obvious that we have a faculty that allows us to draw mediate inferences; it may also be obvious that we have a faculty that allows us to ask another 'why' for every 'therefore'.[3] But what is not at all obvious is that finite beings like us should have an infinitely demanding faculty that directs us to ask all the 'why' questions for all experiential phenomena. As Schafer notes, Kant's formulation of the Principle of Reason, as demanding totalities, raises 'quite forcefully the question of why Kant takes [...] reason to demand *so much*' (2023: 153).[4]

Perhaps, says the second answer, Kant simply adopts the inventory of mental faculties, including Reason, from his predecessors. But while it is true that many faculties in Kant share the name with those of his predecessors – e.g., sensibility, imagination, understanding, and yes, reason –, this should not be taken to mean that Kant doesn't innovate. Not only does he disagree with both Wolff and Crusius about the number of mental faculties (see Chapter 4), but his account of discursive cognition as requiring both sensibility and understanding is generally thought to mark a clear departure from both his empiricist and rationalist predecessors.[5]

But what about Reason specifically? After all, Kant states in the Introduction to the Dialectic that Reason, in its logical use, 'has already been explained by the logicians' (A299/B355). Although this may be true for reason as a faculty of mediate inference, Kant himself must have considered his account of Reason, as a faculty that *seeks* rather than *asserts* totalities, to be original;

[3] Des Hogan made this point to me in conversation. See also Schafer (2019: 183).

[4] Yet even if it were obvious that we have Reason, that still wouldn't answer the Epistemic Question, because obviousness, unlike self-evidence, doesn't indicate a type of justification. It may be obvious that $2 + 2 = 4$, but we can still ask how we come to know that this is so.

[5] Gomes notes that 'both [Kant's] rationalist and empiricist predecessors would have found [Kant's account of discursive cognition] highly controversial, it would clearly be preferable if we were able to read him as having an argument that would not blatantly beg the question against his predecessors' (2017: 24). See also Heßbrüggen-Walter (2004), Dyck (2014), Faltudo (2014), and Kraus (2020: 24-5).

for therein lies his self-proclaimed departure from the Leibnizo-Wolffian tradition (A497-9/B525-7). And so, by his own lights, Kant cannot simply presume that we have Reason, but must provide his own justification. He even clarifies that the *Critique* is not 'a critique of books or systems, but of the faculty of reason in general' (Axii).

So let us look at the third answer. Could we come to know that we have Reason analytically? The short answer is no. The claim 'that we have Reason' makes an existential claim about a faculty, Reason, that Kant claims we possess. Yet given Kant's own metaontology, no existential claim can ever analytically true: 'every existential proposition [*Existenzialsatz*] is synthetic' (A598/B626).[6] But there is a better version of the third answer. Perhaps we could synthetically establish, by whatever means, the existence of some faculty X, which would then analytically entail the existence of Reason; just as the existence of bachelors analytically entails the existence of unmarried men.[7] Here is a schema for this kind of argument:

Analytic Argument

(P1) We have faculty X. [synthetically true].

(P2) If we have faculty X, then we have Reason. [analytically true[8]].

(C) *Therefore*, we have Reason.

Indeed, Kant can be seen toying with an instance of the Analytic Argument in the Introduction. Specifically, he can be read as suggesting that the faculty of mediate inference (= faculty X), which we can presume as obvious, analytically entails Reason. The idea is as follows: by drawing inferences from the conclusion of a syllogism to the (major) premise, we seek a condition for a given conditioned (see Chapter 4). Now, Kant claims that this condition is 'subjected to the same attempt of reason', which now seeks 'the condition

[6] For an in-depth discussion of Kants metaontology, see Bader (2020).

[7] This argument takes inspiration from Gomes, Moore, and Stephenson, who have entertained, though not endorsed, the idea that it is 'analytic of discursive cognition as such that' we have an understanding governed by the twelve categories (2022: 148-50).

[8] As Gomes, Moore, and Stephenson (2022: 148n) note, Kant distinguishes between concept analysis and faculty analysis (A65/B90). I will leave it undecided which of the two applies to premise P2.

of the condition (by means of a prosyllogism)' and this is done 'as long as it happens [*angeht*]' (A307/B364). Kant seems to think that this shows that we have a faculty that seeks the totality of conditions.[9,10]

But this argument is too bad to be charitably attributed to Kant. The faculty of mediate inference simply doesn't analytically entail Reason; for to say that it does would be to repeat the error of first-order metaphysics, but at the level of its faculty metaphysics. Kant is clear that 'the conditioned analytically relates to *a* condition but not to the' *totality* of conditions, to which it only relates 'synthetically' (A308/B364; my emphasis). On Kant's own terms, then, a faculty that seeks *a* condition for a given conditioned should not analytically entail a faculty that seeks the *totality* of conditions for a given conditioned.

But even if the faculty of mediate inference analytically entails Reason, the argument would still be bad. For now, it would be as unobvious that we have the faculty of mediate inference as it is unobvious that we have Reason. By analogy, if it turned out that the concept <dog> analytically entails the concept <animal that breathes fire>, we shouldn't accept the existence of fire-breathing animals, but we should take a cold hard look at our grounds for assuming the existence of dogs. The Analytic Argument merely sweeps the philosophical dirt from one corner to the other; for either way we need a robust synthetic argument to support the existence of an infinitely demanding faculty.[11]

[9] Kraus hints at the argument, when she writes: 'Reason has a tendency to search for completeness', a tendency which 'is already *implied* in [reason's] basic logical use' of mediate inference (2020: 177). See also Heßbrüggen-Walter (2004: 258) and Willaschek (2018: 28).

[10] This argument might also explain why Kant claims that 'the following sentence [is] clear and undoubtedly certain: that if the conditional is given, just by this a regress in the series of all conditions to the same is given to us as a task [*aufgegeben sei*]', and especially why he adds that '[t]his is sentence is also analytic' (A497-8/B526). See also (A201/B246), (A783/B811), and (4:270). This point is further discussed by Grier (2001: 121-2) and Anderson (2015: 11, 278, 281n).

[11] The analytic argument has a close cousin that suffers from the same disease. One might attempt the following argument: we are humans; by definition, all humans have Reason; therefore, we have Reason. But if the concept <human> really had capital-R Reason baked into it, we shouldn't blindly accept that we have Reason, but rather double-check whether we are humans, in that rarefied sense.

This brings me to the fourth possible answer to the Epistemic Question: we can come to know that we have Reason on synthetic grounds a priori. One version of this answer draws on what Kitcher calls 'apperceptive act-awareness' (2011: 160). The general idea is that apperception makes us aware of the activity constitutive of the higher faculty of cognition. In the *Anthropology*, for example, Kant writes that the higher faculty of cognition 'is characterised [...] by the spontaneity of apperception, that is, of *pure awareness of the activity that constitutes thinking*' (7:140-1; my emphasis).[12] Since Reason is part of the higher faculty of cognition, it would seem that apperception also provides us with a pure awareness of the activity of reasoning, which is how we come to know that we have Reason.

Setting aside general problems with Kant's account of apperceptive act-awareness – specifically, whether such awareness is epistemically robust; and if so, how it integrates with Kant's first-order epistemology[13] –, we still face particular problems regarding Reason. While the synthesis of judgments into syllogisms is unified by apperception, which makes us aware of a faculty of mediate inference, the same point doesn't easily generalise to Reason. Reason, after all, seeks to successively synthesise ever more distant conditions into a totality of conditions. But this activity can never be completed and so, it would seem, cannot be unified by apperception. It is unclear, then, how apperception could ever make us aware that we have Reason.[14]

[12] Heimsoeth calls it a 'immediate consciousness [*Unmittelbarkeitsbewusstsein*]' (1956: 249); Longuenesse 'pure action awareness' (2017: 86); and Schafer 'active self-consciousness' (2019: 182). Other relevant passages are (B277n), (A375), (A546/B574), and (28:268).

[13] The literature on act-awareness can be roughly sorted into two camps, which Brennan (ms) calls 'deflators' and 'inflators'. Deflators claim either that act-awareness is entirely 'empty' (A355) because 'no manifold is given' (B135) and so lacks all content (Kraus 2020; Kraus & Freitag 2022); or they claim that act-awareness, though having content, is pre-conceptual or otherwise falls short of genuine cognition (Brook 1994: 254; Klemme 1996: 216). On these views, act-awareness turns out too weak to do any robust metacritical work. Inflators, on the other hand, insist that act-awareness provides robust epistemic access to our noumenal self and its faculties (A546/B574; R6001, 18:420-1), either in the form of some special intellectual intuition (Waxman 1991: 283) or via non-cognitive knowledge (Pippin 1987: 459, 469, Wuerth 2014). These views typically struggle to reconcile Kant's theory of act-awareness with Kant's doctrine of noumenal ignorance (B153; B156-8; R4674-84, 17:643-73; and Brook 1994: 246; Heßbrüggen-Walter 2004: 16; Gomes 2017: 10; Kraus 2020: 31), or with Kant's first-order epistemology (Heßbrüggen-Walter 2004: 15; though see Watkins & Willaschek 2020).

[14] Patricia Kitcher suggested to me in conversation that we might be implicitly aware of Reason's quest for totalities by being aware that acts of mediate-ascending inference can be

Another version of the fourth answer might say that we can establish Reason by means of transcendental argument. After all, we can come to know cognitive faculties in general because they are necessary conditions for the possibility of experience.[15] But while this strategy may work for sensibility and the understanding, there is near-universal agreement among Kant scholars – Geiger (2003) being the famous exception – that Reason as a totality seeking faculty is not a necessary condition for the possibility of experience. We simply cannot come to know that we have Reason by means of transcendental arguments.

But perhaps we might tweak the structure of transcendental arguments to establish Reason. Not being a necessary for the possibility of experience, Reason could be introduced as necessary to explain the errors of speculative metaphysics. Kant argues throughout the Dialectic that rational psychology, cosmology, and theology tempt us precisely because Reason seeks totalities of conditions. I believe there is a kernel of truth in this argument. The big problem, however, is that Reason evidently is not *necessary* to explain the errors of speculative metaphysics. These errors could just as easily be attributed to a faculty that asserts rather than seeks totalities, or indeed, to any number of cognitive corruptions.

1.3 Empirical Psychology

We thus see that all four answers face serious problems. And while some iteration of the fourth answer may be salvageable, it remains unclear how this would work. How could we overcome the unique metacritical challenges of Reason on a priori grounds alone?[16] These difficulties motivate an alternative, albeit initially unlikely, answer to the Epistemic Question. The 'Main Claim'

iterated. While not unpromising as a strategy, this answer would need to be fleshed out in detail to be convincing.

[15] I group transcendental arguments under the fourth solution because, although some people think that the fact of experience analytically entails its necessary condition, the fact of experience itself is synthetic.

[16] On this point, see also Heßbrüggen-Walter (2015: 2484).

of my thesis is that, *according to Kant, we can come to know that we have Reason empirically by way of empirical psychology*.[17] That is, empirical psychology allows Kant to assert the existence of a faculty governed by the Principle of Reason. (Importantly, I am not claiming that this is the only way how we could come to know that we have Reason.)

The Main Claim uses 'empirical psychology' in Kant's technical sense of the term. Kant defines empirical psychology as a '[*p*]*hysiology* of inner sense' (A347/B405),[18] which means to say that empirical psychology considers all 'appearances of inner sense' (4:337),[19] and tries to explain them with reference to causal powers and their laws. That's why, Kant also defines empirical psychology as the discipline that studies the 'sum total of inner perceptions under natural laws' (7:141). Empirical psychology thus effectively seeks to explain our inner appearances as standing in the causal order of an inner nature.[20]

But if empirical psychology is the physiology of inner sense, what is inner sense for Kant? Roughly speaking, inner sense is Kant's take on introspection or 'self-directed observation' (28:227). Via inner sense, 'the mind [*Gemüth*] intuits itself and its inner state' (A22/B37; see also B156); and this intuition stands in temporal succession, as time is the pure form of inner sense (A34/B50-1).[21] Since all inner states are representations for Kant, inner sense is also defined as 'the sum of all representations' (A177/B220;

[17] With 'know' I mean 'epistemically access'. As we will see in Chapter 6, empirical psychology only ever justifies opinion (*Meinung*) but never knowledge (*Wissen*), in Kant's strict sense of these terms (A822/B850).

[18] Kant defines (immanent) physiology in general as the discipline that 'regards nature as the sum total of the objects of the senses' (A846/B874). Physiology of inner sense thus regards nature as the sum total of objects in inner sense.

[19] Sometimes Kant also uses the term 'inner appearance'. See (A492/B520), (A673/B701), (A691/B719), and (A722/B800). For a further discussion of inner appearances, see Kraus (2020: 33-4).

[20] On Kant's notion of empirical psychology, see also (4:295), (7:143), (20:308), (28:222-4), (28:656), (28:670), (29:754-6), and (29:954). For a discussion, see also Hatfield (1990: 83-5), Brook (1994: 4), Sturm (2001: 176-7), Heßbrüggen-Walter (2004: 183), Faltudo (2014: 58), Gomes (2017: 15), and especially Frierson (2014).

[21] On the relation between empirical psychology and inner sense, see also (Bxl-xli), (B275-6), (4:471), (7:143), (7:398), (R6313, 18:614), (R6315, 18:620), (25:252), (25:863-5), and (28:224), as well as Frierson (2014: 5).

see also A98-9).[22] Based on the intuitions of inner sense, we thus get to make judgements of the form 'I have representation R, at time t'. Inner sense allows us to judge what we think, desire, and feel.[23]

Kant contrasts inner sense with apperception: 'Inner sense is not pure apperception, awareness of what we are doing; [...]. It is, rather, awareness of what we undergo as we are affected by the play of our own thoughts' (7:161). In inner sense, we do not observe ourselves as the spontaneous '*I as subject*' that thinks, as we do in apperception, but as the passive '*I as object*' that has thoughts (20:230; my emphasis). In inner sense, we also cognise ourselves 'as we appear' and not, as in apperception, 'as we are in ourselves' (B152)[24] – it was Descartes mistake, Kant argues, to confuse these two ways of relating to oneself (Longuenesse 2008; 2017: 110).

So contrasted, inner sense compares with outer sense. Just as inner sense allows us to represent ourselves and our inner states in time, '[b]y means of outer sense [...] we represent to ourselves objects as outside ourselves, and all in space.' (A22-3/B37). Given that objects in space are by definition corporeal for Kant (A7/B11), Kant further states that '[p]sychology is [...] a physiology of inner sense or of thinking beings, just as physics is a physiology of outer sense or corporeal beings' (28:224). Kant thus sees a symmetry, at least in part, between (the physiology of) inner and outer sense, psychology and

[22] As Kraus puts it, Kant defines 'inner sense as the faculty for intuiting one's own representations' (2020: 36).

[23] Kant's account of inner sense suffers from several general problems, which is why it has been called 'dark' (Pippin 1987: 464), 'mysterious' (Chignell 2017: 139), or simply 'a mess' (Brook 1994: 78). Specifically, it is unclear (i) how we can be 'internally affected by ourselves' (B156; see also B67-8, B155); (ii) how time as a form of inner sense becomes a 'mediate condition of outer appearances' (A34/B50; see also A124, A155/B194); and (iii) how Kant can say that representations in inner sense are in time without violating his doctrine of transcendental idealism (21:416). But since these are general problems, they call for general solutions. I will therefore put these problems in brackets here and refer the reader to the helpful discussions by Hölder (1873), Reininger (1900), Kitcher (1990, 2011), Powell (1990), Mohr (1991), Brook (1994: 77-8), Klemme (1996), Valaris (2008), Bader (2010, 2017a), Green (2010), Schmitz (2015), Stephenson (2017), Kraus (2020: 18, 32-3, 61-2, 80), and Benossi (2021).

[24] See also (A116), (A117n), (B132-3), (B139-40), (B144) (B155-6), (B208), (B430-1), (A546/B574), (4:451), (7:134n), (7:140-1), (R4338, 17:511), and (28:227). For a further discussion of these two types of self-awareness and their difference in Kant, see Deleuze (1983/2008: viii), Pippin (1987: 462-7), Kitcher (1990: 22-3), (2011: 21-4), Brook (1994: 55, 66-9, 210, 241-9), Cassam (1997), Carl (2014), Frierson (2014: 3, 24), Gomes (2017: 9, 13), (2024: Ch. 2), Leech (2017a: 184), Longuenesse (2017: 1), Walker (2017: 205), Kraus (2020: xii, 3, 9, 36, 80), and Brennan (ms).

physics, as well as minds and bodies.[25]

Indeed, the Main Claim of my thesis is largely driven by the conviction that there is a robust, although imperfect, symmetry between 'the inner' and 'the outer' in Kant. Specifically, I will be defending two symmetry theses. 'Ontological Symmetry' say that, on Kant's view, mental powers, Reason included, are powers of inner nature, just as physical powers are powers of outer nature; both types of powers are constituted by causal laws.[26] Building on Ontological Symmetry, 'Epistemic Symmetry' says that, on Kant's view, we can come to know mental powers, Reason included, via empirical psychology, just as we can come to know physical powers via empirical physics.[27]

Both of these symmetries are stated in general, holding for all mental powers, and I do suspect that much of my readings generalises to other faculties (though I make no such claim). The reason I focus on Reason is that empirical psychology, construed in analogy to empirical physics, provides powerful tools for addressing the metacritical challenges unique to Reason. My reading also has a decidedly Kantian twist. For, I will argue, it is Reason itself that, in seeking complete explanations for phenomena in inner sense, hypothesises a faculty that is governed by the Principle of Reason. In seeking totalities, Reason has the power to come to know itself, as a power that seeks totalities. That is the power of Reason.

My Main Claim stands, as the saying goes, on the shoulders of giants. Against the widespread anti-psychological sentiment in the second half of the 20th century – Strawson famously called Kant's psychology an 'imaginary subject' (1966/2005: 5) –, Kitcher's *Kant's Transcendental Psychology* (1990) paved the way for scholarship like mine that takes Kant's faculty psychology

[25] On this symmetry, see also (A381), (4:295), (20:308), (28:223), (28:656), and (29:754). For a further discussion of the relation of inner and outer sense, their parity and disparity, see Hatfield (1992: 210), Vogel (1993), Brook (1994: 54), Frierson (2014: 4), Chignell (2017), Kraus (2019, 2020: 4, 32-3), and Brennan (ms), among others.

[26] Heßbrüggen-Walter also notes this symmetry. He argues that, on Kant's view, '[m]ental faculties and powers bring about mental changes just as physical faculties and powers bring about changes in the physical world' (2004: 23). I say more about Kant's causal account of mental faculties in Chapter 2.

[27] I thus agree with Brook, who observes that 'Kant wanted symmetry between knowledge of the self and knowledge of the world' (1994: 249). See also Waxman (1991: 278) and Kraus (2019), (2020).

very seriously.[28] In this respect, my thesis has also benefited greatly from Heßbrüggen-Walter's *Die Seele und ihre Vermögen*, *Kants Metaphysik des Mentalen in der Kritik der reinen Vernunft* (2004), which undertakes the Herculean task of carefully piecing together Kant's sparse notes on the metaphysics of mental faculties.[29]

More recently, scholars have also begun to engage seriously with Kant's *empirical* psychology.[30] Most notable in this regard are Frierson's *Kant's Empirical Psychology* (2014) – the title speaks for itself – and Kraus's *Kant on Self-Knowledge and Self-Formation*, with the important subtitle *The Nature of Inner Experience* (2020). Kraus, specifically, ventures into a comprehensive discussion of how we can gain 'empirical self-knowledge' based on 'inner experience' (2020: 1). Both authors and their works have opened my eyes to the important role that Kant ascribes to inner experience and, thereby, empirical psychology in the quest for self-knowledge.

It is notable, however, that neither of these authors is particularly bullish on using empirical psychology as a metacritical tool. While Kraus thinks that 'inner experience' can give us insight into our 'thoughts, memories, imaginings, feelings and desires', and even into 'personality traits [...], attitudes, commitments and values' (2020: 1), she stops short of claiming that this insight extends to our faculties.[31] Frierson, who is more optimistic than Kraus, suggests that empirical psychology 'describes the operations of [human faculties] "from without"', but adds the limiting remark that 'the findings of empirical psychology provide no direct support for [Kant's] transcendental philosophy' (2014: 2).[32]

---

[28] Of course, Kitcher was not the first to be right about this. Kemp Smith is on record saying: 'No interpretation which ignores or under-estimates this psychological [...] aspect of [Kant's] teachings can be admitted as adequate' (1918/1962: 52).

[29] And which is yet to be translated!

[30] Kitcher's transcendental psychology often has a slightly empirical flavour: 'Transcendental psychology [...] has striking affinities with empirical psychology' (1990: 26). See especially Kitcher (1990: 21-6, 139-40).

[31] See also Kraus (2020: 4, 17, 24, 39, 43, 171, 184, 212, 228, 246). At one point in her book, Kraus explicitly states that Kant's tripartite division of the mind into the faculties of cognition, desire and feeling must be based on 'transcendental justification' (2020: 25). In conversation, Kraus told me that she deliberately refrained from extending her account of inner experience to include faculty-knowledge.

[32] To be fair to Frierson, however: he claims that the goal of empirical psychology 'would

Most interpreters, however, are downright pessimistic about the use of empirical psychology in Kant. Kitcher calls introspection a 'dubious means' of gaining self-knowledge, and moreover 'a potential route to lunacy' (1990: 6). Heßbrüggen-Walter concedes that empirical psychology may help us identify down-to-earth faculties like 'motherly wit', but denies that it can help us establish more aloof faculties, like Reason (2004: 10).[33] And then there are also people like Brook, who confidently proclaims that when it comes to faculty-knowledge, 'empirical psychology is a nonstarter' (1994: 9). That's just the tip of the pessimistic iceberg.[34] My Main Claim should therefore be as original as it is controversial.[35]

1.4 Four Major Problems

The defence of my Main Claim will be an uphill battle. There is good reason to be pessimistic about the role of empirical psychology in Kant, especially when it comes to answering the Epistemic Question. Setting aside Kant's dismissive comments on empirical psychology in the *Metaphysical Foundations of Natural Science* (4:470-1) and the *Anthropology* (7:133; 7:161) – more on those in Section 6.3 –, there will be a general worry that the fleeting stream of representations we observe in inner sense, this soup of thoughts, desires, and feelings, is too weak to establish a faculty as strong as totality-seeking Reason. This general worry can be translated into four specific problems, which I will address over the course of this thesis.

The first of these four problems is the 'Problem of Lawfulness'.[36] The

be a comprehensive account of the causal laws that govern [...] various powers of the soul' (2014: 8). Thus, by learning about these laws in empirical psychology, we would also learn about our mental faculties. I do agree with that.

[33] That said, Heßbrüggen-Walter is also on record with the following statement: 'Kant draws on introspective certainties about the structure of mental life to justify his architectonics of faculties' (2004: 263).

[34] See, for example, also (Dyck 2014: ch. 7) and Teixeira (2019: x).

[35] Howard (forthcoming) endorses a version of both Ontological Symmetry and Epistemic Symmetry. However, his reading remains more programmatic. Specifically, he does not address the problems I will discuss below, nor does he explain how the symmetries can be used to come to know Reason specifically.

[36] For a version of this problem, generalised to all faculties, see Deleuze (1983/2008: 11).

problem goes like this: we come to know that we have Reason by identifying a faculty governed by the Principle of Reason. The Principle of Reason, in turn, is a universal law. So, to come to know that we have Reason via empirical psychology, we would have to be able to identify universal laws based on the changes we observe in inner sense. But, the concern goes, this is not possible because inner sense is a lawless Wild West, in large part because Kant claims that there are no substances in inner sense that could ground law-governed causal powers.[37] The problem can be put precisely as follows:

Problem of Lawfulness

(P1) If the Principle of Reason is a universal law, then we can come to know that we have Reason via empirical psychology only if inner sense stands under universal laws.

(P2) The Principle of Reason is a universal law.

(P3) Inner sense does not stand under universal laws.

(C) *Therefore*, we cannot come to know that we have Reason via empirical psychology.

The second problem is the 'Problem of Normativity'. The concern is that even if we could identify laws in the inner sense, they would have to be descriptive causal laws of nature, which tell us that representations of type A *are being* followed by those of type B – remember that empirical psychology studies 'sum total of inner perceptions *under natural laws*' (7:141; my emphasis).[38] But the Principle of Reason, says the problem, isn't descriptive but

[37] The schema of the category of substance is 'the persistence of the real in time' (A144/B183). But nothing persists in inner sense, because the empirical awareness of ourselves is 'forever variable' (A107), which is why the category of substance cannot be applied to inner sense. This conclusion is independently confirmed by Kant's Refutation of Idealism, as well as Kant's discussion of the first Paralogism. I will say more about this in Section 2.3. Pointing to (A54/B78-79), (9:14), (24:792), Martens (ms) raises a different but related problem. She suggests that empirical psychology only ever gives contingent natural laws, and never the absolutely necessary laws of thinking, which would include the Principle of Reason.

[38] Kant leaves no doubt on this point: 'psychology [...] explains what happens, [and does] not prescribe what ought to happen' (R5864: 18:371). See also (5:182), (24:25), and (24:339). Frierson suggests that empirical psychology is descriptive whereas transcendental psychology can be normative (2014: 45). Hatfield (1990: 86), Heßbrüggen-Walter (2004: 184), Gomes (2017: 17), and Schafer (2018a: 524n) formulate a similar concern.

normative, demanding that we *ought* to seek the totality of conditions for a given conditioned (A307/B364).[39] There is thus a mismatch between the laws of the inner sense and the Principle of Reason, which makes it impossible to establish Reason via empirical psychology. Again, put precisely:

Problem of Normativity

(P1) If the Principle of Reason is a normative principle, then we can come to know that we have Reason via empirical psychology only if laws of inner sense are normative.

(P2) The Principle of Reason is a normative principle.

(P3) Laws of inner sense are not normative.

(C) *Therefore*, we cannot come to know that we have Reason via empirical psychology.

The third problem is the 'Problem of Perfection'. Even if there were a descriptive version of the Principle of Reason, it would likely have to specify how a perfect reasoner actually reasons. But such a principle is something we could never identify through empirical psychology, because our actual reasoning in inner sense is clearly imperfect.[40] Not only do we draw invalid inferences, but as finite beings we have never completed the search for an infinite totality of conditions, and never will; all we can see is that we have sought this or that condition for a given conditioned – just as we can only have apperceptive act-awareness of synthesising some rather than all conditions. Stated as a strict argument:

---

[39] Kant is clear that Reason makes 'demands' (A656/B684; see also A699/B727). Kraus thus notes that the principles 'of reason are *expressions of normative demands*' (2020: 182); and that they have '*prescriptive force*' (2020: 182n). See also Heßbrüggen-Walter (2004: 24).

[40] Kant notes that 'man often uses his understanding and thinks differently than he should think' (24:25). Kant also thinks that our reasoning falls prey to a 'natural and inevitable illusion', which makes us assert rather than seek the totality of conditions (A298/B354). Kant is, of course, not the only one to note our tendency to disobey reason. Russell quips: 'Man is a rational animal – so at least we have been told. Throughout a long life, I have searched diligently for evidence in favour of this statement, but so far I have not had the good fortune to come across it' (1921: 95). Based on their empirical study of probabilistic reasoning, Kahneman & Tversky come to reject the idea that 'inside every incoherent person there is a coherent one trying to get out' (1983: 313).

Problem of Perfection

(P1) If the Principle of Reason specifies perfect reasoning, then we can come to know that we have Reason via empirical psychology only if reasoning in inner sense is perfect.

(P2) The Principle of Reason specifies perfect reasoning.

(P3) Reasoning in inner sense is not perfect.

(C) Therefore, we cannot come to know that we have Reason via empirical psychology.

The fourth problem is the 'Problem of Aprioricity'. The problem simply insists that the Principle of Reason must be knowable a priori; after all, Kant's discussion of Reason is part of transcendental philosophy, broadly speaking, which is an a priori enterprise (A11-2/B25; A56/B80-1). However, inner sense can, by definition, provide no a priori justification for any principle or law, and so neither can empirical psychology (B152). That's also why Kant banishes empirical psychology from metaphysics strictly speaking (A848/B876).[41] Put as argument:

Problem of Aprioricity

(P1) If the Principle of Reason is an a priori principle, then we can come to know that we have Reason via empirical psychology only if inner sense provides a priori grounds.

(P2) The Principle of Reason is an a priori principle.

(P3) Inner sense does not provide a priori grounds.

(C) Therefore, we cannot come to know that we have Reason via empirical psychology.

All four problems directly threaten the Main Claim of my thesis. However, they also interact in interesting ways with the two symmetry theses introduced above. The first and second problem put pressure on Ontological Symmetry because they indicate that mental powers, Reason included, differ

[41] Kitcher writes: 'The focus on *a priori* origins is clear in the text. [...] Kant [insists] that accounts of *empirical* origins [...] are without philosophical interest' (1990: 15).

from physical ones either by not being part of a law-governed inner nature, or by not being governed by descriptive laws. The third and fourth problem, in turn, put pressure on Epistemic Symmetry insofar as they suggest that empirical psychology faces epistemic limitations that empirical physics doesn't.

There may, of course, be other objections to my Main Claim, and I will address some of them in passing throughout this thesis. One issue I won't address, however, is the following circularity concern: we cannot establish any faculty in Kant's transcendental philosophy from experience, because these faculties are meant to explain the very possibility of experience in the first place.[42] While this concern has pull with sensibility and the understanding, it doesn't arise for Reason, which is not necessary for the possibility of experience. (And even if it were, my answer to the Problem of Aprioricity will also pre-empt any such a circularity concern.)

1.5 Chapter Overview

Here then is the structure of my thesis. The primary goal of my thesis is to defend the Main Claim and thereby answer the Epistemic Question. While I will provide some positive support for Ontological Symmetry and Epistemic Symmetry, much of my thesis is dedicated to addressing the four major problems introduced above. However, there is no one-to-one mapping of problems to chapters. The structure of my thesis is a bit more complicated. Let me explain.

In Chapter 2, I motivate Ontological Symmetry and address the Problem of Lawfulness. I first develop Kant's causal account of powers, suggesting that this account should extend to mental powers. Powers of inner and outer sense are constituted by causal laws. I then address the problem that inner

[42] Strawson elegantly articulated the circularity concern when he writes that 'transcendental psychology [...] 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' (1966/2005: 13). Frierson remarks that empirical psychology must hence presuppose transcendental psychology (2014: 45). On the circularity concern, see also Pippin (1987: 456), Kitcher (1990: 177), Brook (1994: 4-5), and Leech (2017a: 184).

sense cannot stand under such laws because nothing persists in inner sense, yet persistence is the schema of substance, which is required as bearer of causal powers. Drawing extensively on Kant's discussion of the regulative use of the idea of the soul, I respond by arguing that we are justified in assuming an inner substance, and so inner powers, in the mode of what Kant calls 'doctrinal Belief' (A825/B853).

In Chapter 3, I address the Problem of Normativity, by denying that the Principle of Reason must be normative. Kant is committed to the principle that 'ought' excludes 'must', according to which we only ought to φ if we are able to not-φ. I then reconstruct Kant's theory of faculty failure and argue that Reason never fails on its own, but only through the corrupting influences of sensibility. The Principle of Reason is therefore descriptive of Reason but can be normative for finite rational beings like us, that can fail to abide by the Principle of Reason. The chapter touches on the Normative Question, without, however, attempting to answer it.

Chapters 4 and 5 take a small detour to address the Material Questions: what is the exact content of the Principle of Reason? In Chapter 4, I argue that the demand of Reason to seek totalities of conditions is in a qualified sense identical to the demand to seek a system of all empirical laws, as Kant discusses it in the Appendix. Building on this, Chapter 5 considers how, according to Kant, we should go about establishing a system of empirical laws. My positive suggestion is that we hypothesise laws to explain observable phenomena, and that we should adopt those hypotheses that have the greatest probability given the evidence.

Chapter 6 tackles the two remaining problems, the Problem of Perfection and the Problem of Aprioricity. Here is the rough story: Reason, for Kant, seeks to explain all phenomena, including those of inner sense. To explain inner phenomena, Reason hypothesises mental faculties and their laws. Our inner urge to ask why-questions, Kant thinks, is best explained by hypothesising a faculty that demands complete explanations, i.e., Reason. Moreover, the principle of this faculty, the Principle of Reason, can be confirmed a priori, in the same way that some empirical laws in physics can be confirmed a priori.

2

Laws of the Mind

In this chapter, I tackle the Problem of Lawfulness by rejecting its third premise: there are laws in inner sense, laws that govern our mental powers. Here is the plan: I first reconstruct Kant's causal account of powers (§ 2.1), suggesting that it motivates Ontological Symmetry – even though higher mental powers are meant to be spontaneous (§ 2.2). I then introduce what I call the 'Asymmetry Argument', which rejects Ontological Symmetry on the basis that there are no substances in inner sense that could be a bearer of low-governed powers (§ 2.3). Finally, I address the Asymmetry Argument by arguing that inner substances can be the object of both knowledge and rational belief (§§ 2.4 and 2.5).

2.1 Causal Powers and Faculties

Kant rarely explicates his notion of *power* (*Kraft*). Heßbrüggen-Walter complains that the first Critique 'does not give a definition of power or faculty' and it 'also does not specifically introduce the key term of a faculty of the soul [*Seelenvermögen*]' (2004: 152). But Kant isn't lazy. He simply draws on the 'the hitherto known textbooks' (A204/B249), especially Baumgarten's *Metaphysics*, which operate in a broadly Aristotelian framework.[43] Since Kant borrows the concept of power from his predecessors, he doesn't change it

[43] For the historical backdrop to Kant's notions of power and faculty, see Heßbrüggen-Walter (2004: chs. 1 and 2) and Dyck (2014: 202-5).

throughout his lifetime.[44] This consistency, in turn, allows us to draw on passages from the entire corpus, including the pre-Critical texts, to clarify Kant's notion of power.

Kant sees power as an essentially *causal* concept. The concept of power derives, as 'a *predicable*[...] of the pure understanding', from the category of causation (A82/B108).[45] Since 'the concept of causation carries that of laws with it' (4:446), powers must be governed by causal laws; indeed, no 'power [...] can deviate from its law', which defines its unique activity (24:396).[46] (I will return to this point in Chapter 3.) But not only are there no law-less powers, but there are also no power-less (causal) laws (Watkins 2005). On Kant's account, causal laws are not free-floating functions that range over sets of events, but they are principles that govern powers. Thus, if there were laws in inner sense, they would have to be laws of powers.[47]

Kant's concept of power also leads to the concept of *activity* (*Handlung*),[48] which too is a predicable of causation (A82/B108). An activity results from the exercise of a power, 'From power, activity can be derived' (29:772).[49] For example, attraction is the activity that results from the power of attraction; cognition from the power of cognition. The examples show that Kant's notion of activity – which he adopts from Crucius, who defines an activity (*actio*) as the success (*Erfolg*) of a power (1745/1964: § 64, 114)[50] – must be taken in a thin sense. While intentional actions are activities in this thin sense, not all

[44] Gerhardt (1986) and Heßbrüggen-Walter (2004: 25; 152) have emphasised that Kant's concept of power remains largely unchanged throughout the corpus.

[45] Kant also establishes the link between power and causation in (A204/B249), (A648/B676), (4:257), (28:431), and (28:564). For a further discussion, see Gerhardt (1986), Friedman (1992a: 162), Heßbrüggen-Walter (2004: ch. 4.4), and Frierson (2014: 6).

[46] See also (R2244, 16:283), (24:84), (24:102), (24:316), (24:396), (24:402), and (25:577). Heßbrüggen-Walter notes: 'Insofar as a natural power causes an activity, this has to happen according to a law [...]. But not to act according to the laws is then impossible; [it] would mean not to act at all' (2004: 190). See also Frierson (2014: 6) and Hutton (2021: 93-4).

[47] Only changes of substances are governed by laws (A182-3/B224-5); yet changes in substances, as we will shortly see, are the result of powers. On Kant's causal account of powers, see also Heßbrüggen-Walter (2004), Kraus (2020: 61), and Indregard (2017).

[48] I am translating 'Handlung' as 'activity' to indicate that it does not presuppose any intention, or indeed, agent.

[49] See also (A204-5/B250-1), (4:257), and (28:564-5). For a further discussion, see Gerhardt (1986), Friedman (1992a: 162), and Heßbrüggen-Walter (2004: 14), (2015: 2483).

[50] For a further discussion, see Heßbrüggen-Walter (2004: 94, 135).

activities are intentional. As we will see in a moment, activity is more closely connected to actuality than agency.

Causal powers need *substances* as their bearers. The 'causality of a substance [...] is called power' (A648/B676).[51] Unlike Baumgarten, Kant distinguishes the bearer of a power from the power itself: 'the substance *is* no power, but *has* a power' (28:25).[52] For example, matter has the power to attract; the soul or mind has the power to cognise; etc. The activities that result from the exercise of a power are then said to be accidents that inhere in the substance: 'The relation of a substance to its *accidente inhaerente* is the [...] power" (28:25). In that sense, Kant can be said to have a causal notion of inherence.[53]

Kant's notions of substance, power, and activity (or accident) are mutually entailing. Kant writes: 'Where there is activity [...] and power, there is also substance' (A204/B250); and a few lines earlier: 'Causality leads to the concept of activity, which leads to the concept of power, and thus to the concept of substance' and this 'concerns the analysis of [...] the concepts' (A204/B249).[54] The inverse holds too because every substance has at least one accident – there are no bare substances for Kant –, which must result from a power, and so: 'Every substance has a power is an identical sentence' (R4056, 17:400).[55] Substance, powers, and activity thus come as a conceptual package for Kant. This analytic entailment will matter in Section 2.4.

---

[51] See also (A204-5/B249-50), (R4056, 17:400), (28:431), and (29:772). For a discussion, see Gerhardt (1986) and Heßbrüggen-Walter (2004: 14-5, 154-5).

[52] 'E.g. the light rays have a warming power, but I cannot say they are a warming power' (28:431). See also (28:262), (28:671), and (29:771). For a further discussion, see Heßbrüggen-Walter (2004: 134-6), (2015: 2481), and Ferrarin (2015: 45, 290). Ameriks notes that the mind must be distinguished from its powers (2000: 65).

[53] See also (2:416), (R3785, 17:292), (R4056, 17:400), (28:431), (29:770-1), and (29:823). While actions inhere in a substance in virtue of their power, powers themselves cannot inhere in a substance, or else they would have to be the activities of yet another power, and so on *ad infinitum*. Thanks to Andrew Chignell for pointing this out to me.

[54] Kant thus thinks that causation requires substance, which is further discussed by Sturm (2001: 169-71), Watkins (2005), Heßbrüggen-Walter (2004: 154-7), Frierson (2014: 22), and Ferrarin (2015: 290).

[55] Heßbrüggen-Walter (2004: 58) suggests that Kant adopts this idea from Wolff, who thought that all substances have (at least) one power.

Kant often uses 'power' interchangeable with 'faculty' (*Vermögen*) – and so will I.[56] Like powers, faculties are predicables of causation and so governed by causal laws (A82/B108). But there is a subtle difference between the two that we should be aware of. Drawing on Wolff's modal distinction, Kant states that, '[i]n the case of a faculty, we represent the *possibility* of an activity' (28:434; my emphasis); by contrast, '[t]he determining ground for the *actuality* of an activity is [...] called power' (28:515; my emphasis).[57,58] Kant further specifies that a power is the ground of actuality because it is 'internally sufficient' whereas faculties 'do not contain the sufficient ground of their activity' (28:434).[59]

But how should we understand Kant's notion of internal (in)sufficiency? I propose the notion is best understood if we think of faculties *in general* as conditional dispositions to manifest an activity φ when condition c obtains.[60] We can say then: a faculty to-φ-when-c is internally sufficient iff condition c obtains; and it is internally insufficient iff condition c does not obtain. Internally sufficient faculties are powers, which make φ-ing actual; internally insufficient faculties are a '*mere* faculties" (R3588, 17:75; my emphasis), which make φ-ing merely possible. Powers, then, are species of faculties in general – 'power is [a] faculty' (29:823) –, but they are not mere faculties – 'faculty and power must be distinguished' (28:434).

---

[56] Kant indiscriminately uses 'faculty' and 'power' in many places, like (2:60), (5:46-7), (R5864: 18:371), (28:262). On this point, see also Ameriks (2000: 242), Heßbrüggen-Walter (2004: 139), Faltudo (2014: 24), and Ferrarin (2015: 291).

[57] Wolff draws the same distinction in his *German Metaphysics* (1751/1983: §117, 62), *Rational Psychology* (1740/1972: § 54, 35-6), and his *Philosophia Prima* (1736/1962: § 722, 542). For a discussion of Wolff's distinction, see Heßbrüggen-Walter (2004: 57), (2015: 2481), Dyck (2014: 30-33), Faltudo (2014: 9-11), and Ferrarin (2015: 287). Falduto (2014: 10) notes that some of Wolff's successors have collapsed the distinction. For example, Crucius identifies faculties and powers in his *Entwurf* (1745/1964: § 29, 46). However, Zedler (1731–1754: 1328) distinguishes the two.

[58] Kant repeats the distinction between faculties and powers in (17:73), (R3785, 17:292), (28:27), (29:772), and (29:823). See also Heßbrüggen-Walter (2004), (2015: 2483), Faltudo (2014: 11, 24), Ferrarin (2015: 45, 291), Willaschek (2018: 21n2), and Schafer (2019: 179n12). I thus disagree with Kitcher, who claims that 'a "faculty" is simply a short way of referring to a set of powers or capacities' (1990: 10). See also Walsh (1966).

[59] See also (R3588, 17:75), (R3589, 17:76), (28:27), (28:515), (29:772), and (29:823). For a further discussion, see Heßbrüggen-Walter (2004: ch. 4.2).

[60] Other authors have touched on such a dispositional analysis. For example, see Ameriks (2000: 65), Heßbrüggen-Walter (2004: 91, 139, 142, 167), Ferrarin (2015: 287), and Engelhard (2018: 8).

Reading faculties in general as conditional dispositions also enjoys independent textual support. Not only does Kant align 'disposition [*Disposition*]' with 'faculty [*facultas*]' (1:131),[61] but the conditional analysis fits quite well with Kant's account of causation. After all, faculties are governed by causal laws, which for Kant have a conditional form, connecting cause and effect. It is natural to identify the condition c with the cause, the resulting activity φ with the effect. The causal laws governing faculties may thus take the form: 'Necessarily, when condition c obtains, substance S φs'.[62] Insofar as S's φ-ing is the trigger condition for another faculty, this also allows us to stitch together causal chains.

While all powers are internally sufficient, Kant suggests that powers can 'either be externally insufficient [...] or externally sufficient' (R3588, 17:75). A power to-φ-when-c is externally sufficient iff there is no real opposition to its φ-ing; and it is externally insufficient iff there is real opposition to its φ-ing.[63] Kant's notion of real opposition is demanding, but the rough idea is that the effects of powers can (partially) cancel each other out, so that the φ-ing, while actual, doesn't show. Kant calls externally insufficient powers *dead* (*tot*) and externally sufficient powers *living* (*lebendig*). Kant articulates these points clearly in his *Metaphysics Volckmann*:

> All power is either living, which acts (and thus is internally and externally sufficient, it is the cause of the effect or the cause of the accident), or dead, which is internally sufficient and yet externally insufficient, or whereby no activity takes place, because there must be an external cause to the contrary, e.g., every body has a power to fall, but this does not happen if another opposing power resists it; so far as there is no opposition, all powers are living. (28:434)[64]

The distinction between dead and living powers raises epistemic chal-

[61] At times, however, Kant also identifies 'disposition' with 'inclination', as he does, for example, in (R1494; 15:756).

[62] For a similar analysis, see Friedman (1992a: 163-4).

[63] Kant is clear that the distinction between external sufficiency and insufficiency only applies to powers, i.e., internally sufficient faculties in general (R3588, 17:75). On Kant's notion of real opposition, see Heßbrüggen-Walter (2004: 73-4, 177-9).

[64] See also (R3588, 17:75), (17:76, R3589), and (28:27). The distinction is further discussed by Heßbrüggen-Walter (2004: 73-4; 177-9), (2015: 2483).

lenges. One is that we are unable to detect powers whose activities are persistently cancelled.[65] But more worryingly, a power's activity can also be partially cancelled or otherwise distorted. So while we are able to analytically infer from an apparent activity φ to *some* power, we cannot infer to a power to-φ-when-c (for some c) because the observed φ-ing may be the aggregate result of multiple powers. I call powers that are partially cancelled 'zombie powers' – they are neither living not fully dead. Reason will turn out to be a zombie power that comes to haunt us (see Chapter 3). Kant's taxonomy of faculties in general is depicted in Figure 1.

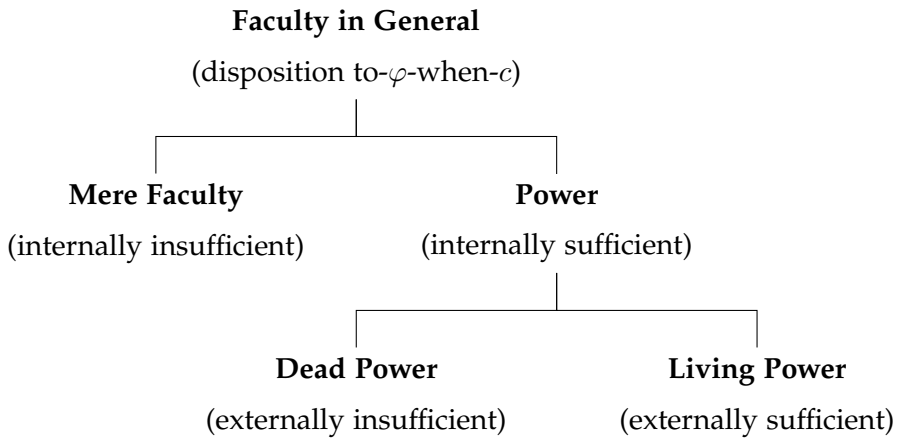

Figure 1: taxonomy of faculties

The causal account reconstructed in this section recommends individuating all faculties in general via their governing law. While we can, of course, also individuate faculties via their unique activity φ – as Kant does when he introduces the faculties of cognition, desire, and feeling (5:177; 20:205-6)[66] –, individuating faculties by their laws, which includes condition c, generates a more fine-grained picture that allows us, for example, to distinguish between

[65] As Kant puts it, 'from the power does not always follow the effect (logically)' because the power can be dead (28:26).

[66] See also (A19/B33). Gomes (2017: 7) and Willaschek (2018: 21) suggest that Kant individuates mental faculties by their representation. Heßbrüggen-Walter (2004: 260) conversely claims that Kant individuates representations by the faculties from which they arise.

different faculties of cognition, like understanding and Reason.[67] I thus agree with Schafer, who notes: 'As Kant understands them, [faculties] are individuated by principles which [...] both describe and govern their activity' (2019: 184).[68] Reason is individuated by the Principle of Reason.

2.2 Ontological Symmetry

Nothing in Kant's causal account of faculties and powers – I resume treating both terms interchangeably – suggests that it is limited to physical powers only. The account is clearly meant to apply to mental faculties also. After all, Kant explicitly calls mental faculties powers, '*powers* of the soul' ('*Seelenkräfte*') or '*powers* of the mind' ('*Gemüthskräfte*').[69] Some faculties even carry 'power' in their name, like the power of imagination (*Einbildungskraft*) and the power of judgement (*Urteilskraft*). Moreover, Kant is clear that mental powers are powers of our 'mental nature [*geistige Natur*]', just as physical powers are of 'bodily nature [*körperliche Natur*]' (A684/B712).[70] The *Jäsche Logic* expresses this symmetry vividly:

> Everything in nature, in the lifeless and in the living world, happens according to rules [...] – The water falls according to laws of gravity, and with the animals the locomotion also happens according to rules. The fish in the water, the bird in the air moves according to rules. The whole nature in general is actually nothing else than a connection of phenomena according to rules; and there is everywhere no lack of rules. [...] The exercise of our [mental] powers also happens according to certain rules. (9:11)[71]

[67] Tolley (ms) tries to account for the difference between understanding and Reason by distinguishing different kinds of cognition, suggesting that understanding produces level-four cognition and Reason produces level-seven cognition as per the stepladder in the *Jäsche Logic* (9:65). Even if successful, my point still stands as the individualisation of faculties by law will be more fine-grained.

[68] On faculty identity, see also Deleuze (1983/2008: 3), Heßbrüggen-Walter (2004: ch. 5), Schafer (2018a: 515-6), Willascheck (2018: 21), and Kraus (2020: 25-6).

[69] I searched the Academy Edition and counted 119 hits for 'Gemütskräfte' – more than double the 50 hits suggested by Faltudo (2014: 25n) – and 40 hits for 'Seelenkräfte'. Both searches included variations in spelling and grammar. Relevant passages include (2:360), (4:368), (4:435), (5:359), (5:350), (5:355), (7:181), (7:225), (16:135), (18:371), (20:245), and (20:247).

[70] Kant also calls reason a 'natural disposition [*Naturanalge*]' (5:62).

[71] See also (24:790), (24:693), and (27:503).

Insofar as our mental powers are part of nature, their rules must be causal laws of nature. Indeed, Kant thinks there are 'inner causes' that give rise to '[o]ur representations [...], which, as modifications of the mind, belong to inner sense' (A98-9).[72] To illustrate this point, Heßbrüggen-Walter suggests that we can think of 'a syllogism as a sequence of mental states' connected by causal laws (2004: 182). For example, consider the sequence of states (i) to (iii) below – all states I can observe in inner sense. My faculty to draw modus ponens inferences takes (i) and (ii) as its causal condition c, and manifest (iii) as its activity φ:

(i) I represent 'if p, then q' at time t_1.
(ii) I represent 'p' at time t_2.
(iii) I represent 'q' at time t_3.

These considerations strongly motivate Ontological Symmetry: Kant's causal account of powers is supposed to apply equally to physical and mental powers.[73] This causal understanding of mental powers pushes back against the Problem of Lawfulness by questioning its third premise. Why should inner sense not be subject to causal laws? Insofar as inner states inhere in a mental substance, they must be an activity that results from the exercise of one or more mental powers. Mental powers, in turn, are governed by causal laws, so their activities be subject to those laws. We should thus expect inner sense, like outer sense, to stand under causal laws.

However, Ontological Symmetry is not without its problems. Before we turn to the Asymmetry Argument (Section 2.3), let me quickly address one salient worry: Kant thinks that the so-called 'higher faculties' are spontaneous (7:141, 9:36). This includes Reason. Yet spontaneous faculties, the worry goes,

[72] See also (4:471) and (20:237). Kant's notion of 'inner causes' is also discussed by Sturm (2001) and Chignell (2017: 155-6).

[73] I am not the first to endorse Ontological Symmetry. Howard, who translates 'Kraft' with 'force', notes: 'Kant holds a single concept of force [...], that he applies [...] in the domains of physics and psychology' (forthcoming). Kraus similarly remarks that Kant describes 'mental faculties as causal powers' (2020: 61). See also Hatfield (1992: 208, 223), Brook (1994: 240), Ameriks (2000: 241), Heßbrüggen-Walter (2004: 168, 183, 236-8), Frierson (2014: 6, 22, 53-4), Wuerth (2014: 183-205, 260), Ferrarin (2015: 287), Chignell (2017: 140), and Indregard (2017).

cannot stand under laws of nature because Kant identifies '*spontaneity*' with the ability 'to start a series of appearances [...] *by oneself*, that is, transcendental freedom' (A446/B474).[74] Since freedom starts a causal chain, laws of freedom, unlike laws of nature, don't invoke an antecedent causal condition (A840/B869; 4:446-7). Spontaneous mental powers must therefore drastically differ from physical powers that aren't spontaneous.[75]

There are two potential replies to this worry. The first reply, ventured by Frierson (2014: 14-8), leverages Kant's transcendental idealism. The distinction between causally determined phaenomena and free noumena allows Kant to claim that 'inner and outer actions' can be fully causally determined and yet be free at the same time (5:99). Frierson suggests that this point generalises to the powers from which these actions arise. We can consider our faculties from two perspectives, 'an empirical-psychological perspective [...] and a free-practical perspective' (2014: 16). Yet both perspectives pick out 'the very same faculties of the soul' (2014: 44).[76] This reading draws support from the *Groundwork*:

> [A] rational being [...] has *two standpoints* from which it can observe itself and recognise *the laws of the use of its powers*, and consequently of all its actions: first, so far as it belongs to the world of the senses, under *natural laws* (heteronomy); second, as belonging to the intelligible world, under *laws* which, *independent of nature*, are not empirical, but are founded merely in reason. (4:452; my emphasis; see also 29:862)

[74] See also (A533/B561), (A548/B576), (4:452), (5:48), (5:99-101), (5:186), (5:197), (5:406), (5:411), (6:24), (6:50), (6:143), (7:134), (7:141-2), (8:223), (8:417), (9:36), (9:76), (21:470-1), and (23:261). Gomes suggests that our faculties have a 'logical and transcendental structure' that places them outside the causal order of nature (2017: 6).

[75] This worry about spontaneity has two close siblings. First, one might argue that higher faculties don't fit the causal account of powers because they have a teleological structure. However, this concern has already been addressed by Frierson (2014: 16-8; 28-32) and Schafer (2019: 186), so I won't deal with it here. Second, one might object that higher faculties cannot be governed by causal laws because their principles are normative. That's the Problem of Normativity, which I address in Chapter 3.

[76] In a similar spirit, Kraus suggests that 'Kant frequently introduces both a transcendental and an empirical variant of major faculties' (2020: 25). See also Hatfield (1990: 84) and Faltudo (2014: 39).

The same power can thus be viewed from two standpoints: one where it is governed by laws of nature, and one where it is governed by laws of freedom. This would allow us to maintain Ontological Symmetry for the first of these standpoint. That said, I have two issues with this reply. First, if we individuate faculties by their laws – as I suggested in Section 2.1 –, then it is impossible for the same faculty to be governed by both laws of nature and laws of freedom, unless these laws are somehow identified. Second, for Kant there is only one law of (transcendental) freedom, namely the moral law (4:446-7). But the moral law is the 'fundamental law of pure practical reason' (5:30), which leaves unclear how other faculties could be governed by laws of freedom.

This leads me to the second reply, which I prefer. As noted by Sellars (1970), Kant distinguishes between *relative* and *absolute* spontaneity. Whereas 'absolute spontaneity' requires the 'transcendental concept of freedom', relative spontaneity requires only what Kant calls 'psychological freedom' (29:267-8). But what is psychological freedom? Unlike transcendental freedom, which initiates a causal chain, psychological freedom implies antecedent causes. Kant gives the example of a watch whose ticking is causally determined by an 'internal principle', namely its spring; the watch is relatively spontaneous even if the internal principle itself depends on an 'external principle', in this case the watchmaker (ibid.).[77]

But if relative spontaneity is consistent with antecedent causes, how does it differ from mere receptivity? The contrast is instructive because Kant is surprisingly talkative about the metaphysics of receptive faculties.[78] Receptivity requires that '2 substances' interact (28:52).[79] In the *Metaphysics*

[77] See also (4:544), (28:285), and (28:448-9).

[78] Strictly speaking, receptivity is not a faculty. Like Wolff (1736/1962: §§ 536-7, 713-4), Kant draws the following contrast: 'The possibility of acting is faculty, the possibility of suffering is receptivity' (29:722; my emphasis). See also (7:140), (25:474), (28:26), (28:443), (29:823). But Kant also suggests that sensibility is a 'faculty of receptivity' (9:36). Receptivity must thus be a faculty in a broader sense. Since, as we shall see, receptivity fits well with the dispositional analysis of Section 2.1, I suggest that this broader sense is that of a faculty in general. See also Heßbrüggen-Walter (2004: 175-6, 234).

[79] Kant reaffirms that receptive faculties require two substances in other places. See (29:772) and (29:823). See also Kant's definition of 'influence' (B111) as well as Heßbrüggen-Walter (2004: 25).

Herder, Kant offers the example of a trumpet: the first substance is the trumpet, which has 'the power [to produce] music' when one blows into it; the second substance is the mind or soul, which has 'its own power to listen', i.e., the power to produce auditive representations when affected by music (ibid.). In this example, the mind's power to listen is meant to be the receptive faculty.

The examples motivate the following proposal. A receptive faculty is one whose causal condition is the state or activity of another substance – the cause of the *mind*'s representation is the *trumpet*'s playing. Put precisely, the faculty to-φ-when-c of a substance S_1 is receptive iff c is the ψ-ing of a substance S_2 and $S_1 \neq S_2$. By contrast, a relatively spontaneous faculty requires that the causal condition is the state or activity of the same substance – the cause of the *watch*'s ticking is the *watch*'s state of the spring. Precisely, the faculty to-φ-when-c of a substance S_1 is relatively spontaneous iff c is the ψ-ing of a substance S_2 and $S_1 = S_2$. Both types of faculties are thereby contrasted with absolutely spontaneous faculties, which don't depend on a causal condition c.[80]

This proposal naturalises relative spontaneity. Relatively spontaneous faculties, like receptive ones, don't start causal chains but they invoke antecedent causes. As such they don't stand under laws of freedom, but laws of nature. Mental faculties that are relatively spontaneous are thereby part of our mental nature, and their activities are observable in inner sense. Indeed, when Kant introduces relative spontaneity, he specifies not only that it can generate 'activities of the soul' (29:268), but also that the 'psychological freedom' with which it is identified 'is treated in empirical psychology' (29:267). Things don't get more explicit than that.

An implication of my reading, however, is that we must limit Ontological Symmetry (and thus also Epistemic Symmetry) to those faculties that are not absolutely spontaneous. Absolutely spontaneous faculties differ ontologically from natural powers, whether physical or mental, in that they are governed

[80] However, see the excellent work of Brink (ms) for an alternative interpretation of relative and absolute spontaneity in Kant.

not by laws of nature but by laws of freedom. And although laws of freedom are causal laws for Kant (4:446-7) – that said, I reserve the term 'causal laws' to denote causal laws of nature –, they don't place absolutely spontaneous faculties in the natural order, and so activities in nature must not be explained with reference to absolutely spontaneous faculties (29:862).

But this limitation is a small price to pay, for there is only one absolutely spontaneous faculty, and that is pure *practical* reason. Crucially, Reason – with a capital 'R' – is defined as *theoretical* reason in the narrow sense, and should therefore not be absolutely spontaneous, but only relatively so. Ontological Symmetry and Epistemic Symmetry hence apply to Reason, but not pure practical reason. Indeed, Kant remarks in the *Critique of Practical Reason* that '*experience* entitles us to assume' theoretical reason; but pure practical reason and its governing principle, the moral law, can only be known via the 'fact of pure reason, of which we are *a priori* aware' (5:47; my emphasis; see also 4:391).[81]

Let's take stock. In Section 2.1, I have suggested that Kant generally conceives of powers as conditional dispositions that are governed by causal laws. In this section, I have defended Ontological Symmetry, by arguing that this causal conception of powers is meant to apply equally to physical and mental powers. Even so-called 'higher faculties' can be governed by causal laws, insofar as they are relatively spontaneous. The upshot is that even if our inner states follow one another without immediate rhyme or reason, we should still expect them to be governed by causal laws because they arise from mental faculties governed by causal laws. But why then is there be a Problem of Lawfulness?

[81] Against this reading it may be objected that theoretical and practical reason are 'one and the same' faculty of reason because they share a 'common principle' (4:391). But it is impossible for the same principle to be both a law of nature governing a relatively spontaneous faculty and a law of freedom governing an absolutely spontaneous faculty. So either theoretical and pure practical reason are both relatively spontaneous, or both are absolutely spontaneous. That's a formidable objection, to which I have no good answer.

2.3 The Asymmetry Argument

Where there is power, there is substance. Physical powers are grounded in bodies and matter, as substances in space; mental powers are grounded in the 'mind' (*Gemüth*) or 'soul' (*Seele*), as substances (only) in time.[82] Being in time, these mental substances are not objects of metaphysics, but point to the empirical 'self' or 'I' as observed through inner sense (28:265). The soul, so understood, is the object of empirical psychology, which Kant also defines as '*doctrine of the soul* [*Seelenlehre*]' (A381). Frierson makes this point well: 'In studying this empirically accessible self, empirical psychology treats the mind/soul as a substance with various causal powers, each of which is governed by a distinctive (set of) causal law(s)' (2014: 49).

That's how things should work. But there is a problem. Within the context of Kant's transcendental philosophy, there can only be a substance in inner sense if we are able to synthesise the inner manifold by applying the category of substance; yet we can only apply the category of substance via its schema, which is persistence: 'The schema of substance is the persistence of the real in time' (A144/B183; see also B412–3).[83] Now here is the problem: the category of substance does not apply to inner sense because Kant thinks that nothing persists in inner sense. Our inner states are in a constant flux. Kant makes this point in numerous passages, like this one from the A Deduction:[84]

> The consciousness of oneself according to the determinations of our state in inner perception is merely empirical, changeable at any time, there can be *no standing or abiding self in this stream of inner appearances*, and is usually called the inner sense, or the empirical apperception. (A107; my emphasis)

[82] While Kant occasionally contrasts mind (*animus*) and soul (*anima*) – e.g., see (7:399), (12:32n), (R4550, 17:590), (25:247-8), (25:474) –, the two are often discussed interchangeably. Brook notes that 'Kant used these terms pretty much as extensional equivalents' (1994: 11). Faltudo (2014: 24-27, 120) notes subtle differences between these two terms, which, however, won't matter for my thesis. I too will use 'mind' and 'soul' interchangeably.

[83] See also (B224) and (R4045, 17:399).

[84] Other relevant passages are (A23/B37), (A90/B123), (B133), (B145), (A240-1/B291-2), (B275), (A347-8/B406-7), (A350), (A381), (B412–3), (B420) and (4:336).

So while Kant insists that the pure 'I' of apperception is identical in all representations (B133-4), he seems to side with Hume on the empirical 'I', suggesting that it is an elusive bundle of ever-changing states – and hence no substance.[85] But if there is no inner substance, there can also be no law-governed powers that underpin the inner states, and so inner sense cannot be governed the laws of these powers. Thus, inner sense is too chaotic to carry out any meaningful metacritical work. This line of reasoning, which threatens Ontological Symmetry and establishes the third premise of the Problem of Lawfulness, is summarised in the following Asymmetry Argument:[86]

Asymmetry Argument

(P1) Inner sense stands under universal laws only if states in inner sense result from law-governed powers.

(P2) States in inner sense result from law-governed powers only if there is a substance in inner sense.

(P3) There is a substance in inner sense only if something persists in inner sense.

(P4) Nothing persists in inner sense.

(C2) *Therefore*, inner sense does not stand under universal laws.

How to counter the Asymmetry Argument? One strategy is to tilt the textual scale. For example, Chignell (2017:142-8) insists that Kant *must* have been committed to inner substances for three reasons. First, Kant defines affection as a causal relation between outers object and a mind substance – indeed, this notion of affection fits well with the causal account of receptivity

[85] On this Humean interpretation of Kant, see, for example, Deleuze (1983/2008: viii), Guyer (1987: 283-4, 308), Friedman (1992a: 182), Brook (1994: 9, 77), Allison (2004: 292), Heßbrüggen-Walter (2004: 155-6), Dicker (2008: 81), Kraus (2016: 18), (2020: 154, 195, 207), Ferrarin (2015: 291), Bader (2017b), Longuenesse (2017: 109-10), and Rosefeldt (2017: 230). Or as Dyck puts it, 'the application condition for the category of substance (i.e., persistence) is not fulfilled in inner experience' (2014: 213; see also 64, 220).

[86] Versions of the Asymmetry Argument have been discussed by Gouaux (1972: 239), Washburn (1976), Friedman (1992a: 182-4), (2013), Nayak & Sotnak (1995), Klemme (1996), Ameriks (2000: ch. 4), Sturm (2001: 169-80), Westphal (2004: 232), Frierson (2014:22), and Kraus (2016), (2020: 6, 80, 211). Based on this argument, Ferrarin (2015: 291) concludes that reason, like all mental faculties, is a power without a substance and thus not part of the natural order.

developed in Section 2.2. Second, Kant thinks that there can be so-called 'inner experience'.[87] Yet experience (*Erfahrung*) in Kant's demanding sense requires the application of the categories, including the category of substance. And third, Kant repeatedly claims that there are inner substances, as he does, for example, in the A Paralogisms:[88]

> [I]n the connection of experience, matter, as substance in appearance, is actually given to the outer sense, *just as the thinking I, likewise as substance in appearance, is given to inner sense*; and in the connection of our outer as well as our inner perceptions, appearances on both sides must be connected among themselves into one experience according to the rules that this category [i.e., the category of substance; C.B.] establishes. (A379, my emphasis)

To further tilt the scale, one might also seek to reinterpret passages like A107, where Kant seems to deny the existence of an inner substance. For example, Chignell (2017: 140-1) contends that these passages merely establish that substances in inner sense – much like those in outer sense – can never be directly *given*. But inner substances, like outer ones, can still be *cognised*.[89] Dicker (2008: 98) takes a different approach, suggesting that it is a contingent empirical fact that inner sense does not exhibit sufficient order to apply the category of substance. Thus, there is no in principle asymmetry to outer substances.

These reinterpretations fare better with some passages, than other.[90] But even so, it is unclear whether they really tilt the scale because there are two heavy reasons that weigh in favour of the Asymmetry Argument. The

---

[87] For example, see (B277-9), (A672/B700), (7:141-2), (7:161-2), (17:43), and (R5453, 18:186). Kraus notes that '[t]he notion of inner experience is [...] a ubiquitous theme in the first *Critique*' (2020: 4). On 'inner experience', see also Frierson (2014: 5).

[88] Chignell (2017: 142-5) further identifies the following passages as relevant: (A34/B50-1), (B68-9), (B153-6), (A342/B400), (A370), (B415), (B427), (4:295), (4:335-6), and (29:982). In addition, Kraus (2020: 8) notes that (i) Kant claims that we can cognise our empirical selves – see (Bxl), (B139), (B155), (B158), (B277-9), (B400), (B431), (4:336), (7:142), and (7: 161); and (ii) that there are objects in the inner sense – see (A342/B400), (A357), (A368), (A385), (B403), (B415), (A443/B471), (A846/B874), (5:95), (4:467), (4:542), (7:142), and (R6313, 18:614). Moreover, Ameriks notes that Kant identifies the object of the inner sense with the soul (A342/B400), and that he identifies the soul with a simple substance – see (20:308), (28:590), (28:684), (28:691), (28:755), and (28:830).

[89] Frierson (2014: 24) explores a similar strategy.

[90] Passages like (B412-3) and (Rlxxx, 23:30) seem especially recalcitrant.

first reason derives from the First Paralogism. Paralogisms are syllogisms in rational psychology, which Kant claims are invalid because they equate the middle term (A402-3). Without getting too much into the weeds,[91] the First Paralogism aims to establish that 'a thinking being, considered merely as such, [...] exists only as [...] substance' (B410-1; see also A348; 12:32n). It does so via the following inference:

First Paralogism

(P1) What cannot be thought of other than a subject exists as a substance.

(P2) A thinking being cannot be thought of other than a subject.

(C2) *Therefore*, a thinking being exists as a substance.

Kant denies that this inference is valid. We cannot cognise the thinking being as substance. This critique has traditionally been thought to only afflict the 'I think' of apperception, as it is treated in rational psychology,[92] but recent readings suggest that the critique generalises to the 'I think' of inner sense, as it is treated in empirical psychology. For example, Rosefeldt (2017) has suggested that the First Paralogism equivocates the schematised with the unschematised category of substance. To cognise the thinking 'I' as substance, we would need to apply the schematised category to intuition, but this is impossible because (i) the pure 'I' is no intuition; (ii) and the empirical 'I' is an intuition but does not persist – it is impossible 'to intuit myself through inner sense as something that persists' (2017: 229).[93]

Rosefeldt's reading of the First Paralogism dovetails with a historical observation made by Dyck (2014). Dyck acknowledges that the primary target of Kant's critique are rationalist efforts to cognise the 'soul's substantiality [...] by means of reason alone', as has been tried by Descartes (2014: 90). But Descartes cannot have been the only target, Dyck argues. That's because

[91] For an in-depth discussion of the First Paralogism, see Kitcher (1982: 520-1), Powell (1990: 80), Stuhlmann-Laeisz (1990: 67), Brook (1994: 160), Ameriks (2000: 68), Allison (2004: 336-7), Bird (2006: 630-1), Rosefeldt (2017), and Proops (2021: ch. 3).

[92] See, for example, Ameriks (2000: 2), Frierson (2014: 23-4), Chignell (2017: 146-7), and Gomes (2017: 10).

[93] For a reading in the same spirit, see Kitcher (1990: 190), (2011: 193), and Westphal (2004: 232).

'since Wolff [the assumption had prevailed] [...] that the soul is given as an object of inner experience' (2014: 69). Thus, for his critique to have currency, Kant would also have had to deny any 'empirical basis for the cognition of the soul's substantiality' (2014: 90). We cannot cognise substances in inner sense.

The second reason weighing in favour of the Asymmetry Argument – specifically its fourth premise that nothing persists in inner sense – derives from Kant's 'Refutation of Idealism' (B274-9; and also 28:681). The Refutation is an ambitious argument in which Kant seeks to rectify the 'scandal of philosophy', namely 'to accept the existence of things outside us [...] merely on belief [*Glaube*], and [...] to be unable to offer any satisfactory proof' (Bxxxixn). The Refutation offers this missing proof of the existence of outer objects. The argument is ambitious, but for our purposes a toy version will suffice, which goes like this:[94]

Refutation of Idealism

(P1) I am conscious of my existence as determined in time.
(P2) I can only be conscious of my existence as determined in time if there is something that persists either in inner sense or in outer sense.
(P3) Nothing persists in inner sense.
(C1) *Therefore*, something persists in outer sense.
(P3) If something persists in outer sense, then there are objects outside me.
(C2) *Therefore*, there are objects outside me.

While the details of the Refutation remain controversial, I take there to be broad agreement that the proof requires the third premise. Nothing persists in inner sense and Kant is clear about this. In the second remark following the Refutation, he states that 'we have nothing at all persistent that we could underlie the concept of a substance, as an intuition, other than matter' (B278).

[94] For more through discussions of the Refutation, see Guyer (1987), Keller (1998: 210), Allison (2004: 300), Westphal (2004), Emundts (2006), Dicker (2008), (2011), (2012), Strum (2009: 249-50), Chignell (2010), (2017), Bader (2017b), Longuenesse (2017: 86-92), and Kraus (2020: 52).

But only spatial substances in outer sense are matter. So the argument of the Refutation succeeds only if the category of substance has no application to inner sense. Chignell (2017: 149-156) concedes this point, and accepts it as a problem for his interpretation.[95]

Note also that the Refutation, which was added in the B edition, motivated Kant to limit the scope of the First Analogy. In the A edition, the First Analogy states that '[a]ll appearances' – suggesting all outer *and* all inner appearances – 'contain the persistent (substance) as the object itself, and the changeable as its mere determination' (A182). However, in the margins of his personal copy of the A edition, Kant notes: 'Here the proof must be conducted so that it applies only to substances as phenomena of outer senses' (RlxxxE, 23:30).[96] This note reaffirms Kant's commitment to the claim that nothing can persist in inner sense.

What we find, then, is a deep ambivalence in Kant's text. On the one hand, there are passages, such as A379, in which Kant claims the existence of inner substances. And Kant is committed to this claim by his causal account of powers, which, according to Ontological Symmetry, applies to both physical and mental powers. On the other hand, however, Kant also claims that nothing persists in inner sense and so that we cannot apply the category of substance. This claim is further supported by Kant's discussion of the First Paralogism and the Refutation, which thereby motivate the Asymmetry Argument, specifically its fourth premise.[97]

How should we respond to this ambivalence? We shouldn't settle the issue by trying to further tilt the scale. We already have heavy textual weights

[95] That said, Chignell (2017) thinks that the problem must be solved by future interpretations – there must be a possible reading of the Refutation that does not rely on the third premise. But until we have such a future interpretation, I maintain that the problem persists. This also goes against Frierson, according to whom the Refutation does not imply 'that there can be no inner substances' (2014: 22n).

[96] Kraus correctly remarks that 'the First Analogy, which explicates the category of substance as persistent through time, cannot be applied so as to find something persistent in inner intuitions' (2020: 131). On this point, see also Friedman (1992a: 183-4), Ameriks (2000: 292-3), and Bader (2017b).

[97] Dyck remarks that there is an 'apparent inconsistency on Kant's part in making explicit causal claims with respect to inner appearances while asserting that [the] category [substance] only bears proper application to external appearances' (2014: 220). On this inconsistency, see also Ameriks (2000: 292-3) and Kraus (2020: 33).

on each side, and so the scale may just break from strain.[98] We also shouldn't try to time-slice the first *Critique* by distinguishing between the A and B edition.[99] While Kant's denial of inner substances features more prominently in the B edition, it is already present in the A edition. Instead, we should take the ambivalence serious, and explain how Kant can claim both things – that there are and aren't inner substances – *at the same time*. That's what I will do in Sections 2.4 and 2.5.

2.4 Substance Without Persistence

I resolve the textual tension by distinguishing different epistemic registers. My proposal is as follows: when Kant claims that there are no inner substances, he denies that we have *cognition* (*Erkenntnis*) of inner substances. This is because, without the schema of persistence, inner intuition cannot be determined according to the category of substance. By contrast, when Kant claims that there are inner substances, he allows for *knowledge* (*Wissen*) of an indeterminate number of inner substances; moreover, he also allows for the regulative assumption of a numerically determined single substance. This assumption – I argue in the Section 2.5 – amounts to so-called *doctrinal Belief* (*doktrinaler Glaube*).[100]

To see why we cannot have cognition of inner substances, we must first get clear on Kant's notion of a schema. What does a schema do for Kant? In his own words, a schema is a 'mediating representation' (A138/B177) that enables the 'application of the categor[ies] to intuition' (A138-9/B177-8). This works because the schema is 'a *rule for the determination* of [the unity of] our

[98] Having said this, Friedmann (1992a: 183-4), (2013: 316-24) and Kraus (2020: 155-7) rightly point to another reason for the Asymmetry Argument. In his proof of the First Law of Mechanics in the Metaphysical Foundations (4:541-2), Kant seems to say that only three-dimensional spatial objects can satisfy the conversion law required of substances; time, being one-dimensional, can't. However, see Sturm (2001: 174-8) and Chignell (2017: 144) for a response.

[99] As suggested by Bader (2017b).

[100] I follow Chignell (2007a, 2007b) in translating '*Glaube*' not as 'faith' but as capitalised 'Belief', to indicate that this is a technical term for Kant that has a broader sense than mere (religious) faith.

intuition in accordance with a [category]' (A141/B180; my emphasis).[101] Kant thus identifies a schema with a rule that allows us to apply a category to intuitions and, in doing so, to determine those intuitions in accordance with the category.

Without a schema, we can't have cognition. Cognition, according to Kant, requires two ingredients: intuition and concepts. Watkins & Willaschek put this point well: 'Cognition [...] arises when one determines an object given in intuition by applying a concept to it' (2017: 86).[102] So in order to cognise that an object a is F, we need two things: (i) a being given in intuition, and (ii) the concept 'F' being applied to the intuition, thereby determining a as being F. Since a schema provides the rule that allows us to apply a category to intuition, without a schema we cannot have cognition that accords with the category.

Given this understanding of cognition, I claim – pace Chignell (2017) – that we cannot cognise substances in inner sense. After all, Kant denies that something persists in inner sense. Yet persistence is the schema of the category of substance. Without the schema, however, we cannot apply the category of substance to inner intuition. But if we can't apply the category, we cannot cognise substances in inner sen se. And accordingly, we cannot *determine* inner intuition in accordance with the category of substance.

Put the other way around, inner intuitions are *indeterminate* with regard to the category of substance. But what does this mean? Kant explains this indeterminacy in the A Paralogisms. There, he notes that persistence guarantees 'numerical identity of an [...] object [...] in time' (A361-2). If an object a persists, I am able to identify and re-identify a at different times. Because nothing persists in inner sense, I am not able to (re)identify myself as an 'object of inner sense' (A362). Inner intuitions therefore leave it indeterminate whether inner states – i.e., representations – inhere in one substance or in

[101] As Kraus puts it, 'transcendental schamata determine the application condition under which a sensible manifold can be synthesized in accordance with the categories, and hence define procedural rules for installing the categories as the rules of synthesis for intuitions' (2020: 198).

[102] For example, see (A51/B76-77), (A92/B125), (B146), and (24:752).

many substances. Kant articulates this point in the 'elastic ball' metaphor:

> An elastic ball, which hits a similar one in a straight direction, communicates to it its whole movement, consequently its whole state (if one looks only at one place in space). If now, according to the analogy with similar bodies, *substances are assumed, one of which imparts representations to the other, together with its consciousness*, then a whole series of them can be imagined, the *first* of which communicates its state, together with its consciousness, to the *second*, this its own state together with that of the previous substance, to the *third*, and this just as the states of all previous ones, together with its own and its consciousness. The *last substance* would therefore be conscious of all states of the substances changed before it as its own, because those were transferred together with the consciousness into it, and in spite of that, it *would not have been just the same person* in all these states. (A363-4n; my emphasis)[103]

However, while Kant denies that inner sense gives us epistemic access to substances that are (numerically) determined, he leaves open the possibility that we have access to inner substances that are (numerically) *in*determined.[104] Indeed, this is plausible. Recall from Section 2.1 that inner states, understood as mental activities, analytically entail not only the existence of a power from which they result, but also the existence of a substance in which they inhere. However, just as we cannot analytically infer to any specific power – thanks to zombie powers –, we cannot analytically infer to any specific substance.

By way of analogy, the inference from inner states to inner substances mirrors the inference from appearances to things in themselves. It is often thought that the concept of an appearance contains the concept of something that appears, i.e., a thing in itself; and thus, that by observing appearance, we can analytically infer the existence of things in themselves.[105] But we can't

[103] For an in-depth discussion of the elastic ball metaphor, see Dyck (2014: 163).

[104] I am not the first to suggest this. For example, Longuenesse (2017: 86-92) argues that we can have an indeterminate perception of ourselves. Walker (2017: 212-4) suggests that we can make indeterminate claims about inner substances by using the unschematised category of substance. Kraus claims in a similar spirit that, 'textually, none of the discussed passages explicitly excludes the in-principle applicability of the category of substance to inner intuition; rather, the passages leave open the option of applying the concept of substance in a less demanding sense than substantial persistence' (2020: 158).

[105] Such a reading has, for example, been advanced by Chignell (2007a) and Watkins & Willaschek (2020).

say anything determinate about things in themselves, especially not about their numerical identity, because they are only a 'transcendental object $= x$' (A109). Similarly, then, by observing inner states, we can analytically infer the existence of a substance in which they inhere, but only as a numerically indeterminate 'thing that thinks, [...] $= x$' (A349/B404).[106,107]

I claim that these analytic inferences establish knowledge, in Kant's technical sense. Kant defines knowledge as an assent (*Fürwahrhalten*) that rests on 'a ground of cognition that is both objectively and subjectively sufficient' (9:70).[108] While subjective grounds depend on the constitution of the subject, objective grounds rest on 'the constitution of the object' (A821/B849).[109] So understood, objective grounds are truth conducive.[110] Moreover, an objective ground for p counts as sufficient iff it guarantees that p is true, and so that $\neg p$ is false: 'the sufficient [objective] ground is that whose opposite cannot possibly be thought [...] to be true' (24:145).[111] More on Kant's notion of objective grounds in Chapter 5.

Importantly, then, knowledge – unlike cognition – does not, qua its definition, require that we determine an intuition by a concept. All we need is that the ground of cognition guarantees truth. Now, Kant would think that we can empirically know that there are states in inner sense; after all, inner

[106] While the quote only references the pure 'I' of apperception, I maintain that Kant's point also holds for the empirical 'I' of inner sense.

[107] This is not to say, however, that the empirical 'I' of inner sense is a thing in itself; after all, it is in time. That said, it is unclear how the 'I' of inner sense can be empirical. After all, it isn't the object of any possible experience (understood as empirical cognition). In fact, Kant clearly states that soul is meant to be a 'simple substance' but that 'the simple cannot come forth in any experience' (A771-2/B799-800).

[108] See also (A822/B850), (8: 141), (9:66), and (R2450, 16: 374).

[109] In the *Blomberg Logic*, Kant likewise notes that objective grounds 'rest in the thing itself [*in der Sache selbst*]' (24:198). See also (A820/B848), (R2450, 16:374), (R2459, 16:378), as well as Chignell (2007a: 326), (2007b: 39).

[110] Objective grounds 'determine truth' (R2450, 16:374). See also See also Gava (2019: 55), (2023), Chignell (2021: 102), Hebbler (2022: 736), and Cooper (2023: 7).

[111] Chignell (2007a, 2007b, 2021) and Cohen (2021) have defended a fallibilist reading of sufficient objective grounds: objective grounds for p can be sufficient and yet p is false. However, see my Benzenberg (ms-d) for why such a reading is textually unattractive, if not impossible. Indeed, most Kant scholars claim an infallibilist reading of sufficient objective grounds. See Mattey (1986: 435), van der Schaar (2003: 9), Willaschek (2010: 189), Höwing (2016: 209-1), (2017: 115), Huber (2018: 659), Gava (2019: 56), Willaschek & Watkins (2020: 3207), Kern (2021: 120), and Kohl (forthcoming).

states are what we see when we introspect.[112] Moreover, analytic judgements, which can never be cognition because they only involve concepts, do amount to knowledge, because the conceptual relations guarantee truth.[113] And so we can know that inner states conceptually presuppose substances. Put together, these considerations motivate the following 'Knowledge Argument' (with 'K' standing for Kant's notion of knowledge):

Knowledge Argument

(P1) K(There are states in inner sense).

(P2) K(If there are states in inner sense, then there are inner substances).

(C) *Therefore*, K(There are inner substances).

By distinguishing between cognition and knowledge, we are therefore able to resolve the textual tension observed in Section 2.3. When Kant claims that there are no inner substances, he is only denying that we have cognition of inner substances. We must therefore also revise the third premise of the Asymmetry Argument: there can be inner substances even if nothing persists in inner sense, it is just that we cannot cognise them. By contrast, then, Kant's assertion that there are substances in inner sense is advanced in the voice of knowledge. The 'Knowledge Argument' shows that we have sufficient objective grounds for the existence of inner substances.

This saves Ontological Symmetry. Just as there are substances in outer sense that ground physical powers, there are substances in inner sense that ground mental powers; indeed, we know that there are such inner substances. I thus agree with Frierson, who notes that '[Kant] is clear' – or better: knows – 'that there is some substance that underlies the changes observed in inner sense, and whatever this substance should turn out to be, it is sufficient to

[112] Kant notes that experience can licence empirical certainty. See (A836/B864), (R2454, 16:375-6), (R2455, 16:377), (R2457, 16:377), (R2459, 16:379), (R2465, 16:382), (R2485, 16:389), (24:231), (24:438), (24:544, (24:637), (24:639), (24:734-5), (24:857), and (24:891-3). Fallibilists, like Chignell (2007a), (2007b), (2021), have argued that empirical knowledge becomes impossible once objective grounds must guarantee truth, and so that empirical certainty itself is fallible. However, see my Benzenberg (ms-d) for a response to this objection.

[113] On this point, see also Chignell (2007a), (2007b), Watkins & Willaschek (2020), Wang (2023), and Kohl (forthcoming).

justify the ascription of causal laws to psychological states' (2014: 26). The only problem is that we don't know whether 'this substance' is one substance or many substances. But Kant has a solution for that problem too, or so I argue in the next Section.

2.5 Belief in Myself

In the B Preface, Kant proclaims that he 'had to remove knowledge to make room for Belief' (Bxxx). While this dictum has long been read as being limited to moral Belief in God and immortality, recent scholarship has sought to extend it to other species of Belief, especially to what Kant calls '*doctrinal Belief*' (A825/B853).[114] This section advances this broader trend. I argue that, while we don't know that there is a numerically identical substance in inner sense, we are justified in assuming such as substance as doctrinal Belief. My argument in this section draws extensively on Kant's discussion of Reason's regulative use in the Appendix to the Transcendental Dialectic.

But first: what is doctrinal Belief for Kant? Doctrinal Belief, like all Belief, is an assent from grounds that are objectively insufficient, yet subjectively sufficient (A822/B850). The grounds are objectively insufficient in that they don't guarantee truth; but they are subjectively sufficient in that they nonetheless justify the assent. This justification, I suggest, is to be construed in analogy to the justification of moral Belief. While moral Belief is justified as a 'hypothetically necessary' means 'for the realisation' of the final end of practical reason (A823/B851), doctrinal Belief is justified as a hypothetically necessary means to realise the final end of theoretical reason or Reason. Put as a strict definition:

[114] For a discussion of doctrinal Belief, see Beck, (1960: 266n18), Heimsoeth (1971: 781-5), Chignell (2007a), (2007b), (2009a), (2009b), (2023), Pasternack (2011a), (2011b), Pickering (2016), Stang (2016: ch. 9), Gava (2018), (2022), (2023), Yuen (2022), Schafer (2023: ch. 5), Benzenberg & Chignell (2024), as well as my Benzenberg (ms-a), (ms-c), (ms-f), (ms-g).

Definition of Doctrinal Belief

S doctrinally believes that p iff

(i) S does not have sufficient objective grounds that p;

(ii) S does not have sufficient objective grounds that $\neg p$;

(iii) S pursues the final end of theoretical reason, e; and

(iv) S can only realise e by assuming that p.

Doctrinal Belief, according to this definition, is justified not by evidence but by its instrumental use. And so, although it does not amount to knowledge, it is more than mere wishful thinking because it is a necessary means to the final end of Reason. Doctrinal Belief is a form of rational assent. I will now argue that Kant accepts the following claim as doctrinal Belief: *that* all states in inner sense inhere in a numerically identical substance. We already know from Section 2.4 that this assertion satisfies conditions (i) and (ii) of the definition in that it falls short of knowledge. In what follows, I will show that it also satisfies conditions (iii) and (iv).

First, then, what is the final end of Reason?[115] Reason is defined as the faculty that seeks to find the *totality of conditions* 'to the conditioned cognition of the understanding' (A307/B364). While finding such totalities must hence be the final end of Reason, this is not the full story. Throughout the Appendix, Kant also argues that Reason seeks to complete 'the *systematic unity* of the cognitions of the understanding' (A647/B675; my emphasis).[116] However, this is not a second, separate end of Reason, but – as I will show in Chapter 4 – Kant thinks that the two ends are meaningfully identical; Kant identifies *complete* and *systematic* explanation.

In Chapter 4, I will also say more about the structure of this systematic unity. But to foreshadow the main points: Kant defines the systematic unity as

[115] With that I mean the final end of theoretical reason in the narrow sense. In the Canon of Pure Reason, Kant specifies that the final end of all uses of reason, including pure practical reason, is the highest good (A797-819/B825-47).

[116] Kant further specifies: 'that a certain systematic unity of all possible empirical concepts [...] must be sought; is a [...] *logical principle, without which no use of reason would take place*' (A652/B680; my emphasis). In Chapter 4, I will argue that the logical principle is identical to the Principle of Reason.

an all-encompassing taxonomy of empirical concepts and laws that is organised according to genera and species, with more general concepts and laws at the top, and more specific ones towards the bottom. In seeking to complete such a system, Reason aims to unify all concepts and laws under a highest genus, from which all other species can be derived (A642-668/B670-696).

One more thing. By systematising empirical laws, Reason effectively systematises the powers of nature that are governed by those laws. Kant makes this point explicit by stating that Reason seeks to complete 'the systematic unity of the manifold powers' (A650/B678). The highest genus among these powers is then called the '*basic power* [*Grundkraft*]' (A649/B677). Importantly, Reason systematises not only physical powers of extended bodies, but also the 'manifold powers [...] of the human mind' (A648/B676). Kant now claims we can only realise the systematic unity of mental powers – which is part of the final end of Reason – if we assume that there is a simple, numerically-identical substance, the soul, that grounds these powers:

> If I want to seek out the properties with which a thinking thing exists in itself, then I have to ask experience, and *I cannot even apply any of the categories to this object* except insofar as its schema is given in sensible intuition. By this means, however, I never arrive at a systematic unity of all appearances of the inner sense. So instead of the concept of experience, [...] reason takes the concept [...] (idea) of a *simple substance*, which in itself is *unchangeable* (*personally identical*), in community with other real things outside of it [...]. Here, however, [reason] has nothing else in sight than *principles of systematic unity in explanation of the appearances of the soul*, namely: to consider all determinations as in *one single subject*, all powers, as much as possible, as derived from one single basic power, all change, as belonging to the states of *one and the same persistent being* [...]. Nothing but advantage can arise from such a psychological idea [...]; thus the consideration of this *object of inner sense* is undertaken [...] and the investigation of reason is directed towards bringing the grounds of explanation in this subject, as far as it is possible, to a single principle; all of which is best – *indeed only and solely* – effected by such a schema[.] (A682-4/B710-2; my emphasis)

Let's unpack this passage. Since we cannot apply the category of substance to a thinking being without the sensible schema of persistence, Kant suggests that we must use the idea of the soul as an 'object of inner sense' that is 'personally identical'. Note that Kant takes his notion of personality from Locke and Wolff: a thinking being has personality iff it 'is conscious of the *numerical identity* of itself at different times' (A361; my emphasis).[117] Accordingly, Kant also thinks that we must assume that all psychological change belongs 'to the states of one and the same *persistent* being'. And because the soul is assumed as an inner object that persists, Kant also thinks that the idea plays the role of 'a schema'; more on this shortly.

This numerically identical inner substance, however, must not be assumed on a whim. The passage specifies that the assumption is necessary to realise Reason's final end of systematic unity. By assuming a single inner substance, Reason 'has *nothing else in sight* than principles of systematic unity in explanation of the appearances of the soul'. And indeed, we can achieve such a systematic explanation '*only and solely*' via this idea. A few pages earlier, Kant also notes that the idea of the soul, like all ideas of Reason, has an '*indispensably necessary* regulative use, namely that of directing the understanding to a certain end', namely the end of a systematic unity of all its cognition (A644/B672; my emphasis).

The textual evidence thus suggests that conditions (iii) and (iv) of the definition are satisfied. Condition (iii) is satisfied insofar as we pursue Reasons final end of completing the systematic unity of all natural powers, including our mental powers. Condition (iv) is satisfied because Kant claims that we can only realise this unity if we assume that there is a single numerically identical substance in inner sense. But this raises one big question: why

[117] Locke defines a person is 'the same thinking thing in different times and places' (1689/1975: II.xxvii.9). Wolff states that 'a thing is called a person that is conscious that it is the very same thing that it was previously in this or that state' (1751/1983: §924). For the historical background to Kant's notion of personality, see Dyck (2014: ch. 5). Locke's and Wolff's definition correspond to Kant's notion of psychological personality: 'personality can be taken practically and psychologically; practically, if free actions are ascribed to it; psychologically if it is conscious of itself and of the identity' (28:296). On Kant's notion of (psychological) personality, see also (A365) and (B408), as well as Dyck (2014: 141, 162), Kraus (2020: 162-5), and Brennan (ms).

exactly should we assume that such an assumption is necessary to realise Reason's final end? After all, we don't have to assume a single physical substance to realise the systematic unity of physical powers. Wouldn't we likewise be able to systematise mental powers if there were multiple inner substance – as pictured in the 'elastic ball' metaphor?

This question is a difficult one. To see why, let us consider two salient answers. First, one might argue that we must assume a single inner substance because the final end of Reason demands that we systematise *mental* powers as if they arise from a single inner substance. Once the single inner substance is baked into the end, it is trivial that we must assume such a substance as a necessary means. But why should Reason set such a specific end? After all, the final end of Reason does not demand that we systematise *physical* powers as if they arise from a single outer substance. Why would the end of reason be any different for mental powers?

Second, one might argue that we must assume a single inner substance because, if there were multiple inner substances, there couldn't be a basic *mental* power governing all inner states – a highest unifying power presupposes a single substance. Yet Reason seeks such a basic mental power as part of its final end. But again, this answer can't be right. After all, Reason also seeks a basic physical power as part of its final end. However, it would seem that there could be such a basic *physical* power even though there are several outer substances. I don't see why this should be any different for mental powers.[118]

Here then is my positive proposal: I suggest that we can only systematically explain the states we observe in inner sense if we have a determinate mapping from inner states to substances; it cannot be left indeterminate what state belongs to what substance. Moreover, I argue that – lacking the schema

[118] There are two ways to understand Reason's quest for a basic power. Either Reason tries to find a basic power for each substance, or it tries to find the most basic *type* of power across all substances. I think Reason goes for the latter. After all, Reason systematises those empirical laws by which we individuate types of powers. Particular *tokens* of power would additionally need to be grounded in specific substances. More on this later. Note also that Kant claims that we must presuppose the soul's simplicity, not numerical identity, to satisfy the quest for basic powers (A771-2/B799-800). See also (20:308), (28:590, (28:684), (28:691), (28:755), and (28:830).

of persistence – we only have such a determinate mapping from inner states to substances if we assume that all inner states inhere in the same substance. We must therefore assume that there is only one single inner substance. I call this the 'Mapping Argument', which can be stated more precisely as follows:

Mapping Argument

(P1) We can only systematically explain the states we observe in inner sense if we have a determinate mapping from inner states to substances.

(P2) We only have a determinate mapping from inner states to substances if we assume that all inner states inhere in an identical inner substance.

(C) *Therefore*, we can only systematically explain the phenomena of inner sense if we assume that all inner states inhere in an identical inner substance.

I take both premises in turn. The first premise can be motivated by reconsidering the example of a modus ponens inference from Section 2.2. There, I suggested that we can think of a syllogism as a series of mental states: (i) I represent 'if p, then q' at time t_1; (ii) I represent 'p' at time t_2; and (iii) I represent 'q' at time t_3. We may explain this series of inner by assuming a mental power that allows us to draw modus ponens inferences. This power takes (i) and (ii) as its causal condition c, and manifest (iii) as its activity φ. But now consider a revised example where we don't know which object has these representations:

(i*) Some x represents 'if p, then q' at time t_1.

(ii*) Some x represents 'p' at time t_2.

(iii*) Some x represents 'q' at time t_3.

This revised example better captures our situation as we observe states in inner sense. Lacking the schema of persistence, we cannot identify and re-identify any inner substances. And so, while we know that inner states must inhere in some substance $= x$, we don't know in which substance they inhere, and especially not whether they all inhere in the same substance. In this situation, however, we cannot explain the sequence of mental states

simply by referring to a power of modus ponens inference. After all, it could be the case that each of the representations inheres in a different substance, in which case there would be no substance with the power of modus ponens inference.[119]

In light of this example, it seems that Reason cannot (systematically) explain inner states without being able to tell which state belongs to which substance. The explananda themselves would become elusive. One might object, however, that we also don't have a determinate mapping from inner states or activities to their mental powers – owing to zombie powers. But this isn't an insurmountable problem because Reason can still explain a given activity by assuming a (set of) mental power(s) as a probable hypothesis. By the same token, it would seem that Reason could also work with a probable mapping from inner states to substances, not a determinate one.[120]

While this objection is philosophically plausible, is textually mistaken. True, Kant thinks that we (causally) explain an observed state by assuming a power to-φ-when-c and an antecedent condition c as hypothesis: 'Hypotheses [...] serve to [...] explain something' (24:220).[121] Moreover, it is true that hypotheses are only a matter of 'probability, but not of certainty' (24:888).[122] Yet while the hypothesised explanation is merely probable, Kant insists that the explanandum itself must be 'completely certain' (24:888).[123] Kant seems

[119] An important assumption here is that a power cannot be grounded in multiple substance. This follows from two commitments: (i) that states can only inhere in one substance; and (ii) that states result from the exercise of a power (see Section 2.1). Having said that, it might be argued that states (i*) and (ii*) could inhere in one substance, S_1, and state (iii*) in another substance S_2. In this case we could say that substance S_2 has the power of podus ponens inference, but then this power would become a receptive power because it is triggered by states of another substance. But even if this were possible – and I am here skirting the complicated question of whether, for Kant, minds can directly affect other minds –, it would follow that we cannot say which of our powers are receptive and which are relatively spontaneous. That too would be a problem.

[120] Thanks to Maya Krishnan for pushing me on this point.

[121] On Kant's notion of explanation and the link to hypotheses, see also (A83/B115), (A635/B663), (A770/B798), (A772-3/B800-1), (A789-91/B817-9), (A798/B827), (8:53-4), (9:84-6), (16:464, R2676), (21:346), (24:223-4), (24:558-9), (24:647), (24:746-7), and (24:887).

[122] Kant also states or implies that hypotheses are only a matter of probability in (A649/B677), (A775/B803), (1:277), (2:149), (5:465-6), (8:311-2), (16:464, R2676), (24:439-40), (24:528), (24:557-9), (28:605), (24:647), (24:746-7), (24:886-9), (28:1285-6), (29:51), and (29:103).

[123] See also (A770/B798), (A822/B850), (5:465-6), (R2680, 16:466-9), (24:647), and (24:746-7). We find the same commitment in Arnauld's *Logique de Port-Royal* (1660/1854: pt. 4, chs. 15-6). In Reflexion R2676, Kant states: 'The hypothesis must be more probable than

to be concerned that hypotheses, which are themselves epistemically wobbly, should at least rest on a firm epistemic ground. But for this we need a determinate mapping. (I'll have more to say about all this in Chapter 5.)

Onto the second premise of the Mapping Argument. Why must we assume a single inner substance to have a determinate mapping? Why couldn't we, for example, alternatively assume the idea of a bi-soul, understood as the idea of two inner substances? Generally speaking, a bi-soul would be problematic because we still couldn't tell which state belongs to which substance; for, lacking the schema of persistence, any state in inner sense could belong to any of the two substances. It is only if we assume that there is one single inner substance grounding all states that we can tell what state belongs to what substance.

By way of analogy, picture yourself at the end of term when your students hand in their anonymous teaching evaluations, with each student submitting at least one evaluation. So you have a class of students, S_1, S_2, ..., S_n, and their evaluations, E_1, E_2, ..., E_m ($m \geq n$). As long as you have two or more students ($n \geq 2$), you cannot tell which student submitted which evaluation. However, once you have only one student in the class – a common occurrence in the tutorial system –, it doesn't matter how many evaluations they submit, you can tell for certain that all submitted evaluations are theirs. So if $n = 1$, and only if $n = 1$, is there a determinate mapping from evaluations to students.

The same applies to substances, S_1, S_2, ..., S_n, and inner states or representations, R_1, R_2, ..., R_m ($m \geq n$). Only if $n = 1$, do we have a determinate mapping from representations to substances. This idea can be made precise. A mapping is a surjective function from the set of representations to the set of substances – each representation inheres in exactly one substance, and each substance has at least one representation. As long as $n \geq 2$, more than

---

and more clearly [deutlicher] cognised than the explained' (16:464). Taken on face value, this would suggest that the explanandum itself is also only a matter of probability. However, I suggest that the 'more probable' was a slip of pen, and that the main emphasis is on 'more clearly' as is suggested by Kant's own example: 'e.g., the freezing [must not be explained] out of a cold-making matter' (16:464, R2676). See also (A772/B800) and (8:53-4).

one such function is possible; and since we lack the schema of persistence, we don't know which one is actual. However, once $n = 1$, only one function is possible, which then also has to be actual.[124] So in the limit, as $n = 1$, we can we tell which representation inheres in which substance, even though we lack the schema of persistence.

The premises of the Mapping Argument are thus well supported. My central claim is that reason, in its search for systematic explanation, needs a firm starting point. The explanandum must be certain. But without the schema of persistence, we can only get such a firm starting point in the inner sense by assuming that all inner states belong to the same substance. The assumption of a numerically identical single inner substance therefore turns out to be a necessary means of realising Reason's final end of systematically explaining inner states in terms of mental powers. Thus, although we lack epistemic warrant for a single inner substance, we do have subjective-practical warrant that licences doctrinal Belief.

Against such a reading, Kraus has recently objected that doctrinal Belief in a single inner substance 'is too weak [...][,] [f]or a belief, while subjectively warranted, could still turn out to be objectively false' (2020: 189). But this objection is misguided. Not only does doctrinal belief carry with it the highest degree of psychological conviction[125] – we would stake the 'fortunes of the entire life' on its truth (A825/B853) – but, like all Belief, doctrinal Belief gives so-called 'practical certainty', which at a minimum requires that we cannot be proven wrong (24:200).[126] And indeed, lacking the schema of persistence,

[124] Here is an informal proof. Let $\mathcal{R}$ be the set of representations, and $\mathcal{S}$ be the set of inner substances. If $|\mathcal{R}| = m$ and $|\mathcal{S}| = n$ and $m, n > 0$, we must show that the number of surjective functions from $\mathcal{R}$ to $\mathcal{S}$ is 1 iff $n = 1$. The number of such surjective functions can be calculated by multiplying $n!$ with the Stirling number of the second kind, $S(m, n)$, i.e., $k!S(m, n)$. It's easy to see why $n!S(m, n) = 1$ *iff* $n = 1$: (i) $n!S(m, n) = 1$ *if* $n = 1$ because $1! = 1$ and the Stirling number of the second kind is always 1 if $n = 1$, no matter the value of m; and (ii) $n!S(m, n) = 1$ *only if* $n = 1$ because, if $n > 1$, then $n! > 1$, and the Stirling number of the second kind, $S(n, m)$, is always at least 1.

[125] On Kant's notion of degrees of conviction, see my Benzenberg (forthcoming-b) and Silva Jr. (forthcoming).

[126] While practical certainty is not based on truth-entailing grounds, it nonetheless a species of certainty in the broader sense of being based on irrefutable grounds: 'With certainty, it is not at all possible to assume the opposite' (R2468, 16:383). Moreover, the practical certainty of Belief is stronger than the certainty of knowledge in that we don't even listen to opposing grounds: 'Practical Belief is often stronger than all knowledge. With the latter, you can still

we could never come to know or cognise that there is more than one inner substance (A771-2/B799-800).

Turning the tables, I maintain that Kraus misconstrues the way in which the idea of the soul, like all ideas of Reason, can be 'an analogue of a scheme of sensibility' (A665/693).[127] Kraus correctly notes that 'the idea [of the soul] substitutes for [...] the schema of persistence' (2020: 13). But on her reading, 'presupposing the idea of the soul amounts, not to making theoretical claims about what reality is like, but to generating a context of intelligibility' (2020: 172).[128] I disagree. The idea of the soul, used regulatively, does make a claim about what reality, specifically inner reality, is like. The claim is this: *that* all inner states inhere in the same numerically-identical substance.

I do agree with Kraus (2020: 210), however, that the idea of the soul should provide a rule for determination; after all, Kant defines a schema as 'a *rule for the determination* of [the unity of] our intuition' (A141/B180; my emphasis). But on my reading, the idea of the soul provides a very simple rule, namely: *to* proceed in systematic explanation as if all inner states inhered in the same numerically identical substance. Only by following this rule, I have argued, do we have a determinate mapping of inner states to substances. To be clear, however, this determinate mapping does not allow us to apply the category of substance, and so the 'schema of reason is *not likewise a cognition*' (A665/693; my emphasis).

In conclusion, while we cannot cognise our inner 'I', we do know that there must be an 'I' and we have Belief that this 'I' is single. And so, while the Asymmetry Argument works at the level of cognition, we can maintain Ontological Symmetry at the level of both knowledge and Belief. We thus have good reason to think that mental powers are like physical powers, not only in that they are grounded in a substance, but also in that they are

---

hear grounds to the contrary, but not with practical Belief' (24:543). See also (24:634) and (24:747).

[127] See also (A655/B693), (A681/B709), (A670/B698), (A674/B702), (A682-4/B710-2), as well as Kraus (2020: 11-3, 189-216).

[128] Other interpretations of the regulative use of the idea of the soul can be found in Strawson (1966/2005: 100), Klemme (1996), Ameriks (2000: 293), Heßbrüggen-Walter (2004), Serck-Hanssen (2011), Goldman (2012), Dyck (2014: ch. 7), Wuerth (2014), Faltudo (2014: 24-33), Frierson (2014: 25), Chulanon (2022), Gava (2023), and Kitcher (ms).

governed by universal laws. Since mental powers underpin inner states, we should further expect inner sense to be governed by these universal laws, disproving the third premises of the Problem of Lawfulness.

3

Reason Never Fails

This chapter turns to the Problem of Normativity. The problem says that we cannot come to know Reason via empirical psychology because the Principle of Reason is normative, but the laws of inner sense are descriptive. Indeed, Kant states that Reason requires us '*to* find' the totality of conditions (A307/B364), thus making 'demands' on us (A656/B684). The causal laws that govern inner sense, however, 'only indicate how we *do* judge, but do not command how we *ought* to judge' (5:278; my emphasis). The problem points to a broader concern that threatens Ontological Symmetry: (higher) mental faculties cannot be governed by causal laws (of nature), because such laws are descriptive, whereas the laws of (higher) faculties are normative.

I resolve the Problem of Normativity by arguing that the Principle of Reason is (also) descriptive, and not (only) normative. Here is the structure: First, I introduce Kant's principle of 'ought'-excludes-'must', i.e., the principle that any demand to φ requires the ability to $\neg\varphi$ (§ 3.1). I then ask how we are able to violate the demands of Reason (§ 3.2). Against the reading that Reason itself violates the demands of Reason, I offer a new reconstruction of the sources of dialectical error, arguing that Reason never fails on its own (§§ 3.3 and 3.4). Instead, we are only able to violate the demands of Reason because of the corrupting influence of sensibility (§3.5). The Principle of Reason, then, is descriptive of Reason, but normative for us.

3.1 'Ought'-Excludes-'Must'

In the last two decades, there has been a heated debate among Kant scholars about the normativity of logic. The main question has been whether the laws of logic – above all the Principle of Non-Contradiction (PNC) – are normative for or constitutive (or normative-constitutive) of thought. Is it possible to think contradictory thoughts? For example, can I think that the ball is blue all over and red all over? Normative readings say *yes*, because the PNC only tells us that we ought to think consistently;[129] constitutive readings say *no*, because the PNC is constitutive of all thought, and therefore representations that violate the PNC aren't bad thoughts, but no thoughts at all.[130]

The debate has been heated because different passages pull in different directions.[131] But the specifics of the debate need not concern us here because my focus is on Reason and its principle, and not the PNC as the principle of the understanding. That said, the debate on the laws of logic has helped to clarify general aspects of Kant's theory of normativity that will be helpful. Tolley (2006) has noted that laws, for Kant, can only be normative if they satisfy the following condition: 'The "subjects" of a law – those beings which are governed by, or subjected to, the law – must both be able to *succeed* and be able to *fail* to act (or be) in accordance with the law' (375; my emphasis).

Entailed in this condition are two principles, which both enjoy robust textual support. The first principle is Kant's famous commitment to 'ought'-implies-'can' (OIC), which he unequivocally articulates for the moral law: 'if the moral law commands that we ought to [φ], it inevitably follows that we must be able to [φ]' (6:50; my substitutions).[132] The second principle –

[129] For a normative reading, see MacFarlane (2000: 86-7), Stang (2014: 220), Leech (2017b: 352), Lu-Adler (2017: 217-8), (2018: 72-7), Pollok (2017: 8), and Boyle (2020: 128).

[130] For a constitutive reading, see Conant (1992), Tolley (2006), Merritt (2015), Newton (2020), as well as the excellent paper from Martens (ms).

[131] On the one hand, Kant insists that thought is 'bound in its action' to the laws of logic (9:11) and that without these laws 'no use of the understanding takes place' (A52/ B76). On the other hand, however, he also claims that the laws of logic describe 'not how we do think, but how we ought to think' (9:14). See also (A59/B84), (9:51), and (24:826).

[132] See also (A548/B576), (A807/B835), (5:30), (5:142), (5:143n), (6:47), (6:50), (6:62), (6:64), (6:380), and (8:276-9). Due to its presence in Kant's practical philosophy, OIC is also referred to as 'Kant's dictum' (Frankena 1958: 59). For a further discussion of Kant's dictum, see Stern (2004: 53), Kleist (2005), Haji (2012: 3, 24), Rödl (2013), and Kohl (2015: 690). An epistemic

which I will call 'ought'-excludes-'must' (OEM) – is lesser known, but will be my focus. OEM states that a law is normative only on the condition that it can be violated. Kant articulates this principle in the *Groundwork*, again with reference to the moral law:

> The representation of an objective principle, insofar as it is necessitating for a will, is called a command (of reason), and the formula of the command is called an imperative. All imperatives are expressed by an '*ought*' and thus indicate the relationship between an objective law of reason and a will that *is not necessarily determined by* [*that law*] in its subjective constitution (a necessitation). [The imperatives] say that it would be good to do or refrain from doing something, but they say it to a will that does not always do something just because it is represented to it that it would be good to do so. (4:413; my emphasis)[133]

Kant illustrates OEM by contrasting our finite will with God's holy will. Finite willers, like us, can fail to obey the moral law because the subjective grounds of action – our incentives or motives – don't necessarily agree with the objective grounds of the moral law. The moral law therefore takes the form of an imperative for us. But holy willers, like God, necessarily abide by the moral law because, '[w]ith God, the objective practical law is at the same time motive' and so the subjective ground of action (29:605). Kant concludes that 'no imperatives apply to the divine and, in general, to a holy will; the ought is in the wrong place here' (4:414; see also 6:226). Yet OEM should not only apply to the moral law, but also to other laws like the PNC and the Principle of Reason; it can be stated more generally as follows:[134]

'Ought'-Excludes-'Must'

(OEM) A law L is normative for x only if x can fail to abide by L.

---

version of OIC in Kant's philosophy is discussed by Wood (1970: 25-32), Zeldin (1971: 46), Reath (1988: 609), and Willaschek (2016).

[133] See also (4:499) and (4:455).

[134] On Kant's commitment to OEM, see also Lavin (2004: 425), Pollok (2017: 5), Kahn (2019: 2-3), and Marten (ms).

Formulated in this way, OEM highlights a central feature of Kant's theory of normativity. Laws are not normative *per se*, but they are only ever normative *for some* x. We must therefore always ask, for *whom* or *what* is a law normative? Importantly, then, one and the same law, L, can be normative for some agents, while being descriptive or, indeed, constitutive of other agents. We have already seen this with regard to the moral law, which can be simultaneously normative for some beings, finite willers like us, and descriptive for other beings, holy willers like God. That's because we can fail to abide by the moral law, but God can't.

Yet the status of laws can differ not only for different agents. Drawing on my discussion in Section 2.1, we can ask more generally whether a law, L, is normative for a substance (including agential substances), a power, or an activity. Indeed, many – though not all – normative readings of the laws of logic allow the PNC to be constitutive of a *power* or faculty, the understanding, while insisting that it is normative for an *activity*, thought, or the agential *substance* of that activity, finite thinkers.[135,136] In the same vein, the moral law is often taken to be constitutive of practical reason but normative for finite agents.[137] And Schafer articulates the same idea more generally for reason and its principle:

> Thus when we consider reason in the context of a sensibly conditioned finite subject, the principle of reason as a faculty will always be *both constitutive* (of reason's exercise insofar as it is free of illicit sensible influences) *and normative* (for the subject's thoughts and actions given that they are subject to such influences). (2019: 187; my emphasis)[138]

[135] For a reading in this spirit, see Stang (2014: 220), Lu-Adler (2017: 217-218), (2018: 72), and Boyle (2020: 128). These readings are further discussed by Martens (ms).

[136] Lu-Adler observes that norms that apply to activities are merely 'evaluative', whereas norms that apply to agents are 'imperatival' (2017: 1). Still, both are norms and satisfy OEM; as such they are not descriptions.

[137] For example, Sensen argues that '[t]he [moral] law is constitutive of [practical] reason' (2013: 78).

[138] Pollok suggests that 'Kant's principles of reason, generally speaking, are [...] constitutive of certain judgments, and they are normative for finite reasoners using their cognitive faculties – reason, understanding, and the power of judgment – to make those judgements' (2017: 9-10; my emphasis).

The general idea, then, is this: the law, L, of a faculty or power, P, is necessarily constitutive of P, but can be normative for (the activities of) a finite subject, S, who has P. This idea has merit. For one thing, it fits well with the account of faculty individuation I defended in Section 2.1 – if faculties are individuated by their laws, they cannot fail to abide by those laws without ceasing to be the faculty that they are. By contrast, finite subjects are under the corrupting influence of sensibility, which can explain why they can fail (more on this in Section 3.2). The idea also provides a template for a global theory of normativity, which might work for PNC and understanding, as well as the moral law and practical reason.[139]

And, I will argue, the template also works for Reason and the Principle of Reason. Of course, we as finite reasoners can fail to abide by the Principle of Reason, which we do, for example, when we assert rather than seek the totality of conditions. That's why we *ought to seek* the totality of conditions. However, the faculty of Reason cannot deviate from its own principle. The Principle of Reason is constitutive of Reason, and so simply describes its activity: Reason *does seek* the totality of conditions. As such, however, the Principle of Reason can be a merely descriptive law that may causally account for phenomena we observe in inner sense.

But this reading, while attractive, isn't a foregone conclusion. Some authors – notably MacFarlane (2000) and Leech (2017a, 2017b) – have suggested that the PNC is also normative for the understanding, as faculty, which must therefore be able deviate from its own law.[140] And similar concerns arise for Reason. Isn't the whole point of the *Critique of Pure Reason*, specifically the Transcendental Dialectic, to show that Reason itself, by its very nature, tends to assert totalities of conditions and thus to violate its own principle? Isn't Reason, according to Kant's analysis, a hopelessly flawed faculty that gets

[139] The idea sits particularly well with the faculty constitutivsim that has been defended by Schafer in a series of papers. See Schafer (2018b), (2019), (2020), (2021). I have great sympathies for Schafer's reading. However, my argument should not depend on it but also work with a Korsgaard-style constructivism. More on this soon.

[140] Stang (2014) can also be read as endorsing this reading. He writes: 'Logic brings the understanding (or more generally, thinking) into agreement with itself because it teaches *the normative laws that apply to that capacity*' (2014: 216; my emphasis). For a discussion, see Martens (ms).

caught up in endless metaphysical speculation? And wouldn't that make the proper Principle of Reason normative for Reason?

This is a crucial concern. Therefore, the primary task of this chapter – specifically Sections 3.2 to 3.5 – will be to show that Reason cannot deviate from its own principle; and so that the Principle of Reason, qua OEM, cannot be normative for Reason, and instead must be descriptive. But before I show that Reason never fails, I want to address another salient concern. One might worry that even if the Principle of Reason were descriptive of the *power of Reason*, it remains unclear how we can answer the Normative Question: how could the principle ever become normative for our *substantial selves*, if it is just a causal law of nature? Anderson worries that 'normative standards [...] would be threatened by a thoroughgoing naturalism' (2005: 292).[141]

A more general way of voicing this concern is to say that Ontological Symmetry entails a normative symmetry, according to which causal laws of mental powers should have the same normative standing for minds as causal laws of physical powers have for bodies. But this normative symmetry either undergenerates normativity by denying that the laws of our mental powers can be normative for us; or it overgenerate normativity by claiming that the laws of physical powers are normative for bodies. To address this concern, we must somehow thread the needle and deny this normative symmetry, without thereby denying Ontological Symmetry.

Threading the needle is fiddly. For while OEM specifies a necessary *condition* on normativity, it arguably does not specify the *source* of normativity. So even if we were to show that only minds, not bodies, failed to obey the laws of their powers,[142] this would not in itself explain why these laws are

[141] See also Pollok (2017: 6).

[142] One might object that laws of nature, qua being laws of *nature*, are 'exceptionless laws', as Anderson puts it (2005: 292); and thus, that there is no meaningful sense in which we could ever fail to abide by laws of nature. But that's clearly not true. Building on my discussion in Section 2.1, we can say that a substance, S, with a power, P, to-φ-when-c fails to abide by P's law iff S does not φ when c obtains (see also Section 3.2). This is the case, for example, if P is a dead or zombie power, and so if its φ-ing is (partially) cancelled by the real opposition of another power. And while bodies are, of course, also subject to real opposition, Kant sometimes seems to think that only minds are subject to real opposition from powers of the *same* substance. See, for example, (R3589, 17:76) and (R6640, 19:122). If true, this would suggest that there is a meaningful sense in which only minds can fail to abide by their laws,

normative for minds.[143] The mere fact that, say, sensibility leads us to violate some law L, does not make it so that L is normative for us. The gap between law and activity does not itself generate 'oughts'.[144] This point would hold not just for causal laws of nature, but also for the moral law as the causal law of freedom. The mere fact that I can act immorally does not make the moral law an imperative for me.

The literature offers a plethora of accounts as to what the sources of normativity might be. Let me highlight two. In the spirit of Korsgaard's constructivism (1996), one might argue that the law of a mental power becomes normative for us when we reflectively endorse the law as our own. Drawing on Groundwork II (4:412), Pollok suggests that a law or judgement becomes normative for us when we act on the representation of the law: 'What makes judgements normative [...] is that [...] as finite reasoners we do not simply act according to laws, as our bodies are bound by the laws of nature. Rather, we are capable of acting in accordance with *the representation of laws*' (2017: 13; my emphasis).[145]

We can abstract from the specifics of these accounts to observe some general features. Accounts, such as Korsgaard's and Pollok's, identify a source of normativity, X, whereby the law, L, of a power, P, becomes normative for S. Not only does X likely exceed the condition specified by OEM, but it is also supposed to be satisfied only by the laws of mental powers, not by the laws of physical powers, thereby establishing the required normative asymmetry (though this is often left implicit). Surely, bodies can neither reflectively endorse their laws, nor act on the representation of their laws.[146]

---

while bodies cannot.

[143] Baiasu (2016: 1194) observes that we cannot derive a substantive normative theory *analytically* from descriptive premises about the workings of our faculties, simply because analytic entailment cannot add anything that was not already contained in the initial premises.

[144] Though see Martens, who argues that '[t]he normativity of the moral law follows, for Kant, from the gap between practical reason's objectively necessitating a particular form of action and our non-rational faculties making it such that our action is not necessarily determined by reason' (ms).

[145] On the sources of normativity, see also the faculty-constitutivism defended by Schafer (2018a, 2018b, 2020, 2021).

[146] After all, Kant defines 'representations', in the broadest sense, as 'modifications *of the mind*' (A97; my emphasis), and so no body can have a representation of its own state.

Since the focus of this thesis is to answer the Epistemic Question, not the Normative Question, I want to remain largely neutral about the specific source of normativity, X. But note this: while Ontological Symmetry places or mental powers in the order of nature, it does not commit me to a 'thoroughgoing naturalism' (Anderson 2005: 292). It may well be that X draws on non-naturalistic resources to explain why a law L becomes normative for S. But this need not in itself contradict the claim of Ontological Symmetry that L is also a law of nature. For example, we may be able to reflectively endorse a law of nature, or act on a representation of a law of nature, even if that reflection or representation is not itself part of the natural order.

This allows us to thread the needle. Ontological Symmetry does not entail a normative symmetry because it leaves room for (non-naturalistic) sources of normativity, X, which would only be satisfied by the laws of mental powers and not by the laws of physical powers. Whatever X is specified to be, it would explain how the Principle of Reason can become normative for our *substantial selves* – answering the Normative Question. With that out of the way, I turn to the main task of this chapter, which is to show that Reason cannot deviate from its own principle, and therefore, qua OEM, that the Principle of Reason must be descriptive of the *power of Reason*.

3.2 Faculty Failure

There are multiple ways in which we, as substantial selves, can fail to abide by the Principle of Reason to seek the totality of conditions for a given conditioned. For example, we can seek what may appears to be a condition, when in reality it is not; we can also seek only some conditions rather than all of them. Most importantly, however, we can also *assert* rather than *seek* the totality of conditions to a given conditioned. In this assertion lies the dialectical error of speculative metaphysics in rational psychology, cosmology, and theology that Kant discusses throughout the Transcendental Dialectic.

This poses the question as to why do we commit such a dialectical error? There are two possible readings: what I call the 'Sensible Sources

Account' and what Willaschek calls the 'Rational Sources Account' (2018: 3). On the Sensible Sources Account, we fail to abide to the Principle of Reason because we, as substantial selves, are also subject to the corrupting influence of sensibility. On the Rational Sources Account, we fail to abide by the Principle of Reason because *the power of Reason itself* is dialectical and violates its own principle.[147] I will ultimately defend the Sensible Sources Account, but my aim in this section is merely to introduce both accounts and to set out their respective motivations.

I start with the Sensible Sources Account. This account draws strong support from Kant's general metaphysics of powers, as I have reconstructed it in Section 2.1. If powers are individuated by their respective laws, they cannot act contrary to these laws without ceasing to be the powers that they are. Powers therefore can never fail on their own. And indeed, Kant independently confirms this implication of his metaphysics of powers, writing: 'No power of nature can act against its own laws when it acts alone' (R2244, 16:283).

This point applies not only to physical powers, but also to mental powers. Kant is clear about this. For example, he claims: 'No body, therefore, deviates from its laws by itself, but by the mixing of other powers: it is the same with the understanding' (24:402) – a claim Kant repeats many times.[148] Yet not only the understanding, but also Reason and the will cannot deviate from their laws when considered in their pure, isolated form: 'If we had pure reason and pure understanding, we would never err; and if we had a pure will (without inclination), we would not sin' (R2246, 16:284). Even more emphatically, Kant states that it would be a contradiction in terms if reason, as a power of nature, were to deviate from its own laws:

[147] Both accounts have defenders in Kant's time. Heßbrüggen-Walter (2004: ch. 2.1) notes that the question of whether faculties themselves can fail, or whether such failure can only ever be brought about by an external influence, has been debated at least as far back as the 17th century.

[148] See also (2:291), (R2244, 16:283), (R2246, 16:284), (24:84), (24:102), (24:316), (24:396), and (25:577). Heßbrüggen-Walter notes: 'If a power of nature causes an activity, it must do so according to a law [...]. But to act contrary to the laws is impossible, it would mean not to act at all' (2004: 190); adding that a mental power, 'as a natural power, cannot act contrary to its own laws' (2004: 218).

> The understanding, and reason itself, left to themselves, never err, that much is undoubtedly clear, and certain, for this would be a manifest *contradictio in adjecto* [...]; they would then have to contradict their own powers, and no power can contradict itself, but it is only because other, external powers mix with those powers and the laws of our understanding and reason, from which, naturally, a mixed effect arises, and from this, in turn, an error. (24:168)

But if Reason cannot go wrong on its own, what causes us to violate its principle? Kant argues that we generally violate the principles of our higher faculties because, as *finite* agents, we are also subject to the corrupting influence of sensibility. This model is well known from Kant's moral thought where sin results from the influence of inclination.[149] But the same also goes for error: 'In error [...] we judge in a confused [*vermischt*] way, and mix up [*vermengen*] the effects of the other powers of the mind, and from this arises error [...]. For [error] is in fact nothing other than a bastard, as it were, of sensibility and understanding' (24:84).[150]

In view of these general considerations, we should expect that the dialectical error of speculative metaphysics is not the result of reason's own failure, but of the corrupting influence of sensibility. Reason never deviates from its own principle. And indeed, Kant opens the Transcendental Dialectic with a discussion of the sources of dialectical error, in which he rehearses all main lines of the Sensible Sources Account. The passage is remarkable, so I will quote it in full. Sentences are numbered for ease of reference. Kant writes:

> [1] [T]ruth, as well as error, and along with it illusion, as the inducement to the latter, can only be found in judgement, i.e., only in the relation of the object to our understanding. [2] In a cognition that is in complete agreement with the laws of the understanding, there is no error. [3] In a representation of the senses, there is also no error (because it contains no judgement at all). [4] No power of nature can, however, deviate from its own laws of its own accord. [5] Therefore, neither understanding alone (without the influence of another

[149] Heßbrüggen-Walter notes that, on Kant's account, we must 'not regard the sinfulness of the soul as an independent property, but as the result of the interaction of the faculties present in the soul' (2004: 30). I agree.

[150] On the corrupting influence of sensibility, see also (2:202), (9:53-4), (9:561), (R2246, 16:284), (24:80), (24:103-4), (24:162), (24:168), (24:282), (24:401-3), (24:721), (24:817), and (24:824).

> cause) nor the senses alone would err; the former not because, when it acts solely according to its laws, the effect (the judgement) must necessarily agree with these laws. [6] In the agreement with the laws of understanding, however, consists the formal of all truth. [7] In the senses there is no judgement at all, neither true nor false. [8] Since we have no other sources of cognition besides these two, it follows: that error is effected only by the unnoticed influence of sensibility on understanding, whereby it happens that the subjective grounds of judgement flow together with the objective ones , and make them deviate from their determination; [9] just as a moving body itself would always keep the straight line in the same direction, but if another power flows into it at the same time in a different direction, it will deflect into a curvilinear movement. [10] In order to distinguish the distinctive activity of the understanding from the power that interferes with it, it will therefore be necessary to regard the erroneous judgement as the diagonal between two powers that determine the judgement in two different directions, which, as it were, enclose an angle, and to resolve that composite effect into the simple one of understanding and sensibility[.] (A293-5/B350-1)

Let's unpack this passage slowly. The first thing to note is that Kant uses understanding here in the broad sense, as including Reason. This is implied by Sentence (8), which says that 'we have no other sources of cognition besides these two', meaning that we have no other sources of cognition besides sensitivity and understanding. For this statement to be true, the understanding has to pick out the entire higher faculty of cognition, including Reason. Of course, this also makes a lot of contextual sense, since the passage is at the beginning of the Transcendental Dialectic, which is supposed to discuss Reason.[151]

Moreover, the passage motivates a conceptual distinction between two types of failure. First, there is *failure-as-deviation*, which is the sense of failure I discussed earlier and which, according to Sentence (4), would require a deviation from the law of a power. More generally, we can say that x deviates from the law L of a power to-φ-when-c iff x does not φ when c is obtains. But there is also *failure-as-error*, which simply consists in making false judgements.

[151] On Kant's broad and narrow usages of 'understanding' and 'reason', see my discussion in Section 1.1.

The mistakes of speculative metaphysics also involve failure-as-error; for not only do the antinomies commit us to a contradiction, but, as I will argue in section 3.4, the claim that there are totalities of conditions is itself false.

This raises the question: how do these two types of failure relate? Sentences (2), (5), and (6) provide the answer. Kant claims that we cannot make false judgments if we don't deviate from the laws of understanding, because the laws of understanding give 'the formal of all truth'. Again, this point is meant to apply not only to understanding in the narrow sense, but also to Reason. Kant makes this point, for example, in the *Philippi Logic* when he states that 'truth is nothing other than the agreement of cognition with the laws of understanding and reason' (24:84).[152] This is, of course, a very strong claim about the nature of truth, which I cannot elaborate here.[153]

Thus, failure-as-error entails failure-as-deviation. This has an important consequence. For if reason itself were to make false judgments – for example, by asserting that there is a totality of conditions for a given conditioned – it would follow that reason itself would have had to deviate from its own principle, which would, in turn, entail the falsity of the Sensible Sources Account. Note that the quoted passage remains neutral (and so can I) on whether, in the case of the higher faculties of cognition, failure-as-deviation also entails failure-as-error, or whether there can be alethically innocent ways of deviating.[154]

The passage repeats the main claims of the Sensible Sources Account. Sentence (4) states that no power of nature can deviate from its law; Sentence (5) applies this claim to the understanding or Reason; and Sentence (8) states that we deviate from the principles of Reason only because of the 'unnoticed influence of sensibility'. The passage goes further, however, and explains exactly how we should think about this influence. Sentence (9) compares the

[152] See also (R2246, 16:284), and (24:168).

[153] See Grier (2001: 102-3) for an extended discussion.

[154] Note that failure-as-error is always only a failure-as-deviation of our higher cognitive faculties. Other faculties, such as sensitivity, don't even judge, and so their representations cannot (even) be false. See sentences (3) and (5). We could also imagine a power of Unreason whose laws necessarily lead to error; here, then, truth, not error, would be a kind of failure-as-deviation. I will return to the possibility of Unreason in Chapter 6.

influence of sensibility to that of a 'moving body' on another; and sentence (10) goes on to suggest that we must 'regard the error of judgement as the diagonal between two powers' – sensibility and reason. Kant is thinking here of a power (or force) parallelogram, as shown in Figure 2 below.[155]

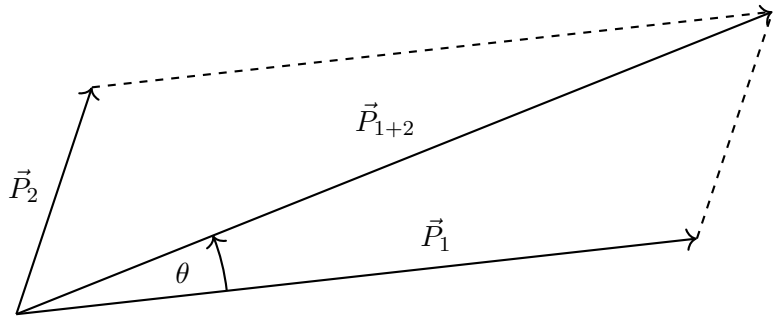

Figure 2: power parallelogram

Let me explain. Say, vector $\vec{P}_1$ represents the power of Reason and its proper activity, as defined by the Principle of Reason. If we adhered to the Principle of Reason, following vector $\vec{P}_1$, we would arrive at true judgements, Kant thinks. However, our minds are also subject to the power of sensibility, represented by vector $\vec{P}_2$, which distorts the activity of Reason. Our minds therefore manifest only the aggregate vector $\vec{P}_{1+2}$, which Kant calls the 'composite effect'. Since $\vec{P}_{1+2}$, or 'the diagonal', deviates from the Principle of Reason, it must yield 'erroneous judgement'. Indeed, Kant seems to think that we can measure the degree of error by the angle enclosed between vectors $\vec{P}_1$ and $\vec{P}_{1+2}$, here represented as θ.

Leaving aside the question of whether the activities of mental powers can be quantified,[156] the general metaphysics of powers developed in Section 2.1 allows us to take Kant's picture of the power parallelogram quite literally.[157]

[155] On the power parallelogram, see also (24:402).

[156] Based on the *Metaphysical Foundations of Natural Science*, one might argue that laws of the mind cannot be quantified 'because mathematics cannot be applied to the phenomena of the inner sense and its laws' (4:471). I will return to this striking pronouncement in Section 6.4.

[157] I thus disagree with Heimsoeth (1966: 8-9), who demotes the power parallelogram, to a mere metaphorical analogy. And unlike Grier, I am not taken aback by Kant's account of

Mental powers are powers in the strict sense, and their activities are subject to real opposition from other powers.[158] Indeed, we have seen that Kant has a special term for powers whose activity is cancelled, calling them dead powers. Now, Reason is not quite dead, because its activity is only partially cancelled – distorted by an angle φ –, but Reason becomes what I have called a zombie power.

The Sensible Sources Account, so specified, boasts several advantages. Apart from the almost decisive textual evidence in its favour, it also dovetails perfectly with Kant's general account of the mental faculties – how they are individuated and how they can be cancelled. As such, it also offers an account of the sources of dialectical error that is consistent with other instances of faculty failure; for example, our failure to abide by the will's moral law due to the influence of sensibility. The account also offers a genuine explanation of why we make dialectical errors; we don't do so because Reason *randomly* decides to disobey its principle, but because, as finite reasoners, we can be led astray by sensibility.

In view of these advantages, it is not surprising that the Sensible Sources Account also has its adherents in the literature. Heßbrüggen-Walter notes that, according to Kant, error requires the real opposition of mental powers: 'Error is [...] a mental actio that – in contrast to the mere absence of knowledge [...] – arises from a *real opposition of powers*' (2004: 211; my emphasis). Likewise, Schafer notes that we only ever fail to abide by the principles of our mental faculties because of external influences, such as sensibility: 'every faculty is paired with an internal principle that explains how this faculty functions insofar as it is free from abnormal, external "hindrances"' (2019: 184).[159]

---

error, which Grier judges to be 'quite "mechanical"' (2001: 104). The power parallelogram is further discussed in Heßbrüggen-Walter (2004: 221).

[158] Kant makes it clear in (A265/B320-1), (A273/B329), and (24:64) that error results from the real opposition of mental powers.

[159] In conversation, Schafer suggested that the corrupting influence can also come from other non-sensible faculties, adding that this non-sensible interference, however, can only occur in beings that have sensibility. In Section 3.5, I will defend a pluralism about the ways we are influenced, and I am open to Schafer's suggestion. On the types of influence, see also Deleuze (1983/2008: 19).

I, too, am an adherent, and I will further defend the Sensible Sources Account in the coming sections. For now, note that the account also solves the Problem of Normativity. For if Reason can never fail to obey its own principle, then, qua OEM, the Principle of Reason must be descriptive of the power of Reason, even if it is normative for us. Moreover, the unique activity of Reason must manifest itself in the inner sense, if only as part of the aggregate activity with sensibility. And so, even though we are only able to introspect this erroneous aggregate activity – which, in turn, motivates the Problem of Perfection – we may be able to isolate Reason in its pure, unadulterated form. I say more about this in Chapter 6.

Despite its many advantages, however, the Sensible Sources Account isn't a foregone conclusion. There are also textual and structural reasons in favour of the Rational Sources Account. Throughout the *Critique of Pure Reason*, Kant repeatedly maintains that Reason, by its very nature, tends to assert totalities of conditions and thus deviates from its own principle of seeking only such totalities. For example, Kant writes that 'reason is driven by a *tendency of its nature* to go beyond the use of experience' (A797/B826; my emphasis); and thus that speculative metaphysics is a '*natural* dialectic of human reason' (A669/B697; my emphasis).[160] Indeed, the A edition of the *Critique* opens with the following striking sentences:

> Human reason has a special fate in a species of its cognitions: that it is troubled by questions that it cannot dismiss, for they are given to it *by the nature of reason itself*, but which it cannot answer either, for they exceed all capacity of human reason. [...] It starts from principles whose use in the course of experience is unavoidable and at the same time sufficiently proven by it. With this, it (*as is also its nature*) rises ever higher, to more distant conditions. But when it becomes aware that in this way its business must always remain unfinished, because the questions never cease, it sees itself compelled to take refuge in principles that transcend all possible experiential use, and yet seem so unsuspicious that even *common human reason is in agreement with them*. But in so doing, it plunges

[160] Throughout the *Critique*, Kant frequently states that the dialectical error is somehow 'natural' to Reason. See (A297-8/B353-4) (A316/B373), (A323/B380), (A406/B433-4), (A584/B612), (A625/B653), (A641-3/B669-71), (A702/B730), and (A877/B849).

> into darkness and contradictions, from which it can indeed infer that *there must be hidden errors at the ground* of them somewhere, but which it cannot discover because the principles it uses, since they go beyond the limits of all experience, no longer recognise any touchstone of experience. The battleground of these endless disputes is now called *metaphysics*. (Avii-viii; my emphasis)

It is because of its transgressions that Reason needs a stern Prussian critique. Considerations such as these have led scholars to support the Rational Sources Account. Foremost among them is Bennett, who calls the Sensible Sources Account a 'flimsy attempt' that 'conflicts with everything else Kant says on the subject' of dialectical error; and instead, claims that 'transcendental reason, as such, is a source of error' (1974: 270). Another prominent defender of the Rational Sources Account is Willaschek , who not only gave the account its name, but also argues throughout his book that 'pure reason is driven by its own need or its nature to answer [metaphysical] questions, even if the answers may not be ultimately warranted' (2018: 3).[161]

The Rational Sources Account is a formidable challenge. Because not only does it fit much of the spirit of the *Critique* – as a critique of Reason, not sensibility –, but it can also allow for sensibility to be *a* source of *some* error, as long as Reason is the source of dialectical error. If the Rational Sources Account was correct, then Reason could depart from its own principle, which might thereby become normative for Reason, in turn, motivating the Problem of Normativity. To reject the Rational Sources Account, I will take a closer look at the sources of dialectical error, and argue that Reason is not the culprit (§§ 3.3 and 3.4).[162]

---

[161] That said, Willaschek's reading is more nuanced in ways that I cannot discuss here. Apart from Bennett and Willaschek, Heimsoeth suggests that the source of speculative metaphysics is to be 'found in the essence of reason itself' (1966: 5). Dyck notes that 'transcendental illusion has its source in reason, and [...] it [...] invites a misuse of the categories of the understanding' (2014: 82). Grier denies that reason is the source of dialectical error, but claims that it is the source of the transcendental illusion that leads us into error (2001: 116).

[162] Like Grier (2001: 101-2), Deleuze (1983/2008) frames the Kant's seeming commitment to both the Sensible Sources Account and the Rational Sources Account as a problem that needs to be resolved. He writes: 'We are now touching on a problem which fully concerns the *Critique of Pure Reason*. How can the idea of illusions internal to reason or of the illegitimate employment of the faculties be reconciled with another idea, no less essential to Kantianism: the idea that our faculties (including reason) are endowed with a good nature' (22). I will

3.3 The Transition Passage

In the Introduction to the Transcendental Dialectic, Kant attempts to provide a general explanation of all dialectical error in special metaphysics.[163] He argues that we mistakenly transition from a principle that tells us to seek the totality of conditions to a principle that asserts these totalities. Kant summarises this transition in a passage that Willaschek has aptly coined the 'Transition Passage' (2018: 18). My aim in this section is to first clarify the principles involved in the passage, so that in the next section we can trace the transition and Reason's involvement in it. Here, then, is the Transition Passage:

> [T]he proper principle of reason in general (in the logical use) is: *to find the unconditioned to the conditioned cognition of the understanding, by which the unity of the same is completed*. This Logical Maxim, however, cannot become a Principle of Pure Reason in any other way than by assuming that *if the conditioned is given, then the complete series of subordinate conditions, which is therefore itself unconditioned, is also given*, (i.e., contained in the object and its connection). [...] But the principles arising from this Supreme Principle of Pure Reason will be transcendent with respect to all phenomena, i.e., no adequate empirical use can ever be made of it. (A307-8/B364-5; my emphasis)

The passage articulates two principles, which I have italicised and which, following Grier (2001: 119-122), I will refer to as 'P_1' and 'P_2' respectively. To unpack these principles, let me begin with a brief terminological note. Aside from a few semantic subtleties, Kant uses 'the unconditioned' in P_1 as roughly equivalent with 'the totality of conditions'. After all, he also writes that 'the unconditioned alone makes the totality of conditions possible, and conversely the totality of conditions is always itself unconditioned' (A322/B379). The totality of conditions, in turn, simply expresses the all of conditions – '*allness*

resolve this problem by arguing that, strictly speaking, the dialectical error is no fault of Reason.

[163] I leave aside the question of whether this attempt is successful. Bennet (1974) has famously been very critical of whether the errors in rational psychology, cosmology and theology all manifest the same mistake. However, see Willaschek (2018) for a more unified reading of the Dialectic.

(universitas) or *totality*' (ibid.) – or, as he puts it in P_2, 'the complete series of subordinate conditions'.[164]

While both P_1 and P_2 are about totalities of conditions, they differ in two important ways. First, P_1 is about '*cognitions* of understanding', whereas P_2 is supposed to be about 'the *object* and its connection'. I have more to say about the meaning of both 'cognition' and 'object' later. Second, while P_2 makes a descriptive claim – 'the complete series [...] *is* also given' – the infinitive construction of P_1 – '*to* find the unconditioned' – is commonly read as articulating 'a demand of reason' (A305/B362).[165] Whether it is also a demand *for* Reason is the big question, of course. In light of all this, we can now state both principles more precisely as follows:

(P_1) For all x: if x is a *cognition* and x is conditioned, then *find* all y, such that y is a cognition and y is a condition of x.

(P_2) For all x: if x is a given *object* and x is conditioned, then all y *are given*, such that y is an object and y is a condition of x.

Kant introduces his own names for the principles: the 'Logical Maxim' and the '(Supreme) Principle of Pure Reason'. There is an overwhelming consensus in the literature that (i) the Logical Maxim is meant to pick out P_1 – so much seems textually uncontroversial – and, more importantly, that (ii) the (Supreme) Principle of Pure Reason is meant to pick out P_2. For example, Willaschek writes that 'the "Logical Maxim" [...] requires us to find a condition for each conditioned cognition' and 'according to [the "Supreme Principle"] if something conditioned is given, then so is the complete series of conditions' (2018: 6-7).[166]

[164] For a nuanced discussion of Kant's notions of 'the unconditioned', 'the complete series of conditions', and 'the totality of conditions' – what they mean and how they relate –, see Heßbrüggen-Walter (2004: 252-9), Willaschek (2018: ch. 3.3), and Kraus (2020: 177-8n).

[165] If only a hypothetical demand, as Proops (2021: 46) claims. See also Dyck (2014: 83).

[166] Grier notes that 'Kant calls [P_2] the "supreme principle of pure reason"' (2001: 122). Dyck writes that 'Kant identifies [P_2] as the "supreme principle of pure reason"' (2014: 83). Proops says that 'Kant refers to [P_2] as "this supreme principle of pure reason"' (2021: 48).

While this reading is straightforward, it is neither necessary, nor necessarily attractive. I agree with (i), but I disagree with (ii). To see why, we must better understand what it means that the Logical Maxim can become a Principle of Pure Reason only by assuming P_2. The locution 'x can *become* y only on condition z' typically does not imply that $y = z$, but instead that x is somehow transformed into y. For example, we might say that I can become an uncle only on the condition that one of my siblings has a child. But this claim does not imply that by becoming an uncle I become the child of my siblings. The same, I suggest, applies to the Principle of Pure Reason, which need not be identified with P_2.

But if the Principle of Pure Reason is not P_2, what is it? The answer to this question, I suggest, lies in the contrast between 'logical' and 'pure' – a contrast Kant highlights by contrasting the Logical Maxim with the 'Principle of *Pure Reason*' (A307/B364; emphasis original). While 'logical' can mean many things to Kant, he introduces the term a few pages before the Transition Passage to mean that which is merely formal: 'There is a merely formal, i.e., logical, use of [reason], insofar as reason abstracts from all content of cognition' (A299/B355). The Logical Maxim, which is '[t]he proper principle of reason [...] in the logical use' (A307/B364), thus concerns only formal cognitions that abstract from all content.[167]

Kant contrasts the logical use of reason, so understood, contrasts with the 'real' or 'transcendental' use of Reason, which does not abstract from all content of cognition (A299/B355). I suggest that the Logical Maxim and the Principle of Pure Reason mirror this contrast, and so that the Principle of Pure Reason concerns real cognitions, i.e., cognitions with content – not formal cognitions, and not objects. Extending Grier's naming scheme, I thus distinguish between two versions of P_1: there is $P_{1.1}$, which concerns formal cognition (= the Logical Maxim); and there is $P_{1.2}$, which concerns real cognition (= the Principle of Pure Reason). Stated precisely:

[167] On this sense of the logical use of reason, see also Willaschek (2018: 31-2) and Kraus (2020: 175).

($P_{1.1}$) For all x: if x is a *formal* cognition and x is conditioned, then find all y, such that y is a *formal* cognition and y is a condition of x.

($P_{1.2}$) For all x: if x is a *real* cognition and x is conditioned, then find all y, such that y is a *real* cognition and y is a condition of x.

My reading differs sharply from the one proposed by Willaschek (2018). While both of our readings imply that there is a regulative version of the (Supreme) Principle of Pure Reason, Willaschek arrives at this conclusion by distinguishing between a regulative and a constitutive version of P_2 (2018: ch. 5). On his reading, then, the Principle of Pure Reason always makes a descriptive claim about objects; and by implication, Willaschek is forced to say that some regulative principles are not prescriptive. On my reading, however, the Principle of Pure Reason never makes a constitutive claim about objects, but always expresses a regulative demand about (real) cognition. (More on the normativity of P_1 in a moment).

At this point, now, one might be tempted to distinguish between the Principle of Pure Reason as denoting $P_{1.2}$, on the one hand, and the *Supreme* Principle of Pure Reason as denoting P_2, on the other. But this is ill advised. Not only do the near-identical names suggest that both principles refer to the same proposition, but the text also allows for it. The Supreme Principle of Pure Reason is introduced several sentences after P_1 and P_2 – indicated by the ellipsis in the quotation of the Transition Passage. Its reference is therefore grammatically indeterminate. Yet if the Principle of Pure Reason denotes $P_{1.2}$, so should the Supreme Principle of Pure Reason.

This reading of the Supreme Principle of Pure Reason also makes sense of Kant's claim that 'the principles [*Grundsätze*] arising from this Supreme Principle [*Prinzip*] of Pure Reason will be transcendent' (A308/B365).[168]

[168] While Kant does not consistently distinguish between 'Prinzipien' and Grundsätze', the terminological difference provides yet another reason to identify the Principle of Pure Reason ('Principium der reinen Vernuft') with the Supreme Principle of Pure Reason ('oberste[s] Prinzip der reinen Vernunft'). Because within the broader context of the Transition Passage, these are the only two instances in which Kant uses the term 'Prinzip' (or its Latin equivalent) as opposed to 'Grundsatz'. Moreover, the terminology also suggests that neither the Principle of Pure Reason nor the Supreme Principle of Pure Reason denote P_2: right after introducing

Insofar, as we naturally transition from P_1, specifically $P_{1.2}$, to P_2 – more on this transition in Section 3.4 –, it makes sense to say that P_2, and with it all the other principles of special metaphysics, 'arise' from $P_{1.2}$. Importantly, the claim that the principles that arise from the Supreme Principle of Pure Reason are transcendent does not imply that the Supreme Principle itself must be transcendent; after all, it denotes $P_{1.2}$.

And there is another crucial advantage. Its lofty title suggests that the 'Supreme Principle of Pure Reason' defines, or at least governs, Reason as a faculty – much like the categorical imperative, as the 'supreme practical principle', governs pure practical reason (4:428).[169] But only $P_{1.2}$ can be a principle that governs the activity of a faculty. For not only does it specify an antecedent condition c – the conditioned cognition –, but the consequent also articulates an activity φ – to find the totality of conditions. By contrast, P_2 does not specify any activity whatsoever, but instead make a claim about the structure of reality – that there is a totality of conditions. It cannot therefore be a principle that governs the activity of a faculty.

These considerations now also allow me to be more precise about the Principle of Reason, as I have introduced it in Section 1.1. Since the Principle of Reason is meant to pick out the principle that defines or governs Reason as a faculty, it cannot be P_2. Moreover, $P_{1.1}$ and $P_{1.2}$ only specify different employments of the same faculty, Reason, to different domains: the domain of formal cognition, and the domain of real cognition. I therefore take P_1, in its domain-neutral formulation, to be *the* Principle of Reason. (And I will further elaborate on the content of P_1 in Chapter 4.)

This much is clear: the Principle of Reason, so understood, is normative *for us*. We ought to seek the totality of conditions, partly because we can fail to do so when we instead assert the totality. The big question now is whether P_1 is also normative *for the power of Reason*. Note that the text leaves this open. While the infinitive construction of P_1 – 'to find' – is usually read as articulating a demand, it can also be read as simply describing the activity

P_2, Kant refers to the proposition as '[e]in solcher *Grundsatz*' (A308/B364; my emphasis).

[169] On the connection between P_1 and the categorical imperative, see Allison (2004: 53).

of a faculty: Reason is the faculty that seeks to find the totality of conditions to the conditioned cognitions. That's just what Reason, qua Reason, does. However, whether P_1 is normative for Reason will depends on whether Reason itself can deviate from P_1.

And whether Reason itself can deviate from P_1 depends on whether Reason endorses P_2. Kant thinks that all dialectical error in metaphysics is based on an argument he calls 'sophisma figurae dictions', which takes P_2 as its major premise and then infers for a given object that there is a totality of its conditions (A499/B528). This conclusion, then, leads straight to the errors in rational psychology, cosmology, and theology, Kant argues. Now, if Reason itself were to endorse P_2, it would thereby be the source of all dialectical error, deviating from its own principle of seeking the totality and instead asserting it. To see if this is the case, however, we must take a careful look at exactly transition from P_1 to P_2.

3.4 Belief in Totalities

Why do we come to endorse P_2? The literature offers several answers to this question, which differ in their assessment of how rational said endorsement is. *Conservative readings*, like those of Grier (2001: ch. 4) and Proops (2021: ch. 1),[170] suggest that endorsing P_2 is entirely irrational, and entirely the result of cognitive confusion.[171] *Moderate readings*, such as those of Willaschek (2018: ch. 1) and Kraus (2020: 177-81), claim that we are permitted to endorse P_2 as a regulative heuristic, while denying that P_2 can be justified. And then there are *liberal readings*, like those of Schafer (2023: ch. 5) and Benzenberg & Chignell (2024), according to which our endorsement of P_2 – or a variant thereof – is rationally justified as a doctrinal Belief.

[170] See also Guyer (2019: 34).

[171] Indeed, Proops (2021: 46-50) reconstructs a threefold confusion to transition from P_1 to P_2. First, we confuse P_1, which for Proops is a hypothetical imperative about cognitions, with a hypothetical imperative about objects; then, we confuse the hypothetical imperative about objects with a categorical imperative; and finally, we confuse a categorical imperative with an assertion about reality, claiming that totalities of conditions is itself given. For a discussion of Proops's reading, see Benzenberg & Chignell (2024).

In this section, I will defend what is perhaps the most liberal reading to date. Specifically, I argue that Reason does, in fact, endorse as doctrinal Belief a variant of P_2 that is much closer to P_2 than the variant discussed by Schafer (2023: 172). However – and this is the main lesson of this section –, even on this most liberal reading, Reason still falls short of outright endorsing P_2 in a way that would lead to dialectical errors. These errors, I argue, only occur when we confuse the subjective grounds of doctrinal Belief for objective grounds of knowledge. And this confusion is the result of transcendental realism, not Reason. Reason never fails.

The core idea of liberal readings is easy to motivate: P_2 is an ideal candidate for doctrinal Belief (as defined in Section 2.5). For not only does P_2 fall outside the domain of possible knowledge, satisfying conditions (i) and (ii) of the definition, but P_2 also satisfies conditions (iii) and (iv) by being a necessary assumption to realise the final end of Reason – 'it is only possible to find something *if it exists*' (Schafer 2023: 166). To be more precise, Reason's final end to find the totality among our real cognitions (as per $P_{1.2}$) cannot be realised 'in any other way than by assuming that' there is a totality of conditions among the objects that corresponding to these cognitions (A307/B364).[172]

Although easily motivated, liberal readings face two challenges. The first is that they seem to reproduce the dialectical errors, especially the antinomies, in the mode of doctrinal Belief. With *doctrinal Belief* in P_2, we can infer via *known* premises to *doctrinal Belief* in the thesis and antithesis of the respective antinomies. We know from experience that there are conditioned objects, and Kant also claims that the series of conditions is either finite or infinite (A418/B445) – the former being the case in the thesis, the latter in the antithesis. From apagogic proofs, we further know that neither the thesis nor the antithesis can be true.[173] Put into an argument (where 'B' and 'K' stand

[172] I am here bracketing the problem that arises from the fact that it is unclear whether the logical conditioning relations that occur between our cognitions can ever correspond to the plethora of real conditioning relations.

[173] It's a difficult question *how* we can come to know these premises. For example, there has been a long-standing debate about whether the premises that appear in the apagogic proofs are dialectically neutral with respect to transcendental idealism and transcendental

for Kant's notions of doctrinal Belief and knowledge):

Doctrinal Antinomies

(P1) B(For all x: if x is a given object and x is conditioned, then all y are given, such that y is an object and y is a condition of x.)

(P2) K(There is an a: a is a given object and a is conditioned).

(P3) K(For all x and y: if y is given and y is an object and y is a condition of x, then either thesis or antithesis).

(P4) K(¬thesis) and K(¬antithesis).

(C) *Therefore*, B(thesis) and B(antithesis).

The Doctrinal Antinomies argument turns on the idea that the conclusion must match the justification of the weakest premise – with the practical grounds of doctrinal Belief being weaker than the truth-guaranteeing grounds of knowledge. This idea is as intuitive as it is false. The practical justification of Belief need not be preserved over entailment. In the present case, specifically, neither assuming the thesis nor the antithesis are *necessary* means of realising the final end of reason (as specified in $P_{1.2}$). We can find the totality of conditions whether it is finite or infinite, and so assuming that the totality is finite or infinite is not necessary. Assuming the thesis or antithesis would therefore violate condition (iv) of the definition.

But the Doctrinal Antinomies argument gives way to a second more difficult challenge to the liberal reading. Premises (P2), (P3), and (P4) of the argument jointly entail that we *know* that P_2 is *false*; that is, we know that there no totality of conditions can be given for a conditioned object. Indeed, this conclusion independently follows from Kant's theory of discursive cognition, according to which antecedent 'conditions are given [only] through the successive synthesis of the manifold of intuition' (A417/B444). This synthesis, however, is never complete, and so the totality of conditions is never given.[174]

realism. See Strawson (1966/2005: 175) and Malzkorn (1999: 118). And while it is of course plausible that the complete series of conditions is either finite or infinite, it is unclear whether Kant takes this claim to be analytically or synthetically justified. What is clear, however, is *that* Kant would claim that we know these premises.

[174] Schafer (2023: 155-6) highlights the same problem.

But if we know that P_2 is false, then we cannot have doctrinal Belief in P_2 as this would violate condition (ii) of the definition.

In fact, this is not only a challenge to liberal readings, but it creates an overlooked problem for Kant's general theory of Reason.[175] If we know that no totality of conditions can ever be given among objects, then we also cannot be required to find it among our real cognitions. For we can only be required to do what is possible – 'ought' implies 'can' for Kant (see Section 2.1) – and yet it seems that we can only find the totality of conditions among our real cognitions if there is a corresponding totality among objects. Thus, by knowing that P_2 is false, Reason's final end, as specified by $P_{1.2}$, becomes null and void for us; in the same way, that the 'impossibility of the [highest good] must prove the falsity of the [moral law]' (5:114).

The general strategy for solving this problem must be to say that not P_2, but a close variant of it, is necessary to realise the final end of Reason, and can thus be held as doctrinal Belief. Schafer (2023: ch. 5) proposes such a variant. He distinguishes between real conditioning relations ('R conditions' for short) that are '*limited to appearances*', and non-R conditions that are '*not limited to appearances*' but instead are 'conditions that apply to both appearances and things in themselves' (2023: 265). On his reading, then, the variant of P P_2 makes reference to both conditioning relations in a disjunctive consequent. While Schafer offers various formulations of this variant, here is a tidied up version:

(P_2^*) For all x: if x is a given appearance and x is R conditioned, then *either* (a) all y are given, such that y is an appearance and y is an R condition of x, *or* (b) a z is given, such that z is a thing in itself and z is a non-R condition of x.[176]

While being ugly, P_2^* seems to help. The antinomies arise only for totalities of R conditions – spatio-temporal, mereological, causal and modal

[175] I thank Ralf Bader for pushing me on this point.

[176] Schafer (2023: 156) thinks that P_2^* is the Supreme Principle. My objection from the previous section applies.

conditions, to be precise – and it is only R conditions that are given by successive synthesis. So the antinomies and our discursive mode of cognition entail that the principle is false only for the first disjunct (a), not for the second disjunct (b); and so, more generally, they don't entail that P_2^* is false. By not being known to be false, P_2^* not only saves the final end of Reason, also satisfies condition (ii) of the doctrinal Belief. And indeed, Schafer argues that P_2^* is justified as a doctrinal Belief (2023: 169-78).

Moreover, P_2^* makes possible an attractive story about the sources of dialectical error. Note first that P_2^* prevents doctrinal antinomies, not only because practical justification need not be preserved over entailment, but also because the second disjunct (b) doesn't concern R conditions. So P_2^* itself is not dialectical. It is only on the condition of transcendental realism – the 'view that appearances and things in themselves are identical' (Watkins 2010: 146) – that the two disjuncts collapse and we end up with P_2, which in turn generates the dialectical errors.[177] Thus, it is not Reason's commitment to P_2^*, but transcendental realism's leap to P_2, that signs responsible for speculative metaphysics.

While Schafer's reading is very impressive, it does not work. I argue P_2^* cannot be held as doctrinal Belief. Schafer himself notes that there are two senses in which an object can be 'given' for Kant: either 'the thing *exists*' or it is 'given to us *as a determinate object of cognition*' (2023: 158).[178] The first sense – which Schafer seems to intend (2023: 24, 174) – might imply that we actually know (b) and so P_2^* to be true; we know via analytic entailment that all appearances must be grounded in things in themselves (see Section 2.4). The second sense of given entails that we know (b) and so P_2^* to be false; no noumenal ground can ever be cognitively given. Whichever way you look at it, P_2^* either violates condition (i) or (ii) of doctrinal Belief.[179]

[177] Although Schafer does not make this point in the book, he suggested it to me in conversation.

[178] On the two senses of 'given' in Kant, see also Willaschek (2018: ch. 3.1).

[179] Nor is it clear how P_2^* can satisfy condition (iv) of doctrinal Belief. Why should it be necessary for us to assume the disjunction expressed in P_2^* in order to realise the final end of Reason? After all, assuming non-R conditions doesn't help us to find totalities of real cognition about R conditioned objects.

So here is my alternative to Schafer, which turns on a simple idea: we can find some x, and so we can be required to find x, even if x *does not yet exist*, as long as x *will come to exist*. For example, one might say that the Israelites were tasked with finding the Messiah, even though the Messiah was yet to be born. The same point, I would argue, applies to totalities of conditions. To find the totality of conditions, it is not necessary that this totality is already given – so P_2 actually violates condition (iv) – but only that it will be given. Again extending Grier's naming scheme, we can thus distinguish between the futurised version of P_2, and a presentist version of P_2 (i.e. the original P_2), both of which only invoke appearances and R conditions:

($P_{2.1}$) For all x: if x is a given object and x is conditioned, then all y *will be given*, such that y is an object and y is a condition of x.

($P_{2.2}$) For all x: if x is a given object and x is conditioned, then all y *are given*, such that y is an object and y is a condition of x.

In a nutshell, I argue that Reason, as governed by $P_{1.2}$, commits us to $P_{2.1}$ as a doctrinal Belief. We have already seen that $P_{2.1}$ satisfies conditions (iii) and (iv) – we can only find the totality of conditions among our real cognition if there will be such a totality among appearances. I will argue that $P_{2.1}$ also satisfies conditions (i) and (ii). Moreover, I contend that $P_{2.1}$ itself does not lead to dialectical errors. It is only when we come to endorse $P_{2.2}$ as purported knowledge that the dialectical errors arise. And this shift from $P_{2.1}$ to $P_{2.2}$ occurs only if we accept transcendental realism as true. Transcendental realism pushes us over the edge. On this general point, I agree with Schafer, and many others.[180]

While $P_{2.1}$ clearly satisfies condition (i) – we cannot know whether a totality of conditions will ever be given –, it might be said to violate condition (ii) on three grounds. First, one might object that we know that no totality of conditions can ever be given (in the cognitive sense of 'given') , since this

[180] The role of transcendental realism in the transition from P_1 to P_2 is also discussed by Malzkorn (1999: 103), Grier (2001: ch. 3), Willaschek (2018: ch. 9), and Proops (2021: ch. 1).

would require an infinite synthesis, which we cannot, however, complete in our finite lives. In response to this objection, note that Kant argues that we must have 'doctrinal Belief in a future life' precisely so that we have enough time to realise Reason's final end (A827/B855).[181] So while it is true that the totality of conditions will *never* be given – i.e., at no finite point in time t –, it might be given over the infinite time of our immortal existence.[182]

Second, one might object that we know that totalities cannot be given, not even over an infinite time, because once the totality was given, we would again run into the antinomies. The worry is that the antinomies would re-emerge at the end of history, entailing that $P_{2.1}$ is false. But this is not obvious. Kant notes that antinomies only arise if 'one leaves undecided whether and how this totality can be brought about' (A417/B445). Arguably, however, this is only left undecided if we assert the totality of conditions without having gone through the successive synthesis. If, however, we had completed an infinite synthesis over an infinite time, it would be decided whether the series of conditions has a first member (= thesis) or not (= antithesis).

Third, and finally, one might object that $P_{2.1}$ is known to be false because it is metaphysically impossible that a totality of conditions could ever be given to discursive cognisers, like us. After all, no unconditioned totality of conditions could ever be given to discursive cognisers, because discursive cognition depends essentially on sensibility (A19/B33; A51/B75); and 'objects of the senses are given to us as conditioned' (A497/B525). While I agree with the argument, all this shows is that we are not *essentially* discursive cognisers. We must have the doctrinal Belief that, over an infinite time, as we find the totality of conditions, we eventually transcend our own discursive nature,

[181] I provide a detailed reconstruction of the argument for doctrinal immortality in my Benzenberg (ms-g). I agree with Englert (2023: 377n) against Proops (2021: ch. 7.8) that the argument is *not* to be identified with the teleological argument for immortality that Kant defends in the B Paralogisms (B424-6).

[182] One way to think of the successive synthesis is as a convergent series that 'asymptotically [...] approximate[s]' the totality of conditions (A663/B691). This can be made precise. Let '$C_n x$' stand for the n-th R condition of x. We can represent the totality of R conditions of an appearance a by the infinite set $\mathcal{T}_a = \{C_1 a, C_2 a, C_3 a, ...\}$. Let $\mathcal{G}_{at}$ represent the set of R conditions of a that are given at time t. Insofar as it takes a discrete time t to complete an act of synthesis for an R condition of a to be given, we have $\mathcal{G}_{at} = \{C_1 a, C_2 a, ..., C_t a\}$. Thus, over infinite time, all R conditions of a will be given, that is, $\lim_{t \to \infty} \mathcal{G}_{at} = \mathcal{T}_a$.

and become non-discursive cognisers.[183]

If I am right, and we don't know that $P_{2.1}$ is false, then all four conditions of doctrinal Belief are satisfied. Reason, as governed by $P_{1.2}$, thus commits us to hold as a doctrinal Belief that there will be a totality of conditions; yet doing so does not lead to dialectical error. But how do we shift from $P_{2.1}$ to $P_{2.2}$? As noted above, I argue that this shift is largely due to transcendental realism, which 'represents external appearances [...] as things in themselves that exist independently of us and our sensibility' (A369). Insofar, however, as appearances exist independently of the conditions of our sensibility, they are not subject to a successive synthesis, but are given in a salient sense all at once.

Transcendental realism, so understood, explains the shift from $P_{2.1}$ to $P_{2.2}$ as a change in *content* and *justification*. The content changes because the totality of conditions, which is only the *future* vantage point of an infinite synthesis, is given *now* already. After all, appearances and their antecedent conditions are claimed to be given all at once without requiring a successive synthesis. Transcendental realism thus 'at once skips over' the successive series of antecedent conditions 'which, at least potentially, one can still come to know [*kennenlernen*] through continued experience' (A773/B801), and instead regards the 'investigation of nature, wherever it may be, as absolutely complete' (A690-1/B717-8).

The justification changes because, on the view of transcendental realism, $P_{2.2}$ isn't justified as doctrinal Belief, but is instead claimed to be known. After all, it is claimed that we can directly cognise totalities.[184]

[183] This is a strong claim, which I defend more fully in my Benzenberg (ms-b), (ms-e), (ms-h). But note that the claim draws independent support from other corners of Kant's thought. For example, in the Appendix to the Transcendental Dialectic, Kant argues that we follow the principle of '*specification*' (A658/B686) when we further determine our concepts, thereby moving from genera to species (see also Section 4.1). While it is true that for every concept there is a lower species (A655/B683), it seems that over infinite time we asymptotically approximate fully determined representations of singular objects. These representations, however, cannot be concepts, but must be intuitions, because they are singular (and immediate) (A320/B377); they also cannot be sensible, but must be intellectual, because they have their origin in the intellect. Thus, in the limit, we do acquire a form of intellectual intuition, which, however, would entail a form of non-discursive cognition.

[184] There is another reason why transcendental realism can claim to know $P_{2.2}$. Kant maintains that $P_{2.2}$ is analytically true of things in themselves (A498-9/B526-7). On this

Transcendental realism thus confuses the (sufficient) 'subjective' grounds of (doctrinal) Belief for (sufficient) 'objective' grounds of knowledge (9:66; see also A822/B850).[185] But only then do we run into dialectical errors, because only objective grounds of knowledge are preserved over entailment, thereby producing the antinomies. This also explains why Kant constantly claims that dialectical error only 'arises when we hold the *subjective* grounds of our judgements to be *objective*' (24:103; my emphasis).[186]

Indeed, we find independent textual support for this reading in the *Prolegomena*. There, speaking of error in theology, Kant makes a point that applies more generally to the sources of dialectical error. He argues that 'the dialectical illusion [...] arises from the fact that we [mis]take the subjective conditions of our thinking for objective conditions of the things themselves [*der Sachen selbst*] and a *necessary hypothesis* for the satisfaction of our reason for a *dogma*' (4:348, § 55; my emphasis). Now, doctrinal Belief is 'hypothetically necessary' to realise the final end of Reason (A823/B851),[187] and dogmas are 'alleged knowledge' (6:69n). The dialectical error thus arises when we confuse doctrinal Belief for knowledge, as does transcendental realism.

---

point, see also Jauernig (2021: 352-3). Insofar as transcendental realism identifies appearances with things in themselves, transcendental realism must hence also think that $P_{2.2}$ is true of appearances.

[185] On Kant's notion of subjective grounds of Belief, see also (A829/B858), (R2446, 16:371), (R2450, 16:373), (R2454, 16:375-6), (R2460, 16:379), (R2470, 16:383 (R2487, 16:391), (R2489, 16:391), (R2499, 16:395), (R2627, 16:442), (R2629, 16:443), (R2714, 16:480), (R2716, 19:482), and (24:198, 747, 852). On Kant's notion of objective grounds of knowledge, see also (A820-1/B848-9), (R2459, 16:378), (R2450, 16:374), (R2459, 16:378), and (24:198).

[186] On the confusion of subjective for objective grounds as the source of dialectical error, see also (2:416-7), (2:88), (A294/B351), (A297/B354), (A396), (A509/B537), (A666/B694), (A792/B820), (7:142). For a further discussion, see Grier (2001: 57-63), Willaschek (2018: 125-6), Kraus (2020: 245), and Proops (2021: 46-9).

[187] Kant generally contrasts necessary hypotheses with contingent ones. For example, he writes: 'Theoretical hypotheses are contingent. A practical one is necessary' (R2692, 16:472). In the case of contingent theoretical hypotheses – which will be the topic of Chapter 5 –, it must 'be possible to explain this [effect] from another [hypothesis]' (28:319). By contrast, '[a] [...] necessary hypothesis is when [...] no other [hypothesis] is possible' (R6236, 18:520). The distinction is also implicit in (A823-4/B851-2), (6:354), (R2688, 16:471), (R2692, 16:472), (R2679, 16:466), (R2688, 16:471), (R6236, 18:520), (21:346), (24:733), (24:750-1), (28:697-8), (28:793), (28:1183), and (28:1184). It is clear from Kant's examples, that necessary hypotheses are meant to be instances of Belief. Indeed, many examples are clearly instances of doctrinal Belief. See (R2689, 16:471), (R3830, 17:305), (R4113, 17:421), (R4580, 17:600,), (R6038, 18:430), (R6249, 18:529), (R6214, 18:502), (R6294,18:562), (21:151), (21:378), (23:43), (27:1457), (28:500), (28:1182), and (28:1259). I reconstruct Kant's taxonomy of hypotheses in my Benzenberg (ms-c).

Where does all of this leave us? In this section, I have defended a very liberal reading of the transition from P_1 to P_2. Specifically, I have argued that Reason, as governed by $P_{1.2}$, takes us all the way to $P_{2.1}$, which we are entitled to hold as a doctrinal Belief. Reason thus comes very close to endorsing $P_{2.2}$ itself. But very close is not all the way. It is only on the assumption of transcendental realism that we actually end up endorsing $P_{2.2}$, and thereby run into the dialectical errors of speculative metaphysics. This, then, is the main lesson of this section: even on the most liberal reading, Reason is not the source of dialectical error.

3.5 Illnesses of the Head

This section seeks to close all open brackets and to further elucidate the corrupting influence of sensibility on Reason. But first: if Reason is not the source of dialectical error, as I have just argued, why does Kant claim constantly throughout the *Critique of Pure Reason* that speculative metaphysics is in the nature of Reason? After all, we had seen in Section 3.2, that there is robust textual evidence in favour of the Rational Sources Account. Yet this evidence can be explained away. Note that Kant claims *human* Reason is dialectical, talking of a 'natural dialectic of *human* reason' (A669/B697; my emphasis).[188] And this is not a one-off claim; consider these three passages:

1. '*Human* reason has a special fate in a species of its cognitions: that it is troubled by questions that it cannot dismiss, [...] but which it cannot answer either, for they exceed all capacity of *human* reason.' (Aviii; my emphasis)

2. 'From the whole course of our critique, one will have become sufficiently convinced: [...] that *human* reason [...] is dialectical by the very nature of its direction' (A849/B877; my emphasis)

[188] See also (A5), (B9), (B21), (A297-8/B354), (A323/B380), (A372), (A407/B433-4), (A569/B597), (A584/B612), (A586/B614), (A625/B653), (A641-2/B669-70), (A702/B730), (A643/B671), (A669/B697), (A804/B832), (A831/B859), (A842/B871), and (A877/B849).

3. '[The] transcendental principle of [*Grundsatz*] pure reason [...] hastily postulates an unlimited completeness from the series of conditions in the objects themselves; [...] its sources [...] are deeply hidden in *human* reason.' (A309/B366; my emphasis)

But what is human Reason? I suggest that human Reason isn't a unique faculty – distinct from Reason proper. After all, human Reason, like all Reason, is governed by the Principle of Reason, which simply defines the faculty. Instead, I suggest that the phrase 'human Reason' is shorthand for 'Reason had by humans, i.e., beings that also have sensibility' – or as Schafer puts it: 'reason [considered] in the context of a sensibly conditioned finite subjects' (2019: 187). The dialectical error, then, does not stem from human Reason, qua Reason, but qua human. Indeed, Kant writes more generally that the dialectical error 'inevitably arises from *human nature*' (A316/B373; my emphasis).

The Sensible Sources Account – which I am defending – can therefore account for the seeming evidence of the Rational Sources Account. Speculative metaphysics is natural to human reasoners because human reasoners are subject to the corrupting influence of sensibility. Indeed, Kant seems to contrast human Reason with pure Reason, that is Reason considered in isolation from sensibility. In a Reflexion, I had already cited in Section 3.2, he notes that: 'If we had pure reason [...] we would never err' (R2246, 16:284).[189] All error, including dialectical error of speculative metaphysics thus stems from our sensible nature.[190]

But this raises a follow-up question: how exactly does sensibility induce error? The Sensible Sources Account must answer this question to have explanatory value. But there is a puzzle underlying the question. For it is the

[189] In view of these considerations, it seems that the *Critique of Pure Reason* is a misnomer and should have been better called the *Critique of Human Reason*, especially insofar as a critique must include a so-called 'discipline'.

[190] My reading is consistent with Grier's suggestion that Reason is responsible for the transcendental *illusion* (2001: 116), so long as the dialectical *error* is ultimately due to sensibility. Indeed, insofar as Reason commits us to $P_{2.1}$ as a doctrinal Belief, this would explain why $P_{2.2}$, which looks very similar, might have an illusory appeal to us.

same sensibility that is both a source of error and a necessary component of (discursive) cognition. Kant makes this point explicit: 'Sensibility, subordinate to understanding, as the object to which the latter applies its function, is the source of real cognition. But the same sensibility, in so far as it influences the act of understanding itself and determines it to judge, is the ground of error' (A294n/B351n). How can this be?

We can solve this puzzle by distinguishing between sensibility in its theoretical and practical varieties. Theoretical sensibility, as the lower faculty of cognition, '[gives] objects to us, and it alone provides us intuitions' (A19/B33); it is one of the 'two stems of human cognition' (A15/B28). By contrast, practical sensibility, as the lower faculty of desire, indicates the unique mode of conative affection that generates desires and inclinations. Kant blames practical sensibility for most of our cognitive mishaps, identifying 'inclination' as one of the main sources of 'error' and prejudice (24:403).[191] If we judge based on inclination, we hold something to be true merely '*because* it appeals to us' (24:167; my emphasis) – that's wishful thinking.

While the errors 'that arise from inclination are manifold' (24:162),[192] it seems that the dialectical error results from a very specific inclination, namely the desire for knowledge (*Wissbegierde*). Kant writes: 'The error is caused by two parts: ignorance and the desire to know, which actually causes [the error]' (24:817). Specifically, we have a desire for metaphysical knowledge. We want to know whether the soul is immortal; whether the world has a first cause; whether God exists; etc. This desire for knowledge, Kant thinks, makes us endorse $P_{2.2}$ as this principle promises answers to all these questions that

[191] See also (9:76), (R2519, 16:403), (R2524, 16:404), (R2531, 16:407–9), (R2540, 16:410), (R2550, 16:411), (R2571, 16:424), (24:165–6), (24:227), (24:426), (24:547–8), (24:641), and (25:46). Kant defines prejudice as the 'the propensity to persuasion' (24:547). See also (9:75), (R2529, 16:406), (R2530, 16:407), (R2541, 16:410), (R2550, 16:411), and (R2647, 16:411). Persuasion, in turn, is defined as 'the assent from merely subjective causes that are falsely taken to be objective' (24:547). See also (A820/B848), (5:461), (9:73), (R2459, 16:379), (R2465, 16:382), (R2486, 16:389), (R2488, 16:391), (20:297), (24:542), (24:547), (24:559), (24:647), (24:747), (24:849), (24:889–90), and (25:1451). Given this definition, it is little surprising that Kant showcases the dialectical error as the paradigmatic case of a persuasion (A821-2/B849-50). In my Benzenberg (forthcoming-a), I argue that all false (assertoric) judgements are instances of persuasions.

[192] Frierson (2014: 204) offers a comprehensive list of all cognitive disorders and defects according to Kant.

Reason puts to us.

Indeed, the desire for metaphysical knowledge leads us not only to endorse $P_{2.2}$, but also to endorse transcendental realism as the metaphysical view that underlies $P_{2.2}$. As I argued in section 3.4, transcendental realism holds that appearances can be given to us independently of the conditions of our (theoretical) sensibility. By implication, we don't have to go to the trouble of finding the totality of conditions through an infinite series of successive syntheses, but the totality is given to us all at once. Transcendental realism therefore promises instant gratification of our desire for metaphysical knowledge. That's why it appeals to us.

But what is the source of this desire for knowledge? Kant sometimes talks as if Reason itself is the source of this desire (A583/B611). But this *façon de parler* is misleading because Reason, as a higher faculty of *cognition*, cannot possibly be a source of a *desire*. Moreover, it is clear that the desire for knowledge does not come out of the higher faculty of desire since knowledge is not demanded by the moral law. The desire for knowledge must therefore originate in practical sensibility. I thus wholeheartedly agree with Heßbrüggen-Walter, who notes that 'the desire for knowledge (which is to be counted among the lower, sensible practical faculty) is responsible for the deviation of the mental powers that causes error' (2004: 217).

Insofar as the dialectical error results from sensibility, it can be said to be a pathology of the mind.[193] A pathology, after all, is a suffering (from the Greek 'pathos'), and suffering, for Kant, indicates the unique metaphysical dependence of receptivity or sensibility (A80-2/B106-8). But one could go further and argue that dialectical error is also a pathology of the mind in the medical sense of 'pathology' as 'illness'. Indeed, Kant foreshadow such a view in his often-overlooked essay *On the Illnesses of the Head* (1763). There, he argues that Reason, as it 'has become perverted in more general judgments' – presumably through sensibility – results in 'madness [*Wahnwitz*]' (2:264).[194]

[193] On the pathology of dialectical error, see also Butts (1997: 309) and Grier (2001: 104-5).

[194] In the *Anthropology*, Kant reassigns madness to the failure of judgement, and instead identifies 'lunacy [*Aberwitz*]' as the unique pathology of 'impaired [*gestörten*] reason' (7:215).

What is madness for Kant? To answer this question, let us first examine the etymology of the German term 'Wahnwitz', which literally translates as 'delusion-wit'. Wit, for Kant, is a talent of the mind (A710/B738; 4:393), which 'pairs [...] heterogeneous representation that by the measures of imagination [...] lie far apart' (7:220; see also 2:132). Wit hence is the ability to draw far-fetched connections, something Kant often associates with positive attributes, such as ingenuity and quick apprehension (2:260; 7:221; R466 15:191-3; see also Nehring 2015).

However, the connections of wit can also fail and lead to madness. Importantly, the madman has not lost all contact with reality; instead, he draws false conclusions from true premises: 'one judges quite wrongly from experiences that may be correct', which distinguishes madness from insanity – the insane person sees ghosts, the madman doesn't (2:267-8). The mad man makes correct judgments of experience, from which he then, however, 'strays into imagined lofty judgments of general concepts in an inconsistent way' (2:268).

Does that sound familiar? The dogmatist, who engage in speculative metaphysics, has all the symptoms of madness. In the 'sophisma figurae dictions' (A499/B528), they infer from a correct judgement of experience, namely that such-and-such appearance is conditioned, via $P_{2.2}$ to far-fetched conclusions about the general metaphysical nature of the soul, the world, and God. Moreover, in the case of the world, these inferences lead to antinomies, and so the dogmatists really judge 'in an *inconsistent* way' (2:268; my emphasis). The dialectical errors of speculative metaphysics manifest a mad mind.

Indeed, this reading is further corroborated in the *Critique of Judgement*, where Kant links madness to enthusiasm (*Schwärmerei*): 'enthusiasm [...] is the delusion of wanting to see something beyond all bounds of sensibility, i.e., of dreaming according to principles (raving with reason). [...] [*E*]*nthusiasm* is comparable to *madness* [...]; madness is an *illness* that deranges [the intellect]. (5:275; my emphasis; see also 7:203). The dogmatist is an enthusiast insofar as they dream in accordance with $P_{2.2}$, thereby claiming to have cognitive

access to the supersensible, transgressing all boundaries of sensibility. This enthusiasm or madness is an illness.

On my reading, then, dialectical error, where we *assert* rather than *seek* the totality of conditions, is never the result of Reason, but instead comes out of our sensibility-subdued human minds, and as such it is literally an illness of the head or mind.[195] Now, since madness is what gives rise to dialectical errors, only human reason is dialectical; pure reasoners, on the other hand, cannot go awry. We can then think of the *Critique of Pure Reason* as offering a *medicina mentis*, literally trying to heal us from our metaphysical madness and 'addiction for expansion [*Erweiterungssucht*]' (A786/B814), which result from our desire for knowledge. The *Critique* is a therapy to improve our mental health.

In this chapter, then, I have defended the Sensible Sources Account.[196] The dialectical error of speculative metaphysics, which implies that we have deviated from the Principle of Reason, is always the result of the corrupting influence of sensibility. Reason never deviates from its principle. By implication of OEM, this shows that the Principle of Reason (in its two variants in $P_{1.1}$ and $P_{1.2}$) must always be descriptive of the faculty of Reason. Reason is simply defined as the faculty that seeks to find totalities of conditions; that's what Reason does. Reason does not, however, assert these totalities. That's what we do as humans.

This solves the Problem of Normativity. For even if the Principle of Reason is normative for us, it simultaneously describes the operation of a mental power whose activities manifests in inner sense. That being said, we only observe the aggregate activity of Reason and sensibility (vector $\vec{P}_{1+2}$ in Figure 2) – Reason is a zombie power, after all –, which motivates the

[195] Kant goes even further, suggesting that a mad mind originates in the body: 'I have only paid attention to the phenomena of the [ill] mind, without wanting to identify its source, which actually lies in the body and may have its headquarters more in the digestive parts than in the brain' (2:270; my emphasis). Perhaps, then, Kant ultimately holds a 'Digestive Sources Account', on which metaphysics is the result of, say, constipation.

[196] Kant amends his account of faculty failure in later writings. In the third *Critique*, for instance, he provides a developmental account of our faculties, implying that faculties can fail when they are not fully developed (5:283). This development account also fits well with Kant in the historical writings, like *Universal History* (1784) and *Conjectural Beginning* (1786).

Problem of Perfection. However, we may hope to solve this problem if we are able disaggregate this composite effect, and 'isolate reason' and its principle (A305/B362). To understand how this can be done, we will first need to answer the Material Question, which is what I do in Chapters 4 and 5. I then return to the Problem of Perfection in Chapter 6.

4

One Principle, Two Formulas

Ontological Symmetry, which I have defended in Chapters 2 and 3, motivates Epistemic Symmetry. For if mental powers are powers of inner nature, just as physical powers are powers of outer nature, then we should expect that we can come to know mental powers, including Reason, just as we can come to know powers of nature. But this raises a general question: how do we come to know powers of nature and their respective laws? This will be the guiding question across Chapters 4 and 5. In Chapter 6, I then apply Kant's general account to answer the Epistemic Question: how can we come to know that we have Reason?

Kant develops his epistemology of powers in the Appendix to the Transcendental Dialectic. There, he argues that we come to know powers and their laws by seeking to complete the 'systematic unity of manifold powers, whereby special laws of nature standing under more general ones' (A651/B679). This system of powers and laws, Kant suggests, is part of a more general effort to complete 'the systematic unity [...] of the manifold cognitions of the understanding' (A648/B676). This general effort, in turn, is demanded by Reason: 'that a certain systematic unity [...] must be sought is a [...] principle without which there would be no use of reason' (A652/B680).

But this creates a textual problem. In Sections 3.3 and 3.4, we have seen that the Principle of Reason, which I had identified with P_1, demands that we seek the totality of conditions. Now, however, we have another Principle of Reason that demands that we seek a systematic unity. How do these two

principles demands hang together? The underlying concern is serious. If, as I have argued in Section 2.1, powers are individuated by their principles, and if we have two Principles of Reason, then we end up with two powers of Reason, not one. In addressing this concern, we also tackle the Material Question: what is the exact content of the Principle of Reason?

In this chapter, I argue that the two principles are just two formulas of the same identical Principle of Reason. Here is the plan: first, I articulate the two principles in more detail (§ 4.1), and I distinguish three different senses of identity (§ 4.2). After discussing and rejecting previous attempts to establish identity (§§ 4.3 and 4.4), I develop and defend my own solution (§ 4.5). In answering the Material Question, we see that seeking the totality of conditions and seeking systematic unity amount to the same thing; and further, that by seeking the totality of conditions, reason can come to know powers and their laws, including its own power and its own laws.

4.1 Totality and Systematicity

I have already said much about the first principle, P_1, in Section 3.3. But here are two more clarifications, the first being that it only applies to '*given* cognitions' (A307/B364; my emphasis). What does it mean for cognitions to be given? Kant answer this question with regard to concepts, which are cognition in the broad sense of 'relating to an object' (A320/B377). Kant contrasts 'given' concepts with 'made' concepts that are stipulated (9:93),[197] as well as 'possible [...] concepts' that we don't yet possess (A652/B680). Given cognition, therefore, simply are those cognitions that we possess at a given point in time and that are not stipulated.

The second clarification concerns the conditioning relation. As I had noted in Section 3.4, $P_{2.1}$ and $P_{2.2}$ are about real conditioning relations, R conditions, that obtains among appearances – such as spatio-temporal,

[197] And in the *Jäsche Logic*, Kant writes: 'All concepts, as to matter, are either given (conceptus dati) or made (conceptus factii)' (9:93). See also (A730/B759) as well as Anderson (2015: 334).

mereological, and causal conditioning. Abstracting from these examples, we can isolate three general properties that conditioning relations have for Kant. (i) Conditioning relations are *transitive*, i.e., for all A, B, and C: if A is a condition of B and B is a condition of C, then A is a condition of C. (ii) Conditioning relations are *asymmetrical*, i.e., for all A and B: if A is a condition of B, then B is not a condition of A. And so, (iii) conditioning relations must be *irreflexive*, i.e., for no A: A is the condition of A.[198]

However, the conditioning relation specified in P_1 (including its two variants $P_{1.1}$ and $P_{1.2}$) cannot be R conditioning because it doesn't obtain between appearances, but between cognitions. The conditioning relations that obtain between cognition – formal or real – are commonly called logical conditioning relations, or simply 'L conditions'.[199] I will say more about the nature of L conditions in Sections 4.3 to 4.5. For now, note that (i) they are meant to track R conditions, and in doing so (ii) they also satisfy the general properties of transitivity, asymmetry, and irreflexivety.[200] With that out of the way, we can now rebrand P_1 as the 'Formula of the Totality of Conditions' ('FTC' for short):

Formula of the Totality of Conditions

(FTC) For all x: if x is a given cognition and x is L conditioned, then find all y, such that y is a cognition and y is an L condition of x.

[198] According to Watkins, 'Kant seems to operate with a generic notion of real conditioning that involves an *asymmetrical* [and] *transitive* [...] [and thus *irreflexive*; C.B.] relation of metaphysical dependence' (2019a: 5; emphasis added). See also Stang (2018: 81-3), Willaschek (2018), and Watkins & Stratmann (forthcoming).

[199] To be clear, Kant never explicitly distinguishes between L conditions and R conditions. This distinction, however, is implicit in his discussion of conditions, both in the Introduction and the Antinomy chapter. The distinction is explicated by interpreters, like Watkins, who distinguishes between 'logical conditioning relations that obtain between cognitions', on the one hand, and 'real conditioning relations' that obtain 'between objects', on the other hand (ms). For a similar distinction, see Watkins (2019a), (2019b), Willaschek (2018: 71-83, 133), Kraus (2020: 238), Watkins & Stratmann (forthcoming), and Chaplin (ms).

[200] Stang suggests that 'it is plausible that the structural principles of grounding [or conditioning; C.B.] in general [also] apply to' the conditioning relations that hold between cognitions, meaning that they are 'irreflexive, transitive, and asymmetric' (2018: 88).

Of course, the main aim of this chapter is to show that FTC is really just another *formula* of the same Principle of Reason that Kant also specifies in the Appendix. And, of course, FTC is only meant to be normative for us, but descriptive of Reason. Figure 3 illustrates the activity prescribed by FTC. Say, we start with a given cognition C_0. To find its totality of L conditions (indicated by the arrows), we first seek C_0's proceeding L condition, which is C_1; once found, C_1 is given, and we seek its proceeding L condition, which is C_2; and so on. Since L conditions are transitive, all cognitions in the series will also be L conditions of C_0. So, by following the 'series of all conditions' (A497/B525), we effectively seek the totality of L conditions of C_0.

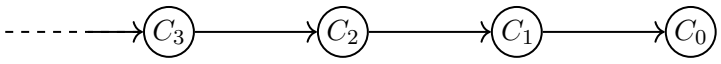

Figure 3: series of L conditions

I now turn to the second formula from the Appendix, which I will call the 'Formula of Systematic Unity' ('FSU' for short). As a first approximation, FSU demands that we complete the 'systematic unity [...] of the manifold cognitions of the understanding' (A648/B676), specifically of 'all possible empirical [cognitions]' (A652/B680). Again, what is a 'demand' for us (A656/B684; A699/B727), is descriptive of Reason. Reason simply is the faculty that seeks to complete the systematic unity of the understanding's cognitions. But wherein does this systematic unity consist? What are its structural features?

In what follows, I identify six features of systematic unity. The first feature is that the systematic unity, like the totality of conditions, consists of the understanding's *cognitions*, specifically its empirical cognitions. Specifically, Kant holds that we ought to seek the systematic unity for two kinds of cognitions: for 'empirical concepts' (A652/B680) – which, as mentioned, are cognitions in the sense of being related to an objects –, and for (our cognition of) 'laws of nature', whereby we also seek a 'systematic unity of the manifold powers' constituted by these laws (A651/B679). What concepts and laws

have in common is that they are both general representations, which come in varying degrees of generality.[201]

This gets me to the second feature. The system of cognitions constitutes a hierarchy, with more general cognitions, *genera*, towards the top and more specific cognitions, *species*, towards the bottom (A655/B683).[202] Kant's notion of genera and species is informed by 'the conjunction model of concepts' (de Jong 1995: 623), which he endorses. On the conjunction model, concepts are conjunctions of differentiae or marks.[203] Since two concepts are identical, or 'reciprocal concepts' (9:98), iff they contain the same marks, we can represent concepts as sets of marks.[204] For example, the concept <human> is the union of all those marks in the concept <animal> plus the mark 'rational', i.e., <human> $=$ <animal> $\cup$ rational.

In light of this conjunction model, Kant defines a species as a concept that is 'contained under' its genus, and a genus as a concept that is 'contained in' its species (24:911).[205] While Kant articulates both containment relations in terms of (proper) parthood (9:95; 24:910), we can represent both of them via the proper subset relation. A concept, C_1 is *contained in* a concepts C_2 iff C_1 is a proper subset of C_2 (i.e., $C_1 \subsetneq C_2$); and conversely, a concept C_1 is *contained under* a concept C_2 iff C_2 is a proper subset of C_1 (i.e., $C_2 \subsetneq C_1$). So genera are proper subsets of their species. For example, <animal> is the genus of <human> because <animal> is a proper subset of <human>.

---

[201] While concepts and laws cannot feature in the same system, it is an open question in Kant scholarship how the two systems relate to each other (if at all). For example, does the system of concepts determine the system of laws, or vice versa? While this question is important, it exceeds the scope of what this thesis can hope to address.

[202] Kant himself invokes vertical terminology when discussing genera and species: 'The *higher* concept, in respect to its *lower* one, is called genus, the *lower* concept in regard to its *higher* one species' (9:96; my emphasis). For a discussion of the hierarchy in the system of cognition, see Kitcher (1994: 259), Geiger (2003: 275), Abela (2006: 413), and Willachek (2018).

[203] See also in (24:261), (24:755), and (24:912). For a further discussion of Kant's conjunction model, see de Jong (1995, 2010), Anderson (2015: 51, 55), Watkins (2019a: ch. 10), and Currie (2022: ch. 6). Kant was heavily criticised for the conjunction model by Frege, who writes: 'Kant seems to think of concepts as defined by giving a simple list of characteristics in no special order' (1879/2020: 100). This criticism might not be fair, however. De Jong (1995: 623) demonstrates how Kant adopts the model from Leibniz, Locke, Wolff, Crucius, and Reimarius.

[204] Marks also cannot be repeated in a concept.

[205] On Kant's notion of genera and species, see also Anderson (2015) and Currie (2022).

As a brief aside, Kant also defines the extension (*Umfang*) and content (*Inhalt*) of a concept in terms of containment relations. The extension of a concept is that which is contained under it, and the content of a concept is that which is contained in it (A654/B682). Since the two containment relations are inversely related, this leads to the inverse relation between extension and intension, which Peirce christened 'Kant's Law' (1932: 76-78).[206] Importantly, this allows us to say that the species is part of the extension of the genus, and the genus is part of the content of the species.

But Kant is committed to the idea that not only concepts, but also laws of nature can stand in genus-species relations. And while laws cannot, strictly speaking, contain each other – for they aren't conjunctions of marks[207] –, Kant gestures at an analogical definition: 'reason presupposes systematic unity among manifold powers, for particular natural laws *stand under* more general ones' (A650/B678; my emphasis). The 'stand under' relation, though not further defined, seems to mimic the 'contained under' relation. So we can say that a law L_1 is a genus of a law L_2, and so L_2 is a species of L_1, iff L_2 stands under L_1. The broader point has philosophical merit: some natural laws clearly are more general than others.[208]

The third feature is that species divide the extension of their genus in a way that is exclusive and exhaustive. Anderson argues that this follows from Kant's commitment to the *dictum de omni et nullo*, according to which 'what belongs to the genus, belongs to everything falling under it, and what is excluded from the genus, is excluded from everything that falls under it' (2015: 125).[209] Moreover, the division must be (at least) dichotomous

[206] Kant explicitly asserts the inverse relation between extension and intension in (9:95), (24:755), and (24:911). For a further discussion of Kant's Law, see Heßbrüggen-Walter (2004), Lu-Adler (2012: 72-75), Anderson (2015: 252), Kraus (2020: 135), and Currie (2022: 176).

[207] It could be argued, however, that natural laws can in some sense echo the conjunction account of concept. For just as we can get from a genus concept to its species by adding a differentia, so we can get from a genus law to its species by adding a minor premise in a hypothetical syllogism.

[208] Although nowadays this generality relation between laws might be specified in terms of determinables and determinates. See, for example, Armstrong (1996).

[209] Anderson (2015: 59; 114-130) argues that this feature aligns Kant's system with that of Wolff and against that of Leibniz, since not all combinations of marks will constitute concepts.

(9:147).[210] For in the case of concepts, we can add either the *affirmation* or the *negation* of a differentia to a genus. By adding 'rational' to <animal> we get <human>; by adding 'non-rational', we get <beast>.

The fourth feature is that the system has one highest genus: '[A]ll the manifold genera are only divisions of a single supreme and universal genus' (A659/B687; see also 24:259). In the case of concepts, Kant identifies this highest genus with a '*conceptus summus*' (9:97), which he defines as follows: 'A highest genus is not contained under another concept; it is itself not a particular concept, that is, it has no further parts' (24:755; see also 24:729; 24:845). Furthermore, Kant claims that this highest concept is the 'concept of an object in general' (A290/B346),[211] which we obtain by abstracting from all marks and which can thus be represented as the empty set, $\varnothing$.[212]

While the highest genus is already given in the system of concepts, this is not the case in the system of laws. As noted in Section 2.5, the highest genus in the system of laws would constitute a 'basic power [*Grundkraft*]' from which all powers and their laws are derived (A649/B677).[213] Basic powers, especially those of the mind, were hotly contested in Kant's times. Wolff and Baumgarten had argued that there was a basic power of the mind, given in the power of representation. Crucius and Tetens had denied this, arguing that there were many irreducible mental powers.[214] By arguing that

[210] Kant's general account of concept division is discussed by Watkins (2013: 286-7) and Anderson (2015: 60). For simplicity's sake, I will assume that all polytomous distinctions can be reduced to dichotomous ones; but nothing in my argument depends on this. All I need is that divisions are *at least* dichotomous.

[211] On the highest genus, see also de Jong (1995: 625), Heßbrüggen-Walter (2004: 209), Watkins (2013: 284-5), Anderson (2015: 60, 210, 368), and Kraus (2020: 234-5).

[212] Breitenbach (2021: 293-6) denies that there can is a highest genus for Kant on philosophical grounds. Yet without a highest genus, we cannot show that the system is connected, as Kant, however, claims: 'there are not different original and first genera, which are isolated and separated from each other (by an empty space)' (A659/B687). For a further discussion of Breitenbach's position, see my Benzenberg (ms-e).

[213] Importantly, the basic power is not to be identified with the soul, as Kraus claims (2020: 210-1). As I had shown in Section 2.1, Kant argues against Baumgarten that the bearer of a power is not itself a power: 'the substance *is* no power, but *has* a power' (28:25). Likewise, we should not think of the *concept* or idea of the soul as a highest genus – as Kraus (2020: 218) claims – because the highest genus is an empty concept, devoid of all content. Moreover, the concepts in the system are empirical concepts, yet the idea of the soul is a priori.

[214] For example, see Wolff's *Deutsche Metaphysik* (1751/1983 §§ 745, 747), Baumgarten's *Methaphysica* (1750: § 744), and Crucius's *Entwurf* (1745/1964: §§ 39, 444). On the historical background to Kant's discussion of basic powers, see also Heßbrüggen-Walter (2004: ch. 4.3),

we must seek a basic mental power that, however, is not yet given, Kant aims to find a middle ground between these two views. This point comes out well in the following passage:

> Whether all powers of the soul can be derived from one basic power, or whether several are to be assumed; we must certainly say: because the soul is a unity [...] it is clear: *that there is* only one basic power of the soul, from which all changes and determinations arise. [...] But that is a completely different question: *whether we are able to derive* all the actions of the soul, and the various powers and faculties of the same, from one basic power (28:261-2; my emphasis)[215]

The fifth feature, which will become important later on, is that 'there can be [no] lowest [...] species' in the system (9:97). That's because 'reason demands in its entire extension that no species be regarded as [...] the lowest; for since each species is always a concept that contains in itself merely what is common to different things, [...] it must always contain other concepts, i.e., subspecies, under itself" (A655-6/B683-4).[216] The argument Kant is making here is that a lowest species would have to be a singular representation, but concepts, by their very nature, are general representations (A320/B377). The same argument generalises to laws, which too are general representations; so there also can't be a lowest law.

The sixth and final feature is that there is also '[no] next species' (9:97). This is to say that 'there are no species or subspecies that are proximate (in the concept of reason), but in-between species [*Zwischenarten*] are always possible' (A659/B688). Kant claims that between every genus and one of its species there must be an in-between species, and so that all species 'touch each other and do not allow for a transition to each other by a leap' (A660/B688). The

---

Dyck (2014: 202-5), Frierson (2014: 7), and Hatfield (2018: 66).

[215] On basic powers, see also (2:416), (5:46-7), (24:82), (28:29), (28:210), (28:145), (28:431-2), and (29:770-82).

[216] See also (A659/B687), (9:59), (9:99), (R2293, 16:303), (24:569), (24:755), and (24:911). Another way to reconstruct the argument is to say that we could always affirm or deny a further differentia of a given concept. As argued by de Jong (1995: 626) and Watkins (2013: 285, 294), Kant's claim that there are no lowest species and thus no fully determined concept is directed against Leibniz, Wolff, and Meier. For a further discussion of Kant's claim that there is 'no lowest species', see also Heßbrüggen-Walter (2004: 123; 251), Anderson (2015: 66-71), and Kraus (2020: 234; 241).

system of concepts and laws is thus not only infinitely extendable towards the bottom, but because there is no next species, it is also infinitely dense.

Taken together, these six structural features exhaustively characterise the systematic unity of cognition, as Kant thinks it is prescribed by Reason.[217] The system is illustrated in Figure 4. Each vertex in the graph represents a cognition: C_0, C_1, C_2, ..., and so on. The edges represent the genus-species relations, with higher concepts being the genera of lower ones. The genera are divided into (at least) two species. The highest vertex, C_0, represents the highest genus from which all other cognition derives. The dashed edges towards the bottom indicate that there is no lowest species; and the dashed edges in between indicate that there is no next species.

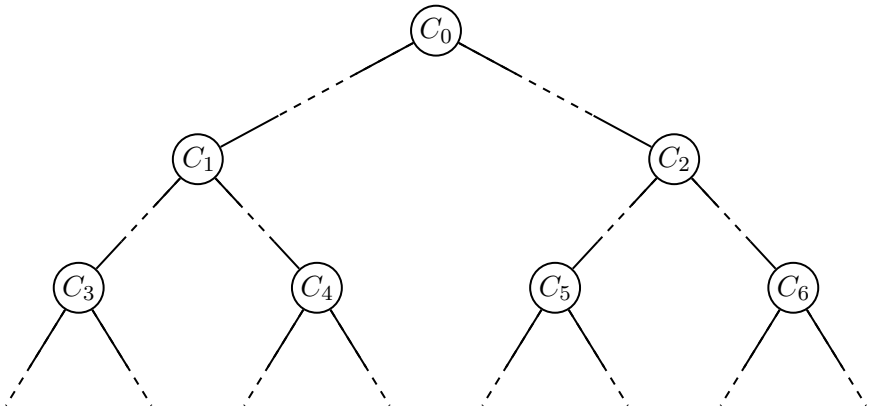

Figure 4: systematic unity of cognitons

[217] Here is a fun fact about the system thus described: it cannot be axiomatised. Let me explain. An axiomatisation of Kant's claim that divisions must be (at least) dichotomous is as follows: for every cognition x, there are at least two further cognitions y and z, such that y and z are species of x, and they are not species of each other. But this doesn't suffice, because it allows for a model in which there is a fourth cognition, u, such that x and y are species of u, but u does not divide z. The only way to axiomatise the dichotomous divisions is to define a 'next species' relation: x is the next species of y iff x is a species of y and there is no z such that y is the species of z and z is the species of y. We can then axiomatise the dichotomous divisions as follows: for every cognition x, there are at least two further cognitions y and z such that y and z are *next* species of x, and they are not species of each other. But this axiom contradicts the straightforward axiomatisation of Kant's claim that there is no next species: for every cognition x and y, if x is the species of y, then there is a cognition z, such that x is the species of z and z is the species of y. In my Benzenberg (ms-e), I show that the system of cognition is therefore best defined as the limit of an iterative, atemporal construction that inserts further intermediate species at each step of the construction.

Reason demands, in line with FSU, that we complete this systematic unity for our empirical concepts and for our (cognitions of) natural laws. This is as much an infinite task, as it is an infinite task to find the totality of L conditions. And just as $P_{1.2}$ presupposes $P_{2.1}$ as doctrinal Belief, Kant claims throughout the Appendix that the demand of FSU presupposes 'transcendental principle[s] [*Grundsätze*]' which say that nature itself is structured in such a way that our cognitions of nature can form a systematic unity (A648/B677).[218] Gava (2018) has shown that these principles are also to be considered as doctrinal Belief. I agree with him and won't pursue this point here.[219]

While the above gives us a good approximation of FSU, it is not yet the final version. How do we realise the systematic unity of cognition? Rather than aimlessly seeking cognitions, Kant suggests that we start from those cognition that are already given to us, and the follow three sub-principles: the principle of '*homogeneity*' tells us to look for higher genera; the principle of '*specification*' demands that we search for lower species; and the principle of '*continuity*' states that we ought to look for in-between concepts (A657-8/B685-6).[220] Once these genera, species, and in-between species are found, they too are given, and we repeat the process. The three principles can be stated more precisely as follows:

Principle of Homogeneity

(PH) For all x and y: if x and y are given cognitions and x isn't a genus of y and y isn't a genus of x, then find a cognition z, such that: z is a genus of x and y.

[218] See also (A650/B678), (A653/B681), (A656/B684), and (A660-1/B688-9).

[219] I agree with Willaschek, who maintains that 'the distinction between the logical and transcendental principles in the Appendix clearly parallels that between [P_1] and [P_2]' from the Introduction (2018: 109).

[220] For an extended discussion of the principles of homogeneity, specification, and continuity, see Wartenberg (1979: 412), Krauser (1987: 179), Guyer (1990: 23), Schiemann (1992: 297-8), Kitcher (1994: 260), Caimi (1995: 316), Grier (1997: 3), Geiger (2003: 276-7), Goldberg (2004: 406), Abela (2006: 411-413), Rajiva (2006: 121), Ostaric (2009: 163), Rauscher (2010: 291-5), Pasternack (2011b: 417), Banham (2013: 19), Watkins (2013: 287-289), McLaughlin (2014: 561-2), Anderson (2015: 371-2), Arias-Albisu (2017: 15), Mudd (2017: 81), and Kraus (2020: 232).

Principle of Specification

(PS) For all x: if x is a given cognition, then find two cognitions y and z, such that y and z are species of x, but y isn't a species of z and z isn't a species of y.

Principle of Continuity

(PC) For all x and y: if x and y are given cognitions and x is the genus of y, then find a cognition z, such that z is the species of x, and z is the genus of y.

Let me illustrate these three principles in relation to concepts. Suppose we start with two given concepts, <human> and <beast>. PH would tell us to look for their common genus, such as <animal>. PS would tell us to look for further species, which in the case of <animal> might be <fish> and <bird>. And PC would ask us to fill in the gaps; for example, having found <animal> to be the genus of <human>, we might find <mammal> as an in-between species. The same point applies to laws, where we seek more general laws – as modern physicists do when they try to unify the four fundamental forces (or powers) of nature –, for more specific laws, and everything in between.

Note that PH, PS, and PC have the form required of principles that govern a faculty. They specify an antecedent condition c, the given cognitions, and a subsequent activity φ – finding higher genera, lower species, and in-between species. However, PH, PS, and PC only specify finite activities. FSU tells us we must apply them iteratively, or so is the idea. This idea can be made premise by defining the notion of 'systematic connection'. A cognition x is *systematically connected* to a (set of) given cognition(s) y iff x can be reached from y by iteratively applying PH, PS, and PC. Given this notion of systematic connection, I propose to formulate FSU as follows:

Formula of Systematic Unity

(FSU) For all x: if x is a given cognition, then find all y, such that y is a cognition and y is systematically connected to x.

The expectation is that by following FSU, starting from our set of given cognitions, we will, in the limit, bring about the system of all possible empirical cognitions that I have described earlier. As Willaschek puts it, when we 'search for higher genera and lower and closer species [...], we asymptotically approach a completely systematic picture' (2016: 231). And this is plausible: by looking for higher genera, lower species, and in-between species, we slowly but steadily converge onto the complete system. While this will turn out to be true, we will see in Sections 4.3 to 4.5 that this convergence is not as trivial as it initially appears.

In closing this section, I make one final point. FSU also applies to laws of nature, including the laws of inner nature that govern our mental powers. So it is by following FSU that we discover mental powers and their laws, which is precisely what we do in empirical psychology. Kant states that '[empirical] psychology amounts to this: to derive, as far as possible, various powers, which we know only by observation, from basic powers' (28:564). Empirical psychology, then, doesn't just explain this or that inner state in terms of mental laws, but tries to explain the '*sum total* of inner perceptions under natural laws' (7:141; my emphasis), thus completing the systematic unity of all mental powers.

4.2 Three Types of Identity

Now that we have a better grasp of FTC and FSU, we can explore their identity. Before delving into various strategies to demonstrate that FTC and FSU are identical, I want to use this section to achieve two upstream objectives: (i) to thoroughly motivate why we should consider FTC and FSU to be identical principles in the first place, and (ii) to distinguish three different senses in which principles of faculties can generally be said to be identical. As we will see, the identity of principles is much less straightforward than the identity of concepts, where two concepts are identical iff they contain the same marks.

So why should we consider FTC and FSU to be identical principles? I have already given the main philosophical motivation at the beginning of

this chapter. Kant introduces both FTC and FSU as principles of reason, calling FTC the 'proper principle of reason in general' (A307/B364) and FSU a 'principle without which no use of reason would take place' (A652/B680). But if, as I argued in Section 2.1, powers and faculties are individuated by their governing principle, then FTC and FSU must be identical, or else we end up with two separate powers of reason.

The identity of FTC and FSU is further motivated by the overall arc of the Transcendental Dialectic. In the Introduction, Kant explains how Reason governed by FTC can be corrupted by sensibility, which is why we endorse $P_{2.2}$. After deriving the ideas of Reason in Book One, Kant then goes on in Book Two to show how $P_{2.2}$ leads to the dialectical errors in rational psychology, cosmology, and theology. In the Appendix, however, Kant tries to redeem Reason by showing it has a 'proper use' as defined by FSU (A642/B671). This story of redemption only makes sense if the power and its principle remain the same throughout.

Aside from these structural considerations, there is also direct textual evidence to support the identity of FTC and FSU. In the Introduction, when Kant introduces FTC, he writes: 'the proper principle of reason in general (in the logical use) is: to find the unconditioned to the conditioned cognition of the understanding, *by which the unity of the same is completed*' (A307/B364; my emphasis). And shortly thereafter, Kant also states that in seeking the totality of conditions, we 'bring the highest possible *unit of reason into our cognition*' A308-9/B365; my emphasis). These passages suggest that by satisfying the demand of the FTC, we also satisfy the demand of the FSU, directly supporting at least one direction of identity.

Lastly, it is not the first time that Kant has offered different formulas for the main principle of a faculty. After all, the moral law, which is the 'supreme practical principle' governing pure practical reason (4:428), and which for us takes the form of a categorical imperative, also comes in different formulas – the formula of universal law or natural law, the formula of humanity, and the formula of the kingdom of ends. But Kant insists that '[t]he three mentioned ways to imagine the principle of morality are basically only many

formulations of the same law' (4:436).[221] This set-up, I suggest, provides a template for thinking about the relationship between FTC and FSU: they are only two formulations of the same principle, the Principle of Reason.

I therefore agree with Willaschek, who, referring to FTC and FSU, states that 'it is obvious that [Kant] has the same principles in mind' (2018: 108).[222] But I also concur with Breitenbach who laments that 'Kant does not explicate how these principles relate' (2021: 291). The two principles should be identical, or else Reason would be servant to two masters, and thus disunified. Yet much like the formulas of the categorical imperative, it is not clear whether or how they can be identical. On the face of it, the demand for a totality of conditions and the demand for systematic unity seem to have little in common. Figures 3 and 4 look nothing alike.

However, before we can determine whether FSU and FTC really are identical, we must first understand what it actually means for two principles of a faculty to be identical. This brings me to the second goal of this section. My main suggestion is that we can think of FSU and FTC as functions that map the set of all given cognitions either onto the set of all L conditions (i.e., the totality of L conditions) or onto the set of all systematically connected cognitions. Let us call the input of a principle '$\mathcal{I}$' and the output '$\mathcal{O}$'. Given these constraints, we can distinguish three increasingly demanding ways in which principles can be said to be identical:

Extensional Identity

For any two principles x and y: x and y are *extensionally identical* iff, for some input $\mathcal{I}$, x and y yield the same output, $\mathcal{O}$.

Intensional Identity

For any two principles x and y: x and y are *intensionally identical* iff, for any input $\mathcal{I}$, x and y yield the same output $\mathcal{O}$.

[221] A helpful discussion of the identity of the various formulas of the categorical imperative can be found in Pollock (1974: 261), Wood (1999: 187) and Geiger (2015: 396).

[222] Kraus likewise notes: 'In the Appendix, Kant then explicates the "completeness" that we approximate by means of an idea as the *systematic unity* of the manifold in cognition' (2020: 196). See also Deleuze (1983/2008: 17).

Hyperintensional Identity

For any two principles x and y: x and y are *hyperintensionally identical* iff,

x and y prescribe (or describe) the same operation.

The different types of identity are best illustrated by examples from mathematics. The functions 'add-10' (i.e., $f(x) = x + 10$) and 'add-12-then-subtract-2' (i.e., $f(x) = x + 12 - 2$) always give the same output from the same input, so they are intensionally identical. However, there is a meaningful sense in which they are not the same operation, and therefore not hyperintensionally identical. Further, the operations 'add-2' (i.e., $f(x) = x + 2$) and 'multiply-by-2' (i.e., $f(x) = 2x$) are clearly not intensionally identical, but they give the same output for the input value 2, which makes them extensionally identical in the sense above.

With these definitions, we can now ask: in what sense, if any, are FTC and FSU identical? Two things are clear. First, FTC and FSU are not hyperintensionally identical, because they prescribe two very different operations – finding the totality of conditions and finding all systematically connected cognitions. Second, FTC and FSU are extensionally identical for irrelevant inputs. For if no cognition is given as input, both FTC and FSU will not be triggered and thus will not return any cognition as output. Likewise, if all possible cognitions are given as input, both FTC and FSU will return all possible cognitions as output.

Therefore, to show that FTC and FSU are meaningfully identical, we must either show that they are intensionally identical, or at least that they are extensionally identical for a relevant input. In the following three sections, Section 4.3 to 4.5, we will see that there is no way to establish an intensional identity between FTC and FSU. However, I will argue that the two principles are extensionally identical for a relevant input. That is, there exists an input, for which the set of all L conditions and the set of all systematically connected cognitions are identical. Before we get there, however, let us look at failed attempts to establish the identity of FTC and FSU.

4.3 Upwards to Higher Genera

In order to establish any identity between FTC and FSU, we need to connect the notions of 'being an L condition' and 'being systematically connected', which in turn is defined by the genus-species relation. A promising way of doing this is to say: *x is an L condition of y iff x is the genus of y*. This is the main claim of what I'll call the 'Upward Solution' – 'upward' because we need to go upward to higher genera to look for further L conditions. The hope behind this solution is that by looking for the totality of L conditions we will thereby find all the other genera and thus slowly but surely complete the systematic unity.

The Upward Solution has garnered support in the literature. Most prominently, Willaschek claims that the condition of a cognition must be more general, and hence its genus: 'The Logical Maxim is a subjectively valid prescriptive principle of the logical use of reason. It directs us to find, for every cognition C_1 that is [...] conditioned, a *more general* cognition C_2' (2018: 128; my emphasis).[223] Moreover, Willaschek suggests that, '[i]deally, the result of this procedure' – i.e., of finding the totality of more general conditions – 'would be a complete hierarchical system of cognitions' (ibid.). That's the Upward Solution in a nutshell.

And indeed, the main claim of the Upward Solution – that genera are L conditions of their species – enjoys exceptionally strong textual support. When Kant speaks of conditions, he often implies that they must be genera. For example, he writes that the 'ground [or condition] always *contains* the consequences *under* itself' (24:910; my emphasis). Remember that 'containment-under' defines the genus-species relation for concepts. And with genera being more general than their species, Kant also notes that 'reason in its

[223] Other defenders of the Upward Solution are Guyer (1990: 19), Anderson (2015: 280), and Kraus (2020: 238). See also and Willaschek (2018: 23, 35, 46, 56, 63, 70, 158). Watkins claims that, according to Kant, 'reason in its logical use aims for comprehensiveness by seeking a single system of syllogisms that would contain all of the conditions from which a given conditioned cognition would logically follow' (ms). Both Willaschek (2018: 61) and Watkins (ms) suggest that Kant adopts this line of thought from his predecessors, Leibniz, Wolff, Baumgarten, and Meier. In fact, see Meier (1752: § 413).

logical use seeks the *general* condition' (A307/B364; my emphasis). Consider also these two passages from the Introduction:

1. [W]e see that reason, in inferring, seeks to bring the greatest manifold of cognition of the understanding to the smallest number of principles (*general conditions*), and thereby affect the highest unity thereof. (A305/B361; my emphasis)

2. [It's] a logical demand [of reason], in the ascent to ever *higher conditions*, to approach their completeness and hence to bring the highest possible unity of reason into our cognitions. (A309/B365; my emphasis)

Moreover, the genus-species relation satisfies all three conditions that Kant applies to any conditioning relation. First, it is *transitive*: if A is the genus of B and B is the genus of C, then A is the genus of C. Second, it is *asymmetrical*: if A is the genus of B, then B cannot be the genus of A. And so, third, it is also *irreflexive*: no A can be the genus of itself. For concepts, these properties are entailed by the proper subset relation, which I used to define the containment relation, which Kant uses to define the genus-species relation. Plausibly, these same properties also apply to the 'stand under' relation that defines the genus-species relation for laws.

In view of the Upward Solution, we can merge Figures 3 and 4 to obtain Figure 5 below. The vertices in the figure still represent all possible cognitions, but the edges now represent *both* the genus-species relation *and* L conditions (as indicated by the arrows). So, for example, C_1 is both the genus and the L condition of C_3. And while the arrows are pointing downwards, we must go upwards to find further conditions. Now, the Upward Solution suggests that by seeking the totality of conditions we would also complete the system, and vice versa. Ideally, then, this would give us an intentional identity between FTC and FSU.

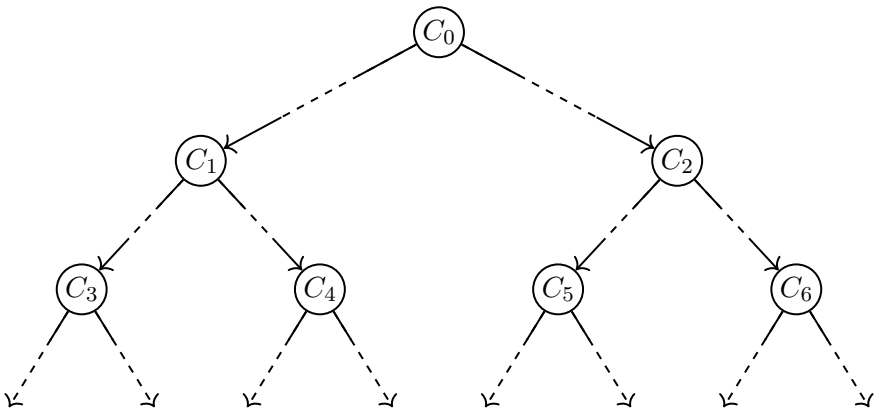

Figure 5: Upward Solution

While the Upward Solution looks attractive from afar – and I will defend a version of it in Section 4.5 –, the version specified above faces serious problems when we look closer. For one, it doesn't establish intension identity between FTC and FSU. Suppose that only conditions C_3, C_4, C_5, and C_6 from Figure 5 are given as input. While FTC will output all higher genera of these cognitions, it won't output all systematically connected cognitions, missing the infinitely many lower species. Thus, given the same input, FTC and FSU produce different outputs. The big problem is that FTC only ever leads us upwards to higher genera, and so will never complete a bottomless system that has no lowest species.

One could try to get around this problem by revising FTC. Rather than seeking the totality of L conditions *for our given cognitions*, one might propose a more liberal version, call it FTC*, which states that for all x: if x is a cognition and there is a y, such that y is a cognition (given or not) and y is an L condition of x, then find y. FTC* would always output the complete system of cognitions, precisely because there is no lowest species. For if there is no lowest species, it follows that all cognitions are genera, and hence L conditions, of another cognition. FTC* would therefore fix the problem that the system lacks a bottom.

But even so, FTC* is both textually and philosophically unattractive. For

not only does Kant explicitly limit Reason's quest for totalities to 'conditioned cognitions of the understanding' (A307/B364), but this limitation is also philosophically motivated. Kant generally characterises Reason as a faculty that seeks to find a 'Because for every Why' (A586/B614). Yet the question 'Why is x the case?' presupposes that x is given to us, and indeed x is given to us as conditioned.[224] On Kant's account, then, only things that are given to us as conditioned call out for explanation. Reason does not seek a Because without a Why.

The main issue, however, is that FTC*, like FTC, also isn't intensionally identical to FSU. For suppose that only cognitions C_3 and C_4 from Figure 5 are given as input. While FTC* (and FTC) would output all their higher genera, including the highest genus, C_0, FSU would not. Here is why: PH tells us to find *a* common genus of C_3 and C_4, such as C_1. However, once we find C_1, we don't get all the way up to C_0 because PH requires *two* cognitions as input. Since C_1 already unifies all our given cognitions, PH doesn't lead us upwards to the highest genus, which hence isn't systematically connected to the cognitions that were given as input. For some inputs, therefore, FTC* (and FTC) output(s) more genera than FSU.

This, however, points to a wider worry with Kant's account of systematic unity. In Section 4.1, I had suggested that it is plausible, *but not trivial*, that we would converge to the complete system of all cognition by iteratively applying PH, PS and PC; in other words, by following FSU. We now see why this convergence is not trivial. On some inputs, we would not actually converge to the complete system of cognition, because PH – or at least my formulation thereof – doesn't get us all the way up to the highest genus. Yet it is only if we reach the highest genus, that we would be able to fill in the rest of the system by iteratively applying PS and PC.

One might take this wider worry to simply count against my original formulation of PH. Instead of taking two cognitions as input, we could

[224] On this point, see also Nisenbaum, who notes that Kant 'does not posit that everything is in need of explanation, but only things that are conditioned' (ms). See also Lu Adler (2021), Watkins & Stratmann (forthcoming), and Chaplin (ms).

reformulate the principle, call it PH*, to require only one cognition as input, stating that for all x: if x is a given cognition, then find a cognition y such that y is a genus of x. By requiring only one cognition as input, PH* would address the wider worry. For it would always output the highest genus from any (non-empty) set of given cognitions as input. And once we have reached the highest genus, we can complete the system of cognitions by iteratively applying PS and PC.

But PH*, like FTC*, is both textually and philosophically unattractive. Kant illustrates the principle of homogeneity with an example from chemistry. Starting with our concepts of 'acidic and alkaline salts', the principle instructs us to find 'a common principle to [...] the salts', i.e., a genus of which acidic and alkaline salts are species (A653-4/B681-2); however, starting from two species of salt, the principle, unlike PH*, does not direct us to find, say, the basic matter of the universe. Worse still, in cases where the highest genus is already given – as is the case with our concept of a thing in general –, PH* would make an impossible demand, namely to find a genus to the highest genus.

These considerations support of my original formulation of PH, which also seems to enjoy some traction in the literature. For example, Anderson argues that the principle of homogeneity directs us 'to seek a unifying genus *for any two* empirical natural kinds or powers' (2015: 371; my emphasis). And Kraus writes: 'The principle of homogeneity [...] guides us in seeking, for every *pair of concepts*, a genus concept [...] such that *both these concepts* can be subsumed under it' (2020: 232; my emphasis). I agree with both Anderson and Kraus, and thus keep my original formulation of PH. I will address the wider worry about FSU in Section 4.5.

We can now close several brackets and return to the original problem. We have seen that there are inputs for which FSU does not output the complete system of all cognitions, whereas FTC* does. So, in addition to all the textual disadvantages, FTC* also claims no philosophical advantage over FTC, which is why I suggest we stick with FTC. It then follows that FTC and FSU are not intensionally identical, not only because FTC always outputs the highest

genus, whereas FSU doesn't, but also because the system lacks a bottom, and so by following FTC we will typically miss lower species, species that we would not miss by following FSU.

In view of this problem, one might try to secure at least an extensional identity between FTC and FSU by creating an artificial bottom. Here is how: suppose we start with an input that contains all cognitions at some level of the system – like, say, C_3, C_4, C_5, and C_6 in Figure 5 – *as well as* all cognitions below this level. Given this input, both FTC and FSU would yield the same output, namely the complete system of all cognitions. Importantly, FTC wouldn't miss a lower species, simply because all lower species are already given as input. So the Upward Solution seems to at least show that FTC and FSU are extensionally identical.

While I agree with this conclusion, I suggest that the input in question is *irrelevant*. This is because it is only a generalisation of the input where all cognitions are given. It doesn't matter, after all, which level of the system we choose to establish the extensional identity. In general, for any level L, if all cognitions at or below L are given as input, then FTC and FSU yield the complete system as output. If we now choose the highest level – i.e., all cognitions at or below the highest genus –, then all cognitions would be given as input. But clearly the input at which all cognitions are given is irrelevant, and so should be its generalisation.

Worse still, these seem to be the only inputs, apart from the empty input,[225] on which FTC and FSU generate the same output. It therefore turns out that the Upward Solution – at least in its current version – not only fails to establish an intentional identity between FTC and FSU, but moreover also fails to show that FTC and FSU are extensionally identical for any relevant input. Once again, the main problem here seems to be that the system

[225] Strictly speaking, we would also get an extensional identity between FTC and FSU for inputs where not every species is given below a certain level, but where for each species it is guaranteed that one of its species is given. Consider, for example, a striped input, where (i) there is at least one level L_1 where all cognitions are given as input, and (ii) for each level L_n there is another level L_m ($m > n$), so that all cognitions at L_m are also given as input. However, inputs like this one are hardly more than play, and thus clearly irrelevant to establishing any meaningful identity between FTC and FSU.

has no bottom, and yet that FTC only directs us upwards. While I think that the Upward Solution can be saved, let us first look at some alternative approaches.

4.4 Alternative Approaches

In this section, I will consider and reject three alternative approaches to establishing the identity between FTC and FSU. The first of these approaches, which I will call the 'Downward Solution', simply reverses the direction of L conditions. The main claim is: *x is an L condition of y iff x is the species of y*. Thus, unlike the Upward Solution, which claims that genera are the L conditions of their species, the Downward Solution claims that species are the L conditions of their genera. So we have to go downwards to lower species to find further L conditions. If we combine Figures 3 and 4, we get Figure 6, with the direction of the arrows reversed.

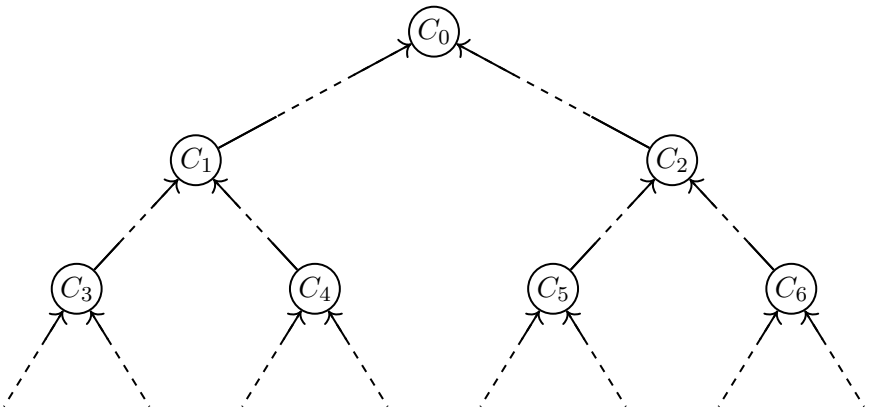

Figure 6: Downward Solution

The Downwards Solution has been entertained by Watkins, who notes that 'if one concept *contains* another *in* itself, then the content of the one conditions that of the other' (2019: 220; my emphasis).[226] Since species

[226] Watkins generally identifies as '[t]he crucial idea [...] that Kant views the containment

contains their genera in themselves, Watkins seems to think that species are conditions of their genera. The Downward Solution can also lay claim to a historical precedent in Wolff, who notes in his *Anmerkungen* that we 'descend from the highest [concepts] to the lowest' (1740/1972: § 53, 118).[227] And while Wolff does not identify this descent with the search for further L conditions, it is this very descent that we perform.

To be clear, the Downward Solution does not set out to establish any intensional identity between FTC and FSU. It's easy to see why. Suppose cognitions C_1 and C_2 were given as input. While FTC would output all species of C_1 and C_2, it would miss the highest genus, C_0, as well as any cognition between C_1, C_2 and C_0. By comparison, FSU would output the highest genus, C_0, because C_0 is systematically connected to C_1 and C_2 via PH. This means that FTC and FSU can produce different outputs for the same input, even on the Downward Solution.

But while the Downward Solution fails to establish intensional identity, it succeeds in establishing extensional identity. The main problem with the Upward Solution was that the system of all cognition lacks a bottom because there is no lowest species. But while there is no lowest species, there is a highest genus. The Downward Solution exploits this asymmetry. By turning the L conditioning relation upside down, the Downward Symmetry can use the highest genus as its bottom: whenever the highest genus, C_0, is given as input, FTC and FSU will output the whole system of cognition, because all other cognitions in the system are both species and, on the Downward Solution, L conditions of C_0.

What's more, the highest genus isn't an irrelevant input either. Quite the contrary. For, as I had noted in Section 4.1, Kant explicitly declares with respect to concepts, that the highest genus – i.e., the 'concept of an object in general' (A290/B346) – is already given. All we have to do in order to obtain this '*conceptus summus*' is to take any concept C, and iterate a process

relation that obtains between concepts of differing levels of generality as a kind of conditioning relation, one which reason given its nature, must seek out' (2013: 291).

[227] Anderson suggests that 'Wolff's vision of a concept hierarchy descends from a common highest genus' (2015: 96).

of logical 'abstraction', in which we subtract from all of C's marks (9:97).[228] Moreover, while the highest genus among our cognition of laws – i.e., the law that governs the 'basic power [*Grundkraft*]' (A649/B677) – is not yet given, it seems that it might well be given, if only over an infinite period of time.

So, what's the problem with the Downward Solution? The main problem, as I see it, is that the Downward Solution outright contradicts the text. We had seen in Section 4.3, that Kant repeatedly suggests that genera are the conditions of their species. Remember that 'reason in its logical use seeks the *general* condition' (A307/B364); and that it's 'a logical demand [of reason to] [...] ascent to ever *higher conditions*' (A309/B365). Passages like these motivated the Upward Solution in the first place, and I think they decisively count against the Downward Solution, despite its many philosophical merits.

This brings me to the second alternative approach, which I will call the 'Bidirectional Solution'. The Bidirectional Solution attempts to avoid the textual difficulties of the Downward Solution by conceding that genera are the L conditions of their species – thereby accounting for the quoted passages – while insisting that species are also the L conditions of their genera. The main claim of the Bidirectional Solution is: *x is an L condition of y iff x is the genus of y or x is the species of y*. On this picture, we have to go both upwards to higher genera and downwards to lower species to find further L conditions. By merging Figures 3 and 4, we now have Figure 7 (see next page), with arrows pointing both up and down.

A version of the Bidirectional Solution has been defended by Breitenbach (2021). Focusing on the system of concepts, Breitenbach acknowledges (i) that 'more general concepts would contain the conditions for more specific concepts'; but also (ii) that 'more specific concepts condition more general concepts' (2021: 292).[229] Breitenbach thinks that both genera are the L

[228] Kant develops a similar derivation of the highest genus in the horizon-passage in the Appendix (A658-60/B686-8). De Jong argues that Kant adopts this derivation from the tradition of porphyrian trees (1995: 624-5, 636-7n). For a further discussion of Kant's derivation for the highest genus, see Watkins (2013: 284-5).

[229] Breitenbach (2021: 292) suggests that Watkins (2019a: ch. 10) also defends the Bidirectional Solution. I'm, however, inclined to read Watkins as defending the Downward Solution because he only ever claims that species are the conditions of their genera, but not that genera are the conditions of their species.

conditions of their species, and species the L conditions of their genera. Moreover, Breitenbach suggests that this establishes the identify between FTC and FSU: 'by searching for [...] all conditions together as a totality [...], reason is not primarily after any one node in the system [...] but after the entire system taken as a whole' (2021: 296).

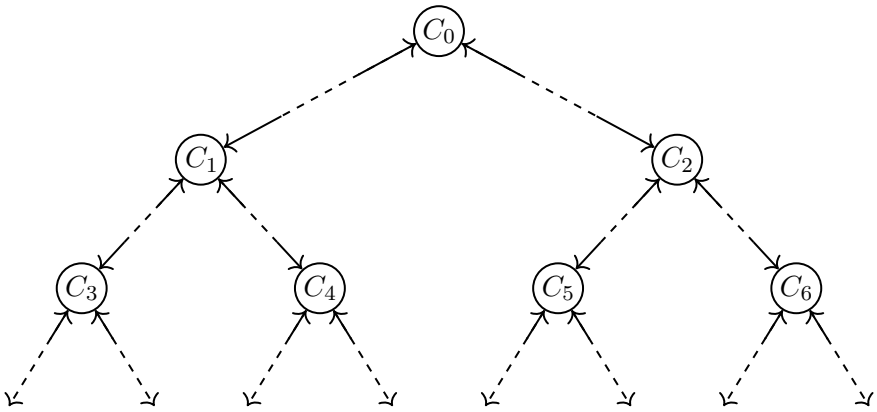

Figure 7: Bidirectional Solution

Just to be clear: even on this permissive interpretation of L conditions, we don't get intentional identity between FTC and FSU. The wider worry I discussed in Section 4.3 still lingers: there are possible inputs on which FTC would get us to the highest genus – just as it did in the Upward Solution –, but FSU wouldn't. For example, if cognitions C_3 and C_4 from Figure 7 are given as input, FTC would again get us to the highest genus, C_0 because C_0 is an L condition of C_3 and C_4. By contrast, FSU only gets us to C_1 because PH requires two cognitions as input, but once we find C_1 as the common genus of C_3 and C_4, FSU doesn't require us to look for any higher genera.

Nevertheless, the Bidirectional Solution claims to preserve the philosophical advantages of the Downward Solution while retaining the textual advantages of the Upward Solution. Like the Downward Solution, the Bidirectional Solution establishes the extensional identity between FTC and FSU, since both principles output the complete system of cognition when the highest genus is given as input. But like the Upward Solution, it also makes

sense of Kant's claim that genera are conditions of their species. Thus, the hope is that the Bidirectional Solution can combine the advantages of the previous two solutions.

But this hope is unfounded. For one thing, it is doubtful whether the Bidirectional Solution really solves the textual problems of the Downward Solution. Kant not only claims that genera are *also* the conditions of their species, but he also implies that *only* genera are the conditions of their species. In this passage from the *Vienna Logic*, he states that the 'ground [or condition] *always* contains the consequences under itself' (24:910; my emphasis). Moreover, there are no passages in which Kant states that species are the conditions of their genera. And while the absence of evidence isn't the evidence of absence, exegetical parsimony recommends that we don't multiply L conditioning relations without need.

But the textual difficulties aren't even the main problem with the Bidirectional Solution. I had noted that, for Kant, conditioning relations must at least be transitive, asymmetrical, and thus irreflexive. But if genera are the L conditions of their species, and vice versa, then L conditions would be symmetrical and (given transitivity) also reflexive. One salient response to this problem is to distinguish between two L conditioning relations, let's call them 'L_1 conditioning' and 'L_2 conditioning'. We can define that the genus is the L_1 condition of its species, and the species is its L_2 condition of its genus. In this way, L_1 and L_2 conditioning satisfy the structural properties that apply to any conditioning relation.

And to be fair, that's exactly what Breitenbach (2021) does. Focusing on concepts, Breitenbach distinguishes between '*containment-in* conditioning', on the one hand, and '*containment-under* conditioning' on the other (2021: 292-3). On her terminology, genera are the containment-in conditions of the species (= L_1 conditions), and species are the containment-under conditions of their genera (= L_2 conditions). Importantly, Breitenbach sees these as two separate conditioning relations, writing that '[f]or Kant, the systematic unity of cognitions consists in the search for a plurality of condition' (2021: 292).

But by distinguishing between two L conditions, the Bidirectional Solu-

tion only makes things worse. Because we would now get two versions of FTC, let's call them FTC_1 and FTC_2. FTC_1 would require us to find the totality of L_1 conditions for our L_1 conditioned cognitions; and FTC_2 would require us to find the totality of L_2 conditions for our L_2 conditioned cognitions. But now we can ask how FTC_1 and FTC_2 can be identical to FSU. Here, then, the Bidirectional Solution faces both the problems of the Upward Solution – for the identity of FTC_1 and FSU – and the problems of the Downward Solution – for the identity of FTC_2 and FSU. So if anything, the Bidirectional Solution combines the worst of both worlds.

This brings me to the third alternative approach: declaring defeat. We have seen that the Upward Solution, Downward Solution, and Bidirectional Solution all fail. However, these three solutions seem to exhaust the solution space: for either genera are the L conditions of their species, or species are the L conditions of their genera, or both. So even if Kant intended FTC and FSU to be meaningfully identical in some sense, it turns out that they are not. There simply is no relevant input for which we can even establish the extensional identity between FTC and FSU that simultaneously remains faithful to the letter of the text.

While declaring defeat can only be an *ultima ratio*, there may be ways to make this work. In the Appendix, Kant notes that PH and PS reflect 'a dual, conflicting interest [of Reason]: on the one hand, the interest of extension (of the general) with regard to the genera, and on the other hand, the interest of the content (of the specific) with regard to the diversity of the species' (A654/B682). A few pages later, however, Kant clarifies that 'there is no real conflict, but merely a different interest of reason' (A666/B694). And this makes sense. For any given cognition, we can, of course, do both: look for their genera and their species.

These remarks might motivate what I call the 'Defeat Reading'. The Defeat Reading combines five claims: (i) genera are the L conditions of their species, as the text explicitly states; and so (ii) FTC entails the demand of PH to seek higher genera; however, (iii) FTC does not entail the demand of PS to seek lower species; (iv) FSU, by contrast, entails both the demand of PH

to seek higher genera and the demand of PS to seek lower species (and the demand of PC for to seek in-between species); and so (v) the dual demand of FSU entails the demand of FTC, making FSU strictly stronger than FTC.

The Defeat Reading turns declaring defeat into a seemingly viable option. For although FTC and FSU are not identical principles, we wouldn't end up with two separate powers of Reason. That's because the demand of FTC would be entailed by the demand of FSU, though not vice versa. On the Defeat Reading, then, we would find that the Appendix specifies an additional demand of Reason – to seek lower species –, a demand that was not entailed by Kant's discussion of FTC in any of the preceding sections of the Dialectic. And insofar FSU is strictly stronger than FTC, the Defeat Reading would also imply that FSU is not just a formula of the Principle of Reason, but the Principle of Reason itself.

However, apart from turning Reason into a Frankenstein faculty – which I find even scarier than Reason being a zombie power –, the Defeat Reading faces two serious challenges. Since genera are the L conditions of their species, and Reason demands via PS that we seek lower species, Reason demands not only that we seek the regressive series of further L *conditions*, but also the progressive series of further L *consequences*. Yet at the beginning of the antinomy chapter, Kant explicitly states that the progressive series of consequences is merely 'an arbitrary and not a necessary problem of pure reason' (A411/B438). But then the demand to seek such a series cannot stem from the principle that *defines* Reason.

The second challenge is again the broader concern I've mentioned several times: there are inputs for which FTC gives us the highest genus, but FSU does not. So while it is true that FTC outputs all the cognitions that PH outputs, it may also output additional cognitions that PH doesn't output. As a consequence, however, it can't be the case that FSU is strictly stronger than FTC. For some inputs, the outputs of FTC and FSU overlap only partially, meaning that some cognitions are in the output of both FTC and FSU, some are only in the output of FTC, and some are only in the output of FSU. So without addressing the wider worry, we can't even declare defeat.

4.5 Ascending from Rock Bottom

In this section, I argue that the Upward Solution can be saved. For a start, I agree with Willaschek on the main claim of the Upward Solution, namely that genera must be the L conditions of their species: *x is an L condition of y iff x is the genus of y*. The text is clear on this point, we have seen. All we need to establish a meaningful identity between FTC and FSU is a bottom layer to the system from which we can ascend upwards to higher L conditions. I will argue that, contrary to first impressions, there is such a bottom layer. To see why, we will need to focus on the system of laws, and return to the system of concepts only later on.

Throughout this chapter, I have suggested that FTC and FSU start with given cognitions as their input. But it makes little sense to say that laws of nature are given to us. We don't just wake up one day and have a cognition of a law of nature given to us. Instead, it makes a lot more sense to say that we start with our observations of singular states, both external and internal, of the form 'substance S φ-s at time t'. For example, we observe that the rock is falling now; or that I represented 'God exists' yesterday. Indeed, Kant remarks that Reason 'does not begin with concepts, but with common experience, and thus takes as its ground something that really exists' (A584/B612).

The general idea, then, is that we infer laws of nature, both of inner and outer nature, inductively from the manifold states we observe. I will say much more about Kant's account of inductive inferences in Chapter 5, but roughly inductive inference work as follows: we observe several sequences of states – say, state S_1 is always followed by state S_2 – and infer the law L, according to which 'necessarily, if S_1, then S_2'. This general idea motivates a reading on which the system of laws doesn't hover over an infinite abyss of lower species, but rests on a solid foundation of singular states. Figure 8 illustrates this reading (where 'S' stands for states and 'L' for laws).

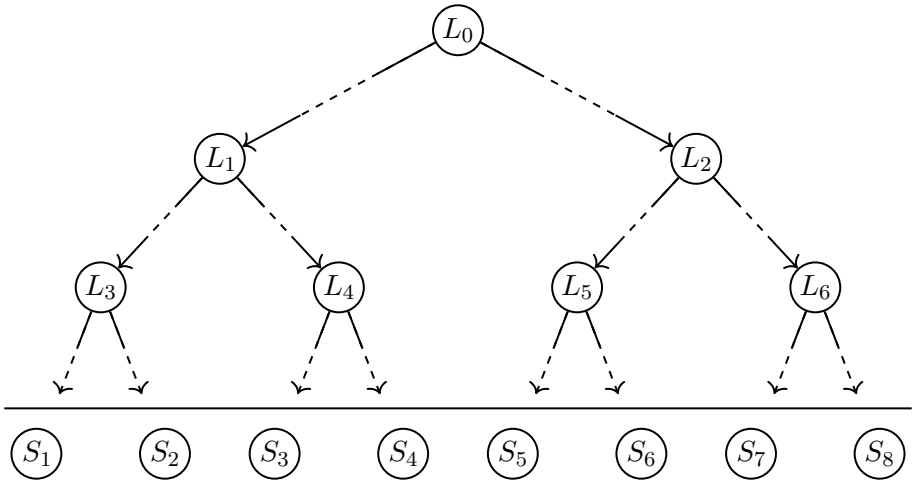

Figure 8: systematic unity of laws and states

This reading not only makes philosophical sense, but it also has textual baking. In Book One of the Dialectic, Kant identifies the premises of a syllogisms as L conditions. He writes: 'I have arrived at a cognition (conclusion) through a series of *conditions (premises)*' (A331/B387; my emphasis). This claim works especially well for the major premise of a syllogism, which Kant states must be a general rule: 'the general rule (major premise)' (A330/B386). This, then, dovetails with Kant's other claim that 'reason in its logical use seeks the *general* condition' (A307/B364; my emphasis). So major premises in syllogisms are L conditions of their conclusions.

Kant distinguishes between '*categorical* or *hypothetical* or *disjunctive*' syllogisms, depending on whether the major premise is a categorical, hypothetical, or disjunctive judgement (A304/B361). I focus on hypothetical syllogisms where the 'major premise is composed of 2 judgements', an antecedent and a consequent (24:587). The most prominent form of a hypothetical syllogism is the 'modus ponens, when the antecedent is true', i.e., when the antecedent is affirmed in the minor premise (24:580; see also A790/B818). The general form of a modus ponens inference can therefore be stated as follows (for any two judgements 'A' and 'B'):

Modus Ponens Inference

(P1) If A, then B.

(P2) A.

(C) *Therefore*, B.

Here is Kant's own example, slightly modified: 'If the earth is an animal, it has motion; now it [is an animal]'; so it has motion (R3262, 16:746). However, both A and B can also represent our cognitions of singular states in nature, S_1 and S_2. The major premise, which is a general rule that connects (our cognitions of) states in nature, can therefore be thought of representing a law of nature that governs these states. Now if both states are cognised, and we infer inductively to the law, we infer from what is L conditioned its L condition. Our cognition of naturals laws are the L conditions of our cognitions of singular states.

But Kant doesn't stop there. He argues that the premises of a syllogism, including the major premise of a hypothetical syllogism, become the conclusion of another *pro*syllogism: 'A chain of rational inferences where the conclusion of one becomes the premise of another; this is ratiocinatio polysyllogistica' (24:595).[230] Although Kant doesn't specify the form of the hypothetical prosyllogism, I suggest that it should itself be a hypothetical syllogism, i.e., have hypothetical judgement as major premise. But unlike a modus ponens inference, it must also have a hypothetical judgement as its conclusion, namely the major premise of the original syllogism. Chain inferences would meet these criteria:[231]

Chain Inference

(P1) If A, then C.

(P2) If C, then B.

(C) *Therefore*, if A, then B.

[230] See also (24:479) and (24:680).

[231] Another way to satisfy the criteria is by using a nested judgement as major premise, as in the following syllogism: (P1) if C, then if A, then B; (P2) C; (C) *therefore*, if A then B.

I will bypass the technical challenges that plague Kant's theory of prosyllogisms and focus on the underlying intention. Kant still intends the major premise to be the general L condition of the conclusion. Thus, when we infer inductively from the conclusion of a chain inference to the major premise, we infer from something that is L conditioned to its L condition. Insofar as the hypothetical judgments in the chain inference represent laws of nature, it follows that our cognitions about laws can be L conditions not only of our cognitions about singular states, but also of our cognitions about other laws.

Kant's theory of hypothetical (pro)syllogisms thus motivates the view that Reason starts its search for the totality of L conditions from our cognitions of singular states. After all, our cognitions of singular states are L conditioned by cognitions of laws, which are in turn L conditioned by cognitions of more general laws, *and so on*. Indeed, Kant suggests that Reason seeks 'conditions via prosyllogisms' (A331/B387), and further that it 'progresses through prosyllogisms to the unconditioned' totality of conditions (A323/B379). We thus have ample textual support to base the system of laws on a bottom layer of states.

The bottom layer, however, does not establish an intensional identity between FTC and FSU. This is because we're still faced with a version of the wider worry that FSU doesn't always output the highest genus, whereas FTC does. Consider the case where states S_1 to S_4 from Figure 8 are given as inputs. FTC will output all the laws that are their L conditions, including the highest genus, law L_0. FSU, on the other hand, will only output the laws that are systematically connected, which gets us up to law L_1, but not higher. Again, that's because PH requires two cognitions as input that are not genera of each other.

However, we can kill two birds with one stone by both establishing an extensional identity between FTC and FSU while also solving the wider worry. Consider the case where the complete bottom layer of states is given as input. Here, both FTC and FSU will output the complete system of laws. After all, FTC would seek all L conditions of all given states, which just are all the laws in the system. Likewise, FSU would seek higher genera for all

states, up to the highest law, which is the genus that unifies all states; FSU will also output ever more specific laws and all the other in-between laws. This shows that FTC and FSU must be extensionally identical.[232]

Moreover – and this brings me to the wider worry – the input in question, the entire bottom layer of states, is also clearly a relevant input. Because it is the input on which FSU really does complete the system of natural laws. Once all states are given, FSU will output the highest genus and all other laws, something that was thought to be plausible but turned out to be non-trivial. And while not all states are currently given, the expectation is that in our empirical progress, more and more states will be given, so that over infinite time we will converge on the entire bottom layer; and we will then also converge on the complete system by following FTC and by iteratively applying PH, PS, and PC, as is required by FSU.

My proposed reading encounters two notable objections.[233] The first objection asserts that the bottom layer of states violates the criterion that there is no lowest species in the system, as states would be considered lowest species. While I agree that states must be lowest species, I argue that there are no lowest species *within* the system because the states (or our cognition of them) are not part of the system – as indicated by the bold horizontal line

[232] One might object that even over an infinite time, we would not encounter enough states to infer all laws in the system. Assuming each state is discovered in discrete time intervals, we would have given a countably infinite number of states over an infinite time. However, the system of laws consists of uncountably infinitely many laws. Here's why: each law corresponds to a vertex in the graph shown in Figure 4, so the number of laws equals the number of vertices. Each vertex can be uniquely identified by a path from the highest genus downward, represented as a unique binary string. Moving left adds a '0' and moving right adds a '1'. Since the system contains infinitely many subordinate laws, there are at least as many paths as there are infinite binary strings. By diagonalisation, the set of all infinite binary strings is uncountable. Therefore, the system must have uncountably infinitely many laws. Consequently, even as immortal beings, we cannot complete the system. Nonetheless, since Kant predated the development of transfinite mathematics, this concern is arguably moot. For an extended discussion, see Benzenberg (ms-e).

[233] In conversation, Andrew Stephenson raised a third objection, challenging the notion of a unique lower layer of states. He noted that states can be described more coarsely or finely depending on the concepts we use to delineate them. This raises the question of whether a unique lower layer of states exists or whether the states themselves form a system analogous to the system of concepts. This question leads to a broader issue: how are the system of laws and the system of concepts supposed to interact? Addressing this issue is beyond the scope of this chapter. For the purposes of the following discussion, I will assume that the layer of states can be uniquely determined.

in Figure 8. The idea is straightforward: the system includes only *general* (empirical) cognition, not all cognition. Cognitions of states are *sigular* and thus lie outside the system, even though the system is built upon them.

By locating the layer of states *outside* the system, it becomes true that there is no lowest species *within* the system. Consider this: we can assign each layer of generality a real number between 0 and 1, with the highest genus assigned a 1 and the bottom layer a 0. Since the 0-layer is outside the system, there is no layer within the system closest to 0. For any n-layer, where $0 < n < 1$, there will be an m-layer such that $0 < m < n$. And so, just as there are infinitely many in-between species between any genus and its species, there will be infinitely many in-between species between a genus and the lowest species. Thus, if the lowest species is outside the system, there will still be infinitely many lower species within the system.

The second objection is that my solution only works for the system of laws – which has been my focus thus far –, but not for the system of concepts: there is no bottom layer for the system of concepts. I disagree. Just as the system of laws rests on a bottom layer of *states*, the system of concepts rests on a bottom layer of *intuitions*. Let me explain. In Section 4.1, I noted that Kant defines the genus-species relationship with reference to containment relations: x is a genus of y iff x is contained in y (or inversely, y is contained under x). While Kant typically discusses containment only between concepts, in the *Prize Essay* he also acknowledges that concepts can contain intuitions under themselves:

> For a representation to be a cognition (though here I mean always a theoretical one), we need to have a concept and intuition of an object combined in the same representation, so that the former [i.e., the concept] is represented as *containing* the latter [i.e., the intuition] *under* itself. (20:273-4; my emphasis)

While this is the most explicit passage, there are several others suggesting that containment relations apply not just to concepts, but also to intuitions. For example, in the Transcendental Aesthetic, Kant argues that our representation of space must be an intuition because 'it *contains* an infinite quantity of

representations *in* itself' – implying that intuitions can be contained under other representations (B40; my emphasis). Moreover, Kant opens the Schematism chapter by stating that when we subsume the representation of an object, specifically an 'empirical [...] intuition', under a concept, the representation of the object 'is *contained under* the concept' (A137/B176; my emphasis).[234]

Intuitions are thus the ideal candidate for the bottom layer of the system of concepts. Since they can be contained under concepts (but not vice versa),[235] it follows that concepts can be the genera and, hence, L conditions of intuitions. And while intuitions, like concepts, are cognitions in the broader sense described in the Stufenleiter passage as 'relating to an object', intuitions are 'singular' representations, whereas concepts are 'general' (A320/B377; my emphasis). As such, intuitions, like states, must be placed outside the system of general (empirical) cognitions.[236] This ensures that there is no lowest species within the system of concept. For every concept, there will be a more specific concept. This addresses the second objection.[237]

In conclusion, this chapter has demonstrated that the Formula of the

[234] I thus disagree with Anderson, who claims that 'a concept's logical extension is made up of specific *concepts*, rather than individual objects', or our intuitions thereof (2015: 50). For remember that the extension of a concept is defined by what is contained under it (A654/B682). And while I agree that every concept contains infinitely many more specific concepts under it, I would claim, based on the cited passages, that concepts *also* contain intuitions of individual objects under themselves, which are hence part of the concept's extension.

[235] Intuitions, as singular representations, satisfy Kant's definition of a lowest species: '*Each species is infima*, in so far as it *contains* no conceptus communes, but only singulars *under* it. Species infima is that which contains only individuals' (24:259; my emphasis).

[236] There is one key asymmetry between the relation of laws to states and that of concepts to intuitions. While laws can be the genera of states, they do not converge onto states as they become more specific; laws and states differ in kind. In contrast, concepts do converge towards intuitions as they become more specific. And while we start with *sensible* intuitions from which we abstract empirical concepts, concepts eventually converge onto intellectual intuitions. For if we were to fully determine a representation through the intellect, that representation would have its source in the intellect, and so count as *intellectual*.

[237] Indeed, the proposed account may even be generalised to the demands of practical reason. Just as Reason, defined as theoretical reason in the narrow sense, aims at a system of laws resting on a layer of states, practical reason might aim at a 'system of good maxims' (5:274), i.e., maxims that can be willed to be '*universal laws of nature*' (4:421), resting on a layer of the 'whole of all ends [...] in systematic connection' (4:433). While I cannot discuss the details here, Beck rightly observes that 'reason serves the same function in the practical realm as it does in the theoretical realm, that of systematising' (1960: 58). See also Guyer (2003) and Pasternack (2011b: 419). And although FTC and FSU are extensionally identical, I find it more fruitful to approach the 'common principle' of theoretical and practical reason (4:391) through the lens of FSU rather than, as Schafer (2023) has done, FTC.

Totality of Conditions (FTC) and the Formula of Systematic Unity (FSU) are, in a meaningful sense, identical. Whether we start with all singular cognitions and seek the totality of their L conditions, or aim to identify all their systematically connected cognitions, the objective is the same: to discover increasingly general laws that account for the given singular cognitions. Thus, by saving the Upward Solution, we can show that FTC and FSU are effectively two formulations of the same underlying Principle of Reason, which gives us a first answer to the Material Question.

My discussion in this chapter also explains Kant's alternative descriptions of Reason. For Kant also identifies Reason as a 'faculty of mediate inference' (A299/B355), which makes sense insofar as Reason pursues an 'ascending series of rational inferences' in prosyllogisms when seeking further L conditions. Kant also calls Reason the 'the *faculty of principles*' (A299/B356), which points to the fact that principles are 'general cognitions [that] serve as the major premise in a rational inference' (A300/B357).[238,239] The goal of the next chapter, Chapter 5, will be to analyse in detail how exactly we infer these general principles or laws from our given cognition of states, and the criteria we must satisfy in this process.

[238] These further descriptions of Reason are also discussed by Anderson (2015: 279-82), Schafer (2019: 188-9), and Kraus (2020: 176).

[239] Throughout this chapter, I have focused my discussion on FTC and FSU. However, there is also the 'Principle of Determination', which arguably also governs the operation of Reason (A571/B599). While much more could be said here, I suggest that we align the Principle of Determination with PS, insofar as both principles require that we further determine our concepts and so move from genera to species. Indeed, when discussing the 'logical law [...] of *specification*', Kant notes that '[t]he cognition of appearances in their *complete determination* [...] demands an ongoing specification of its concepts' (A656/B684; my emphasis). See also (28:410). I thus claim that the Principle of Determination is entailed by FSU qua PS. Importantly, I disagree with Kraus, who falsely aligns the Principle of Determination with PH (2020: 177).

5

Probabilistic Inferences

In the previous chapter, I argued that we seek the totality of L conditions by ascending to increasingly general cognitions. This is true not only of concepts, but also of laws, which will be my focus going forward. Kant thinks that we infer *general laws* from *particular states* that are given by observation. In this way, we also infer laws of the mind from the states we observe in the inner sense.[240] Therefore, to understand how we come to know the mental powers that are governed by these laws, it is vital to decipher how these ascending inferences work for Kant, and in which cases they are valid. That's what I will do in this chapter.

In Section 4.5, I had suggested, as a first approximation, that these inferences from particular states to general laws are inductive. Kant himself states that they are 'based on induction' (A196/B241; see also B3, A300/B357). And this makes some sense: we infer from some observed sequences of A-type and B-type states to a law that says 'Necessarily, when A, then B'. However, Kant also suggests that the inference from 'particular instances [...] to the universality of the rule' is done by 'the hypothetical use of reason' (A647/B675). This would seem to fit Kant's claim that we infer 'hypotheses [...] per inductionem' (24:558; see also 9:84-5).

However, we will see that this focus on induction is a red herring. I argue

[240] I agree with Heßbrüggen-Walter, who notes that 'it is self-evident that dispositional entities like faculties and powers are not directly experienceable [...]. What is accessible to us are only the realisations of the dispositional entities, thus the actually occurring activities, in the case of mental faculties thus representations' (2004: 166).

in this chapter that inferences from observed states to hypothesised laws are, in reality, *abductive inferences to the most probable explanation*.[241] The plan is as follows: first, I define and illustrate Kant's notion of hypotheses and explain how it interacts with his account of probability (§ 5.1). Next, I introduce and justify Kant's three main criteria for legitimate hypotheses: the 'Criterion of Possibility', the 'Criterion of Consequence' (§ 5.2), and the 'Criterion of Unity' (§ 5.3). Building on these criteria, I then assess whether Kant tacitly endorses Bayes's Law (§ 5.4) or whether he measures the probability of a hypothesis in some other way (§ 5.5).

5.1 What Hypotheses Are

Kant underscores the crucial role of hypotheses in our cognitive economy, stating that '[h]ypotheses are indispensable. [...] They are encountered everywhere' (16:463, R2675).[242] Given their importance, it is not surprising that Kant frequently discusses the doxastic and epistemic status of hypotheses in both his published and unpublished works, spanning his Critical and pre-Critical periods. He even dedicates an entire chapter in the *Critique* to hypothetical reasoning: 'The Discipline of Pure Reason in Hypotheses' (A769/B797). What is surprising, however, is that Kant's theory of hypotheses has received little attention in the literature.[243] This chapter aims to address that gap.

Let's begin by reconstructing Kant's definition of a hypothesis. Kant defines a hypothesis as 'a proposition which one assumes [*annehmen*] to

[241] To be fair to Kant, the distinction between *induction* and *abduction* was only properly articulated by Peirce in the 19th century. Without this distinction at hand, it's not surprising that Kant calls inferences by 'induction [...] particularly strange' (24:287). For Kant's theory of induction, see also (9:133), (24:277), (24:558), (24:594), (24:679), (24:771-2), and (24:777).

[242] See also (9:66-7), (9:86), (24:222-4), (24:746-7), (24:888-9), (28:144), and (29:103).

[243] There are a few papers on Kant's theory of hypotheses, most notably Butts (1961), (1962), Vanzo (2012), Anderson (2013), Pasternack (2014b), Demarest & van den Berg (2022), as well as the recent discussion by Cooper & Jones (2023) and Cooper (2023), (ms). There are also some entries on hypotheses in the Kant-Lexikon: de Freitas & Faggion (2015a, 2015b) and Sturm (2015). Yet to my knowledge, there is no single-focus monograph-length discussion of Kant's theory of hypotheses. Lehmann's thesis, *Idee und Hypothese bei Kant* (1908), doesn't clarify what hypotheses are for Kant, and how they ought to be rationally formed.

explain certain phenomena' (29:918).[244] Moreover, the phenomena-to-be-explained are given: 'Hypothesis is something that I assume to explain something else that is given to me' (28:416). Because the explanans is the *ground* and the explanandum the *consequence*, Kant also defines a hypothesis as the 'assent to [...] a ground because of the adequacy for the consequences' (9:84). This definition not only has historical precedent,[245] but Kant also maintains it with remarkable consistency.[246] Put formally:

Definition of Hypothesis

S hypothesises that p iff

(i) there is some q, such that q is empirically given to S, and ;

(ii) S assumes p as a ground to explain q as a consequence.

The definition is best illustrated through examples. Kant frequently acknowledges the cognitive importance of hypotheses in various epistemic contexts: 'Hypotheses are of very great utility, and we cannot banish them [...] from human cognition' (24:222).[247] In medicine, for instance, physicians observe several symptoms ($= q$), which are empirically given, and then formulate hypotheses about the cause of a disease ($= p$) to explain these symptoms (29:103; 24:220). Hypotheses also permeate the practice of law: 'all judges make hypotheses' about who might have committed a given crime ($= p$) to best explain the evidence given in court ($= q$) (24:222). These are two

[244] I translate the German terms 'annehmen'/'Annahme' as 'assume'/'assumption' rather than 'accept'/'acceptance', contrary to Chignell's (2007a, 2007b) suggestion. As argued in my Benzenberg (forthcoming-a), 'acceptance' signals an outright commitment, whereas Kant's 'Annahme' are not always an assertoric assent. Thus, 'assumption' is a more accurate translation.

[245] Wolff defines a hypothesis as 'something that is assumed [...] because it gives a ground for certain phenomena [*ratio reddi possit quorundam phænomenorum*]' (1728: § 126).

[246] For example, see also (A769-72/B797-800), (2:149), (5:126), (5:394), (5:463), (5:466), (5:470), (6:354), (8:53-4), (8:169), (8:399), (R2676, 16:464), (R2678, 16:465-6), (R2679, 16:466), (R2681, 16:469), (R2682, 16:469), (R2690, 16:471), (R2694, 16:472), (R5624, 18:260), (R6236, 18:520), (24:220-3), (24:439-40), (24:557-9), (24:647), (24:746-7), (24:886-9), (28:416), (28:1184), (28:1285-6), and (28:1291). For a further discussion of Kant's definition of hypothesis and its historical context, see Madonna (1992: 39), Trullinger (2013: 393), Pasternack (2014a: 68), (2014b: 67), Sturm (2015: 1059), Cooper (2023), (ms), Demarest & van der Berg (2022), and Techert (ms). See also my Benzenberg (forthcoming-a), (ms-c), (ms-d).

[247] See also (R2675, 16:463) and (24:224).

examples from professional contexts.[248]

However, hypotheses feature most prominent in the natural sciences, where they are 'useful and indispensable' (9:86).[249] This applies not only to the empirical part of physics – 'In [...] physics, [...] hypotheses arise' (28:144) – but also to empirical psychology, where psychologists use 'hypotheses for explaining the phenomena of the human [mind]' (R1525, 15:924).[250] Kant's favourite example of a scientific hypothesis is the Copernican system (= p) which is meant to explain the apparent movement of celestial bodies (= q).[251] I will examine the Copernican hypothesis more closely in Section 5.3.

Paradigmatically, however, science hypothesises laws. While Kant rarely cites examples of hypotheses about laws, he does mention the 'hypothesis of the law of continuity' (2:21).[252] Moreover, allowing for hypotheses about laws also explains why Kant aligns hypotheses with inductive inferences. Like inductive inferences, hypothesised laws draw universal conclusions from finite evidence: to explain the observed pattern of A-type states being followed by B-type states ($= q$), we hypothesise the universal law 'Necessarily, when A, then B' ($= p$). [253] I say more on the underlying notion of explanation in Section 5.2.

Yet because of their ubiquity, it's important to get hypotheses right. Wrong hypotheses in medicine can be lethal. The same is true for law. Kant mentions the case where '[t]hey found counterfeit coins in a mint master's box in England; he was executed and after his death it came out that someone else had tricked him' (24:530). In the natural sciences, false hypotheses can create great epistemic damage: 'It is [...] very sad when in an age all hypotheses prevail, but they are false[.] [...] Such hypotheses do more harm than good, [...] and if one accepts them as true, then instead of true one gets only false

[248] For other examples, see (A662/B691), (5:428n), (5:467), (9:66-7), (9:70), (24:223-4), (24:530), (24:439-40), (24:746-7), (24:886-9), (28:144), and (29:103).

[249] On the role of hypotheses in natural science, see (24:746-7), (24:886-9), and (29:103)

[250] See also (R5560, 18:233-4).

[251] See also (Bxxiin), (9:85-6), (R2675, 16:463), (R2680, 16:468), (24:221-2), (24:647), (24:887-8), and (29:103).

[252] See also (Bxxiin), (A775/B803), and (24:223-4).

[253] For these reasons, I disagree with Butts (1961: 165), (1962: 201), and Techert (ms), who maintain that laws cannot be hypotheses for Kant.

cognition' (24:224).

But this highlights a broader issue: *hypotheses can be wrong*. Kant argues that the inference from observed consequences to a ground of explanation is fallible because there are multiple possible grounds to explain a given phenomenon: 'A hypothesis can never be brought to apodictic certainty. [...] To conclude from the consequences to the grounds gives uncertainty because the same cognition can have more than one ground, and because a consequence flows from more than one ground' (24:647).[254] Perhaps the patient has a different disease. Perhaps the defendant is not guilty, but someone else committed the crime. Perhaps the observed A-B-sequences result from a different law of nature.

In this respect, hypotheses differ from both transcendental arguments and (doctrinal) Belief. Brook states that 'the role of transcendental arguments is to say that they attempt to reveal the conditions *necessary* for some phenomenon to occur' (1994: 12; my emphasis). Transcendental arguments show that experience is *only possible* on some condition X. Likewise, Belief is a 'hypothetically necessary' means to realise an end of reason, i.e., no other means is available to realise the end (A823/B851). And while this makes Belief a so-called 'necessary hypothesis' (R6236, 18:520), necessary hypotheses aren't hypotheses in the strict sense defined above (see Section 3.4 and my Benzenberg ms-c).

Although hypotheses can be wrong, we must still strive to get them right. But which of the many possible hypotheses should we choose? In response, Kant first notes that all inferences from the consequences to the ground of explanation are probabilistic: 'Ascendendo, I cannot actually infer, because one cannot know all the consequences of a cognition; this therefore only gives a *probable conclusion* and takes place with all our hypotheses' (24:528; my emphasis). In light of this, Kant offers a simple rule for choosing hypotheses: 'In a hypothesis one must [...] be convinced that this possible cause is also the *most probable* [*wahrscheinlichste*]' (29:104; my emphasis).

[254] See also (Bxxii n.), (A783/B811), (A830/B858), (5:183), (8:311-2), (24:746-7), (24:886-9), and (28:605).

Two clarifications: first, Kant does obviously not suggest that we are limited to holding only one hypothesis in all of science, specifically the one hypothesis that is most probable. Instead, he specifies that the hypothesis must be more 'probable [...] than the opposite' (24:879), i.e., more probable than all its *competing hypotheses*.[255] Second, the probability of the hypotheses ($= p$), we will see, depends in large part on the given phenomena ($= q$), which therefore provide *evidence* for the hypothesis.[256] We can thus offer a more precise formulation of the rule, for any hypothesis H, evidence E, and the probability of H given E, $P(H, E)$:

Rule of Probability Maximisation

(RPM) It is rational to assume a hypothesis H given the evidence E iff for all hypotheses H_i that compete with H, $P(H, E) > P(H_i, E)$.

RPM raises an immediate question: how does Kant determine the probability of a hypothesis H given the evidence E? This will be the overarching question of this chapter.[257] Kant makes various claims about what affects the probability of a hypothesis, and fitting these claims together is a significant task. Specifically, it is uncertain whether Kant implicitly relies on a version of Bayes's Law (more in Section 5.4). To reflect this uncertainty, I represent the probability of H given E as a two-argument function $P(H, E)$, rather than using conditional probabilities, $P(H|E)$, which are defined as $P(H \wedge E)/P(E)$ and entail Bayes's Law as a theorem.

For a start, let me note a few general things about Kant's notion of probability. Kant directly links probability to certainty. In the *Critique of Judgement*, Kant states that '[p]robability is a part of [...] certainty', to which

[255] Competing hypotheses are both exclusive and exhaustive of the sample space of possible outcomes.

[256] I use 'evidence' in the contemporary sense of the term, which differs from Kant's own. For Kant, evidence (*Evidenz*) means something like intuitive self-evidence, which he also calls '*Augenscheinlichkeit*' (24:437). This kind of evidence can only be achieved in mathematics. '[M]athematical certainty', which is another name for intuitive certainty, 'is also called evidence' (9:70). See also (A790/B818), (2:290-1), (R2465, 16:382), (R2493, 16:393), (24:150), (24:225-6), (24:229), (24:441), (24:546), and (24:857-8).

[257] On Kant's theory of probability, see also Funaki (2002), Pasternack (2014b), Schüssler (2020), and Chignell (2021).

it relates 'as parts to a whole' (5:465). And in the *Jäsche Logic*, Kant notes, 'With probability, there must always be a measure by which I can estimate it. This measure is certainty' (9:82). Taken on face value, Kant can be seen to define degrees of probability as degrees of certainty. On this definition of probability, certainty is the yardstick of all probability, and so always entails probability of 1.[258]

However, Kant defines probability not only in terms of certainty, but also in terms of grounds of assent. He writes: 'Probability is a fraction where the *sufficient ground* of truth is the denominator, but the *insufficient grounds* of assent which I have are the numerator' (24:196).[259] The grounds in question are meant to be objective grounds that rest on 'the constitution of the object' and are thereby truth-conducive (A821/B849; see Section 2.4). Sufficient objective ground entails truth, which is why 'the probability of [...] judgements [...] approximates the truth' (R2595, 16:434).[260] Sufficient objective grounds also entail '*certainty* (for everyone)' (A822/B850), which is why Kant defines probability as a degree of certainty.

Probability, so understood, measures the strength of our objective, truth-conducive grounds. Kant's notion of probability thus corresponds roughly to what we might today call epistemic probability. At the outset of the Transcendental Dialectic, Kant himself contrasts his notion of '*probability*' ('*Wahrscheinlichkeit*') with that of '*illusion*' ('*Schein*'), which doesn't approximate to truth but merely has the semblance of truth (A293/B349). We must also contrast Kant's notion of probability with our present-day notion

[258] For a further discussion of this point, see my Benzenberg (ms-d). Against Chignell (2021), I argue that all certainty characteristic of knowledge must come with probability 1 for Kant, and so that Kant must be an infallibilist about sufficient objective grounds.

[259] Kant more generally defines probability as a degree or part of the sufficient ground, 'If the grounds of the assent are only a part of the sufficient grounds, then it is called probability' (R2452, 16:375). Other relevant passages include (A293/B350), (5:465), (9:82), (R2452, 16:375), (R2583, 16:427), (R2591, 16:432), (R2595, 16:434), (R2602, 16:436), (24:145), (24:194-5), (24:433), (24:436), (24:880), and (24:884).

[260] Kant even calls probability – *horribile dictu!* – 'a degree of truth' (24:143). In the same spirit, he remarks that '[p]robability is a partial truth' (24:507). Yet Kant doesn't mean to say that truth itself comes in degrees, but that the ground on which the probable assent rests is a degree of the complete ground, which entails truth. Here's how Kant puts it in the *Vienna Logic*, 'But if I compare the thing with the sufficient ground of truth: then it is probable' (24:884; my emphasis).

of subjective probability or credence, as a measure of the strength of our psychological confidence or conviction.[261]

Kant's definition of probability also satisfies a version the Kolmogorov axioms.[262] Since I cannot have less than no objective grounds (24:433), no probability can be lower than 0, which gives us the first axiom. And since '[n]o ground can be more than sufficient' (24:196), no numerator can increase the probability beyond 1, giving us the second axiom. Finally, Kant compares the increase in probability with the addition of weights: 'one lot [*Loth*] is always a ground for lifting a pound, even if it is not sufficient, because *several lots together make a pound*' (24:195; my emphasis). This suggests that probability, for Kant, is additive as required by the third axiom.

So much on Kant's general notion of probability; now back to hypotheses. Kant suggests that a hypothesis (about a law) grows in probability the more of its consequences are given by observation. He writes: 'The *more* an opinion suffices to explain the *consequences*, the *greater* the *probability* of the hypothesis. Until the opinion is sufficient to explain all consequences, then it has complete certainty' (24:439-40; my emphasis).[263] Kant's idea seems to be that all possible A-B-sequences constitute the sufficient objective ground for the law 'Necessarily, when A, then B'. So the more of these sequences I observe, the more of the sufficient ground is given to me, and the more probable the hypothesised law becomes.

While a hypothesis would obtain probability 1, and so become certain, if all its consequences were observed, Kant argues that this is impossible. We cannot 'know [*kennen*] all consequences' of an assumed hypothesis (24:440). He justifies this with an astonishing remark, stating that every hypothesis 'has *infinitely* many consequences' (24:220; my emphasis). Kant's thought here must be that a universal law of nature has infinitely many (possible)

[261] For Kant's theory of subjective probability, see my Benzenberg (forthincoming-a).

[262] Kant's theory of probability doesn't satisfy the strict formulation of the Kolmogorov axioms because Kant never defines a sample space of events. But if we replace the sample space with the sufficient objective ground, we get variants of the axioms that Kant could have accepted. However, see Section 5.5 for key differences between the sample space and the sufficient objective grounds.

[263] For the claim that hypotheses become more probable the more consequences are given, see also (A649/B677), (9:84-5), (24:558), (24:886-9), and (29:51).

instances. And as finite beings, we cannot observe or verify infinitely many consequences. Doing so would 'exceed our powers' (A790/B818), at least within a finite time.[264]

Because the sufficient objective ground of hypotheses (about laws) is infinite, we should not measure their probability as a simple fraction with the sufficient objective ground as the denominator. Kant himself appears to deviate from this straightforward formula. As evidence accumulates and more objective grounds are provided, the probability of a hypothesis grows with asymptotic 'approximation towards certainty' (9:84), a claim Kant reiterates in multiple contexts.[265] Therefore, although the probability cannot be represented as a fraction, each additional consequence still increases the probability of the hypothesis, albeit in diminishing increments.

This concludes my discussion for this section. I've established that Kant defines hypotheses as assumptions made to explain phenomena, and that we should adopt the hypothesis most probable given the evidence. I've also shown that the probability of a hypothesis increases as more of its consequences are observed. While this provides a first anchor for Kant's view on the probability of hypotheses, it is not yet the complete picture. As I will explain in Section 5.3, Kant also argues that hypotheses are more probable the more unified they are. Before addressing that, however, I will first discuss two minimal criteria a hypothesis must satisfy to be admissible in the first place: the Criterion of Possibility and the Criterion of Consequence.

[264] Kant thus endorses a simple two premise argument: (P1) a hypothesis would only become (empirically) certain if all of its consequences where given; (P2) it's impossible that all consequences of a hypotheses are given; (C) therefore, a hypothesis cannot become (empirically) certain. Kant repeats this argument in (A647/B675), (A789-91/B817-9), (9:84-5), (R2178, 16:260), (R2678, 16:465-6), (R2679, 16:466), (R2680, 16:466-9), (R2681, 16:469), (24:220), (24:223-4), (24:439), (24:444), (24:528), (24:557-9), and (24:886-9). For a discussion of the argument, see Butts (1962: 196), Madonna (1992: 40), Sturm (2015: 1060), Cooper (2023), (ms), and Demarest & van den Berg (2022: 16).

[265] See (A647/B675), (9:81-2), (21:61), and (24:886-9). Kant notes that all approximation (*Annäherung*]) is asymptotic (*asmptotisch*) (A663/B692).

5.2 Possibility and Consequence

Kant's criteria for admissible hypotheses are meant to navigate the epistemic exuberance of earlier rationalist, and the scepticism of his empiricist predecessors. As an example of the rationalist's exuberance, Kant cites the physico-theological argument, 'which assumes the existence of God initially as a hypothesis' to explain the perceived order in nature (24:222).[266] Kant also mocks the Cartesian vortex theory. Descartes had hypothesised in his *Principia Philosophiae* (1644) that the cosmos operates through vast vortices of invisible ether, with planets and stars moving in swirling patterns. Kant calls this hypothesis a 'mere fantasy of the brain' (24:220),[267] and already lampoons it in his *Living Forces* essay from 1747:

> [The rationalists] have been obliged to tire their imagination with artificially devised vortices, to build one hypothesis upon another; and [...] they confound us with an infinite number of strange motions, which are far more wonderful and incomprehensible than all that is to the explanation of which they are to be applied. (1:60)

The exuberance of the rationalists led to a sceptical backlash against using hypotheses within the empiricist traditions of the 17th and 18th centuries. Bacon (1620/2004), Boyle (1662/1999, 1666/2008), and Hooke (1705) argued that hypotheses are mere fictions of the mind and should be discarded. This view influenced other empiricists, such as Locke (1689/1975: 563, 629), Hume (1740/2000: § 2), Turnbull (1740: 2), and Reid (1785/2002: 50). The most notable expression of this scepticism is articulated in Newton's famous dictum 'hypotheses non fingo', commonly translated as 'I feign no hypotheses' (1687/1999: 764; see also 1715: 222-4).[268]

Kant agrees with the sceptics that we need to temper the exuberance

[266] On Kant's critique of the physico-theological argument, qua *hypothesis*, see also (5:466), (R5484, 18:196), (23:43), and (28:1285-6).

[267] Kant also calls the Cartesian vortex theory a mere 'fiction' and 'philosophical novel' (24:220).

[268] Laudan claims that the Bacon-Boyle-Hook-view was endorsed by 'most scientists and epistemologists' of the time (1981: 10). For a comprehensive discussion of this anti-hypothesis coalition, see Vanzo (2012).

of rationalism, but he doesn't want to discard hypotheses altogether. After all, '[h]ypotheses are of very great utility, and we cannot banish them [...] from human cognition' (24:222). So Kant, being Kant, seeks a carve out a middle ground, rhetorically asking: 'Who will show us the boundary where grounded probability ends, and arbitrary fiction begins?' (1:365). Just as Kant's transcendental philosophy aims to define the limits of metaphysical knowledge, so does his theory of hypotheses aim to set clear limits to our empirical reasoning. With this aim in mind, Kant repeatedly specifies three criteria any admissible hypothesis must satisfy:

> And something must be apodictically certain in every hypothesis, namely 1) the possibility of the presupposition itself. [...] For actualities can be invented, but not possibilities; these must be certain. 2) The consequence. The consequences must flow correctly from the assumed ground; otherwise, the hypothesis becomes a mere chimera. 3) The unity. It is an essential requirement of a hypothesis that it be only one and that it require no auxiliary hypotheses for its support. (9:85)

In this passage, which is taken from the *Jäsche Logic*, Kant states the Criterion of Possibility, the Criterion of Consequence, and the Criterion of Unity.[269] The first two of these criteria are the subject of this section, with the third being separately addressed in the next section. Taken at face value, the Criterion of Possibility and the Criterion of Consequence articulate binary rules for sorting admissible from inadmissible hypotheses. In what follows, I will examine how Kant justifies these criteria, how they relate to each other, and, importantly, how they integrate with his theory of probability and RPM.

[269] The Criterion of Possibility is mentioned here: (A770/B798), (5:394), (5:466), (9:84-6), (R2680, 16:466-9), (R2682, 16:469), (R2686, 16:470-1), (R5739, 18:341), (24:558-9), (24:647), (24:746-7), and (24:886-9). The Criterion of Consequence is mentioned in the these passages: (B115), (A646-7/B674-5), (A770/B798), (8:311-2), (8:399), (9:84-6), (R2178, 16:260), (R2675, 16:463), (R2678, 16:465-6), (R2680, 16:466-9), (R2681, 16:469), (R2682, 16:469), (R2690, 16:471), (R2694, 16:472), (R5739, 18:341), (R5560, 18:233-4), (24:220), (24:223-4), (24:392), (24:439-40), (24:530), (24:557-559), (24:746-7), (24:886-9), and (28:416). For the Criterion of Unity, see also (A653/B681), (A774-5/B802-3), (R267516:463), (R2676, 16:464), (R2678, 16:465-6), (R2681, 16:469), (24:223), (24:439-40), (24:647), (24:746-7), and (24:886-9).While Kant states these three criteria with remarkable consistency, he occasionally experiments with alternative criteria. See, for example, (B115), (A682-4/B710-2), (5:438), (R2676, 16:464), (R5560, 18:233-4) (R5739, 18:341), and (28:416). I will bracket these alterative criteria given their shaky textual standing.

But let me start with a precise formulation of the two criteria. While the Criterion of Possibility mostly speaks for itself, the Criterion of Consequence needs some clarification. What exactly does it mean that '[t]he consequences must *flow* correctly from the assumed ground' (9:85; my emphasis)? In the *Critique*, Kant specifies the criterion by saying that a hypothesis 'must be connected as a *ground of explanation* with that which is actually given' (A770/B798; my emphasis). The idea, then, seems to be that the consequences flow from the hypothesis just in case the hypothesis explains these consequences. We can thus state the two criteria more precisely as follows:

Criterion of Possibility

(CP) It is rational to assume a hypothesis H given the evidence E only if (it is certain that) H is possible.

Criterion of Consequence

(CC) It is rational to assume a hypothesis H given the evidence E only if (it is certain that) H explains E.

I will discuss these criteria in reverse order, starting with CC. There are two plausible justifications for CC. First, it could be justified as a constitutive norm of any hypothesis. Hypotheses are defined as assumptions we make to explain a given phenomenon, so any assumption that cannot explain the evidence fails as a hypothesis. Second, CC might also be instrumentally justified. Kant sometimes refers to hypotheses as epistemic *means* that we adopt 'to explain something' (24:750) – explanation being the *end*. Since it is only rational to adopt means that are sufficient to achieve a given end (4:414-7), it follows that it is only rational to adopt hypotheses that are sufficient to explain the evidence.

With CC justified, we can turn to a more basic question: what does it even mean for a hypothesis H to explain the evidence E? What is Kant's notion of explanation? Kant flirts with the hypothetico-deductive model, on which explanation requires *deducibility*. He states, 'hypothesis is an opinion that something is, because the consequences, which are, *can be deduced from*

it' (R2682, 16:469; my emphasis). And also says, '[t]o the perfection of a hypothesis belongs the [...] *deducibility of the consequences* derived from this assumed grounds' (24:647; my emphasis).[270] If we represent the evidence as a set of propositions $E = \{p_1, p_2, ..., p_n\}$, we can say that H explains E only if for all $p \in E$, p can be logically deduced from H.

Moreover, Kant believes that the primary method to deduce evidence from a hypothesis is through 'modus ponens' inferences, as discussed in Section 4.5 (A790/B818). When we observe an A-type state followed by a B-type state, we hypothesise the law 'Necessarily, when A, then B'. From this law and the antecedent A-state, we can infer the B-state via modus ponens. This reasoning also explains why Kant asserts that explanation requires derivation from laws, which specify the major premise in a modus ponens inference: 'For we can explain nothing except what we can reduce to laws' (4:458-9).[271]

While logical deducibility is a necessary condition for explanation, it is not sufficient. Kant also requires that the hypothesis be at least as intelligible as the evidence. He insists that we must never 'attempt to explain what we do not understand by reference to something that we understand even less' (8:53-4). Such attempts fail to provide any explanation: 'A transcendental hypothesis [...] *would not be an explanation at all*, since what is not sufficiently understood [...] would be explained by something of which nothing is understood' (A772/B800; my emphasis).[272] Put together, Kant's definition of explanation involves two conditions:

---

[270] In the third *Critique*, Kant defines explanation as 'clear and determined derivation [*Ableitung*]' (5:412). See also (A822-3/B850-1), (9:84-86), (24:223-4), (24:558-9), (24:746-7), and (24:887). On Kant's deductive notion of explanation, see also Madonna (1992: 39-40), Falkenburg (2000), Sturm (2015: 1060), Demarest & van den Berg (2022: 16), and Techert (ms), as well as my Benzenberg (ms-a).

[271] In a similar spirit, Kant notes: 'No other things or grounds for explanation can be cited to explain given appearances than those that have been linked to the given ones according to the already known laws of appearances' (A772/B800). See also (A798/B827).

[272] Kant also writes: 'The hypothesis must be [...] more clearly cognised than the explained. So it must not have the same darkness as the latter, e.g., the freezing from a cold-making matter' (R2676, 16:464). See also (A635/B663), (A770-3/B798-801), (A798/B827), (4:458-61), (5:139, 412), (8:53-4), and (21:346). We should accordingly also not explain the sleep-inducing properties of opium via its *virtus dormitiva*, as is Molier's famous example.

Definition of Explanation

A hypothesis H explains evidence E iff for all $p \in E$:

(i) p can be logically deduced from H, and;

(ii) H is more intelligible than p.

This now gets me to CP: a hypothesis is rationally admissible only if it is possible. Note that CP requires more than just the *logical* possibility of the hypothesis, which is determined by the PNC. Kant asserts that 'the mere principle of contradiction [...] which can prove nothing but the possibility of thinking' cannot serve 'as the criterion of this possibility' relevant to hypotheses (5:466).[273] Instead, we must establish the *real* possibility of a hypothesis by determining whether it 'agrees with the formal conditions of experience (in accordance with intuition and concepts)', as outlined in the second Postulate of Empirical Thought (A218/B265).

I propose that CP can be reduced to CC. To be at all intelligible in the sense required for an explanation, it would seem that a hypothesis must at the very minimum be really possibility. Only if we can establish the real possibility of a hypothesis can it serve as an 'explanation for given appearances [...] from known empirical principles', without which 'nothing is understood' (A772/B800).[274] If we couldn't establish the real possibility of the hypothesis, 'there would be no end to empty fantasies' (5:466). Yet empty fantasies are clearly less intelligible than given appearances.[275]

While CP inherits its justification from CC, it is CP that gives CC its bite. Insofar as we cannot establish the real possibility of metaphysical hypotheses, Kant notes that '[t]here are sciences that do not allow hypotheses, such as [...] metaphysics' (9:86).[276] And so 'even the wildest hypotheses, if they are only physical, are more bearable than a hyperphysical one[s]', which is why we

[273] See also (Bxxvin), (A222/B270), (A770-1/B798-9), (5:394), and (R5560, 18:233-4).

[274] See also (A677/B705) and (5:139).

[275] Note that the inverse does not hold: CC cannot be reduced to CP. Some hypotheses may be really possible without explaining the given evidence. For example, the hypothesis that unicorns exist might be really possible, but it wouldn't explain phenomena like 'earthquakes and fire-breathing mountains' (24:558).

[276] See also (A773/B801), (3:409), (4:746-7), (8:53-4), (16:471, R2689), (18:233-4, R5560), and (24:733).

are not rationally permitted to hypothesise 'a divine creator' to explain the '[o]rder and purposiveness in nature' (A772-3/B800-1). For the same reason we are not permitted to 'assume the soul as a simple substance [...] because the simple cannot occur in any experience' and so cannot be shown to be really possible (ibid.).

But CP and CC do not only rule out metaphysical hypotheses. They also render irrational various hypotheses in both psychology and physics. For example, Kant argued that we cannot establish the real possibility of, and therefore cannot assume as an explanatory ground, 'an understanding that is able to intuit an object without the senses, or a power of attraction without any physical contact' (A770/B798). And for the same reason we must reject the hypothesis 'that there are pure, disembodied spirits in the material universe' (5:468). These hypotheses are unintelligible and so, according to Kant, explain nothing.

The above cases are all examples of hypotheses that violate CC by failing to explain the given evidence. Other hypotheses violate CC by outright contradicting the evidence. In these cases, we are not just unjustified in adopting the hypothesis, but we actually know it to be false: 'one wrong consequence, then the hypothesis [is] false' (24:444). Indeed, just as the paradigmatic case of explanation involves a modus ponens inference, Kant argues that we can reject hypotheses based on 'modus tollens' inferences, '[f]or if even a single false consequence can be drawn from a sentence, then that sentence is false' (A791/B819). Kant acknowledges that it can be hard to discard hypotheses, yet he shows zero tolerance for anomalies:

> Hypotheses also have the disadvantage that they [...] become firmly rooted in our minds, and even if we realise their falsity afterwards, we still do not let them go so easily. For example, if this or that hypothesis were sufficient to lead to new, and correct consequences, and if it is only the hundredth, which contradicts the ground, then one rejects this one [consequence], even if it is obviously true, rather than letting go of that hypothesis[.] (24:224)

We can distinguish these two types of CC violations more precisely. Let '$\vdash_e$' represent the explanation relation defined above. In the first type of CC violation, there exists some $p \in E$ such that $H \nvdash_e p$. In the second type, there exists some $p \in E$, such that $H \vdash_e \neg p$. Both types of CC violation imply claims about probability. In the first case, where no objective grounds are given, Kant states that 'no probability takes case' (5:466), suggesting that the probability function, $P(H, E)$, is undefined for these hypotheses. In the second case, where we know H to be false, the hypothesis would appear to have zero probability, i.e., $P(H, E) = 0$.

With this in mind, I propose that not only can CP be reduced to CC, but CC can be further reduced to RPM. Hypotheses with undefined probabilities or zero probabilities cannot be considered to be the most probable. So while Kant presents CP and CC as separate binary criteria designed to restrain the epistemic exuberance of rationalist hypothesising without succumbing to scepticism, I find that, in effect, they are simply illuminating different aspects of the same underlying decision rule, which is RPM: it is only rational to adopt the hypothesis that is most probable given the evidence.

5.3 Unity as Generality

With CP and CC reduced to RPM, I now turn the Criterion of Unity. We have already encountered Kant's formulation of the criterion from the *Jäsche Logic*: 'It is an essential requirement of a hypothesis that it be only one and that it require no auxiliary hypotheses for its support' (9:85). This way of putting the criterion isn't an isolated instance. In the *Pölitz Logic*, Kant writes: 'Unity of the ground. No subsidiary hypotheses' (24:557). Formulation like these presents the criterion as a binary rule, akin to CC and CP: it is rational to assume a hypothesis H given the evidence E only if H can explain E without auxiliary or subsidiary hypotheses.

But this impression is mistaken. The Criterion of Unity should not be reconstructed as a binary rule. For Kant, unity varies in degree – a hypothesis can be *more or less* unified depending on the number of auxiliary hypotheses

it requires. And more unified hypotheses are more probable. Kant explicitly states: 'Unity [...] is an essential requirement of a hypothesis that it be only one and that it require no auxiliary hypotheses to support it. If we have to use several other hypotheses to support a hypothesis, it loses a great deal of its probability' (9:84). And again, in other words: 'The more subsidiary hypotheses are necessary, the less probable the hypothesis is' (24:647).[277] This seems to commit Kant to the following principle:

Principle of Probable Unity

(PPU) For any hypotheses H_1 and H_2 and evidence E, *ceteris paribus*, if H_1 is more unified than H_2 (i.e., H_1 requires fewer auxiliary hypotheses to explain E than H_2), then $P(H_1, E) > P(H_2, E)$.

With PPU in place, we can straightforwardly reduce the Criterion of Unity to RPM. All else being equal, we should rationally choose the more unified hypothesis because, (i) according to PPU, this hypothesis is more probable; and (ii) according to RPM, it is rational to assume the most probable hypothesis given the evidence. Therefore, unlike CP and CC, I won't discuss the Criterion of Unity as a separate criterion. Instead, the main aim of this section is to understand better what it means for a hypothesis to be unified and, in light of this understanding, to explain how Kant can justifies PPU.

To better understand Kant's notion of unity, it helps to see how PPU is applied. Kant often illustrates the principle by contrasting the Copernican hypothesis with that of Tycho Brahe. Briefly, Copernicus proposed a revolutionary model where the Sun is at the centre of the cosmos, with all planets, including Earth, orbiting it, with only the Moon orbiting Earth. In contrast, Tycho Brahe's model places Earth at the centre, with both the Moon and the Sun orbiting Earth, while all other planets orbit the Sun. Figures 9 and 10 (not to scale) illustrate these models, highlighting the orbits of Earth and the Sun in bold.

[277] Kant reiterates the connection between unity and probability of a hypothesis in (R2675, 16:463), (R2681, 16:469), and (29:103).

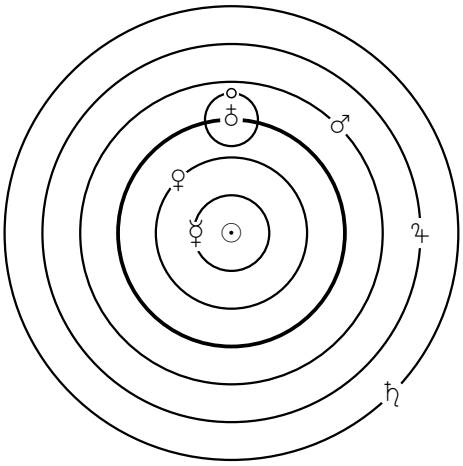

Figure 9: Copernican hypothesis

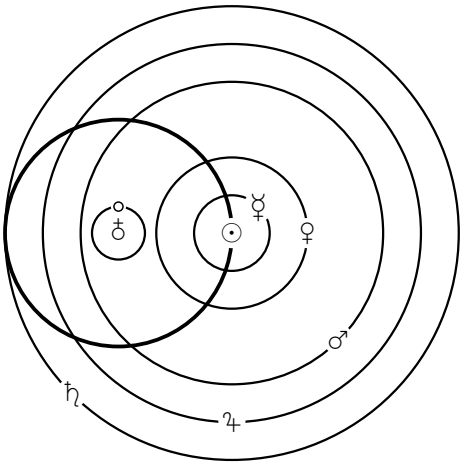

Figure 10: Tychonic hypothesis

The Copernican hypothesis is not only a model for Kant's transcendental philosophy (Bxvi-xvii), but is also more unified than the Tychonic hypothesis: 'Hypothesis subsidiaria, when one assumes several to explain a hypothesis, as Tycho Brahe assumed the epycylos epicyclorum' (24:440).[278] This point is intuitive: Brahe's model requires nested orbits which are expressed in subsidiary hypotheses, from which 'it can already be guessed that this is not the real ground' of celestial motion (24:558-9; see also 9:84-86). Brahe's disunified hypothesis seems less probable and so is less preferrable. Kant rejects the Cartesian vortex hypothesis for the same reason (1:60; 24:220).

Examples like these make PPU intuitively plausible: more unified hypotheses seem more probable because they involve less artifice. However, how does Kant justify PPU beyond its intuitive appeal? One might at first argue that disunified hypotheses are less probable because they violate CC and so lack all probability. Specifically, if a hypothesis H_0 requires one or more auxiliary hypotheses, $H_1, ..., H_n$, to explain for the evidence E, then H_0 alone does not sufficiently explain E. Indeed, Kant notes that auxiliary

[278] See also this passage: 'Tycho Brahe [...] made cyclos, and to explain these deviating cyclos, he assumed cyclos within cyclos, and so on and so forth, and you can see that it goes on and on without end' (24:889).

hypotheses are needed only when not '[a]ll consequences [...] flow from this one assumed ground' (24:746). In the *Jäsche Logic*, Kant seems to be gesture toward this idea when he states:

> Tycho Brahe's hypothesis, for example, *was not sufficient to explain many phenomena*; he therefore adopted several new hypotheses to supplement it. Here it can already be guessed that the hypothesis adopted cannot be the real ground. The Copernican system, on the other hand, is a hypothesis from which *everything that is to be explained can be explained*, as far as it has come before us. We do not need any auxiliary hypotheses (hypotheses subsidiarias) here. (9:85-6; my emphasis)[279]

Yet this justification of PPU is philosophically flawed. We evaluate not individual hypotheses but the entire set of hypotheses, $\mathcal{H} = \{H_0, H_1, ..., H_n\}$. While H_0 alone may fail to explain the given evidence E, the purpose of auxiliary hypotheses is to ensure that H_0 need not be rejected because, when combined with $H_1, ..., H_n$, it can explain E. Kant explicitly acknowledges that 'it is still possible that the Creator made an arrangement of the planets, as the hypothesis of Tycho Brahe shows' (24:888). The Tychonic model, with all its auxiliary hypotheses, can explain the observed motion of celestial bodies – when viewed through the lens of the Early Modern period[280] –, and thus does not violate CC.

Alternatively, one might try to justify PPU inductively. Each subsidiary hypothesis, the ideas goes, is introduced only because the previous hypotheses were not sufficient to explain the evidence. So, the more subsidiary hypotheses are required, the more often the theory has failed to explain past evidence, and the more probable it becomes that it will again fail to explain future evidence. This would also clarify why the probability of a hypothesis is tied to the number of subsidiary hypotheses: 'The more subsidiary hypotheses are necessary, the less probable the hypothesis is' (24:647).

[279] See also (R2680, 16:466-469), (24:647), and (24:886-9).

[280] In historical fact, Tycho Brahe was a meticulous astronomer, conducting extensive and precise measurements of celestial bodies using his own observatory. So if anything, it is his hypothesis, not that of Copernicus, which best explains the given evidence.

More subsidiary hypotheses indicate more past failure. Kant alludes to this inductive justification when he writes:

> In some inferences, one must invent many subsidiary hypotheses. If we have consequences and invent the ground for them, it may often be that we have even more consequences than the ground is sufficient for. We must therefore [again] invent new grounds, and these new grounds are nothing else than the subsidiary hypotheses. (24:223)

The issue with this inductive justification of the PPU is that it ties the principle to the historical development of a hypothesis. We can easily imagine scenarios where a hypothesis starts out disunified, requiring many subsidiary hypotheses, simply because the underlying unity in the evidence hasn't yet been discerned. In such cases, the number of subsidiary hypotheses wouldn't indicate past failure but would instead reflect the apparent disunity in the evidence. In fact, such cases not only seem possible but may be quite common in many scientific fields.

We encounter a similar issue when modifying the inductive justification of the PPU. Instead of linking the inductive base to the history of the *same* specific hypothesis under consideration, one might argue inductively that disunified hypotheses are improbable because *other* disunified hypotheses have often proven false in the past, with greater disunity correlating with higher failure rates. However, this approach still ties the PPU's justification to the contingent twists and turns of a science's history. We can easily imagine instances where disunified hypotheses have, by mere chance, performed surprisingly well in the past.[281]

But if PPU can neither be justified via CC nor via induction, how can it be justified? To answer this question, we must first rethink what it means for a hypothesis to be unified. Kant's emphasis on subsidiary or auxiliary hypotheses is a red herring, I argue. I take my lead from Kant's claim that

[281] One might also worry that the inductive justification of PPU is circular. Kant claims that hypotheses derive from inductive inferences governed by the Criterion of Unity (in addition to CP and CC), which in turn entails PPU. But then, PPU's induction must adhere to its own criterion. However, this circularity can be avoided by denying Kant's claim that hypotheses derive from inductive inferences (in the strict sense).

unified hypotheses explain more than disunified hypotheses. Kant writes: 'If we have to use several other hypotheses to support a hypothesis, it loses a great deal of its probability. For *the more consequences that can be derived* from a hypothesis, the more probable it is; the fewer, the less probable' (9:84; my emphasis).[282]

Kant here claims that unified hypotheses have greater explanatory power, as more consequences can be derived from them. To make sense of this claim, I propose to distinguish between the *evidence* E and the *domain* D of a hypothesis H. The evidence includes all *actual* consequences that have been given by observation, and that can be explained by the hypothesis. The domain by contrast comprises all possible consequences of a hypothesis – such as all *possible* states governed by a hypothesised law. In other words, then, the domain constitutes the sufficient objective ground of a hypothesis, whereas the evidence represents the insufficient ground already observed.

The distinction between a hypothesis's evidence and domain explains why unified hypotheses have greater explanatory power. True, both unified and disunified hypotheses can derive the same number of actual consequences. Paired with their subsidiary hypotheses, disunified hypotheses explain the evidence just as well as unified hypotheses. To use Kant's own example, both Copernicus and Brahe explain the observed motion of the heavenly bodies. But unified hypotheses derive more possible consequences. They have a larger domain because they attempt to explain a wider range of possible phenomena, which need not yet have been observed.

My main proposal is that the more general a hypothesis is, the larger its domain. General laws are higher up in the system of laws, and have more possible states stating under them. They attempt to explain a larger domain of phenomena. We can picture the domain as the shadow a law casts onto the layer of possible states: the higher the law, the larger its shadow. Thus,

[282] Kant repeats this claim: 'In the case of unity I conclude: because many consequences fit a ground, the more probable it is that the ground is true and the right one. But if some but not many consequences can be deduced from the hypothesis, and new hypotheses must always be made to support it, there is little probability' (24:888). See also (R2678, 16:465-6), (R2681, 16:469), (24:558-9), and (29:103).

the unity of a hypothesis can be measured by its generality. If we assign each layer of generality in the system a real number value between 0 and 1 – where the highest genus is 1 and the lowest is 0 (as suggested in Section 4.5) –, we can express the unity of a hypothesis H, $u(H)$, by this value. Figure 11 illustrates this thought.[283]

[283] My reading assumes that every law has a determinate degree of unity or generality. But this is not obvious. Writing about the system of concepts, which structurally mirrors the system of laws, Kant notes, '[O]ne concept is not broader than another because it contains more under itself – for one cannot know that – but rather insofar as it contains under itself the other concept' (9:98). Anderson draws on this passage and suggests that the 'more-general-than' relation is not connected on the set of all concepts: 'if neither of two concepts is contained in the other [...], then it is *not determinate* which is higher' (2015: 56; my emphasis); only if A is a genus of B, is it true that A is more general than B. Anderson's reading has much going for it. Since extension and content are reciprocal for Kant (9:98), a concept would contain more under itself iff it contained less in itself. One might thus be tempted to measure the generality of a concept by the number of marks it contains in itself: fewer marks equals greater generality. But this measure doesn't work. Kant's system of concepts is infinitely dense (see Section 4.1), meaning that every concept (except the highest genus) has infinitely many genera and thus contains infinitely many marks. Because these infinities all have the same size (see my Benzenberg ms-e), all concepts (except the highest genus) would, on the cardinality-measure, be equally general. But this is absurd, and so we must reject the measure. Without a measure, however, there appears to be no fact of the matter which of two given concept is more general. The only exception are genera and their species, which can use the proper-subset relation as a measure for their relative generality. What can we do to overcome this ontological limitation? In my Benzenberg (ms-e), I argue that the system of concept, specifically its infinite density, cannot be defined in the space of graphs; instead, the system must be embedded in the plane (thanks to Jonathan Fraser for pointing this out to me). Once the system is embedded in the plane, we are no longer bound to the cardinality-measure, but can choose from a great variety of different measures to determine the absolute geniality of a concept. We could, for instance, simply measure a concept's geometrical distance to the highest genus. With an alternative measure at hand, there should, pace Anderson, be a determinate fact of the matter about a concept's generality. And indeed, in the quoted passage from the *Jäsche Logic*, Kant only states that 'one cannot *know*' a concept's absolute generality. He does not deny there being a fact of the matter. This epistemic limitation, however, raises a separate concern. If we couldn't know the generality of a concept and (by structural analogy) of a law, we wouldn't be able to apply the Criterion of Unity to any real-world hypotheses (except where one law was the genus of another). But we can resolve this concern by inflating the notion of knowledge in question. Sure, we cannot know (in Kant's strict sense of 'know') the exact generality of a hypothesised law, but we should be able to arrive at a robust approximation. By means of illustration, say, you throw a dart at the [0,1] interval. Looking from afar, you may fail to tell the exact position of the dart to the n-th digit, but you can get a good sense for where it landed. This image also holds philosophical merit when applied to hypotheses. If presented with two hypotheses, I can roughly say which one of the two is more general – even if they don't contain each other. And indeed, that's exactly what Kant himself does when he compares the unity of the Copernican and Tychonic hypotheses. Not only should every hypotheses therefore have a determinate degree of generality or unity, but we should also be able to roughly discern that degree.

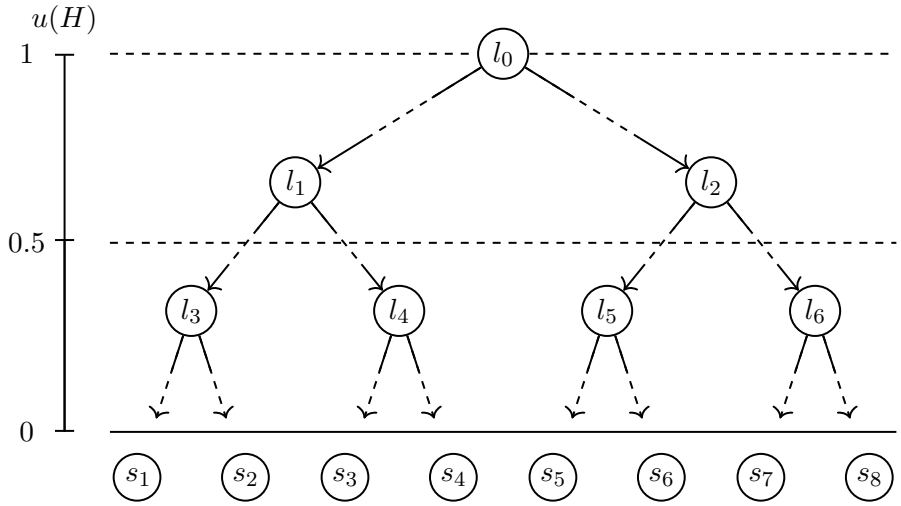

Figure 11: degrees of unity as generality

The proposed reading gains further support from Kant's discussion of 'acidic and alkaline salts' in the Appendix (A652/B680). As mentioned in Section 4.3, Kant argues that FSU, and specifically PH, requires us to seek a common genus under which these two types of salt can be classified as species. If such a genus is found, Kant further argues, it would provide the 'assumed ground for explanation [...] *probability by this very unity*' (A652/B680; my emphasis). Kant thus explicitly accepts the idea that genera possess greater unity than their species – as implied by $u(H)$ –, and reiterates that this unity would also confer greater probability.[284]

One might object to my proposed reading by arguing that a set of disunified hypotheses, $\mathcal{H} = \{H_0, H_1, ..., H_n\}$, can not only explain the same evidence, but also cover the same domain as a unified law. After all, both the Copernican and Tychonic hypotheses cover the motion of celestial bodies. But while this may be true, I argue that if a set of disunified hypotheses covers the same domain as a unified hypothesis, then each disunified hypothesis is less general than the unified one. Figure 12 illustrates this point: the unified

[284] Translating domain-talk into generality-talk has the added benefit of avoiding technical issues, since every law of nature, regardless of its generality, has infinitely many possible states under it. While all laws have the same domain size in this sense, there is an intuitive sense in which more general laws encompass a greater domain.

hypothesis of law L_1 covers the domain of possible states S_1 to S_n, whereas the disunified hypotheses L_2 and L_3, although collectively covering the same domain, individually have smaller domains and so are less general.

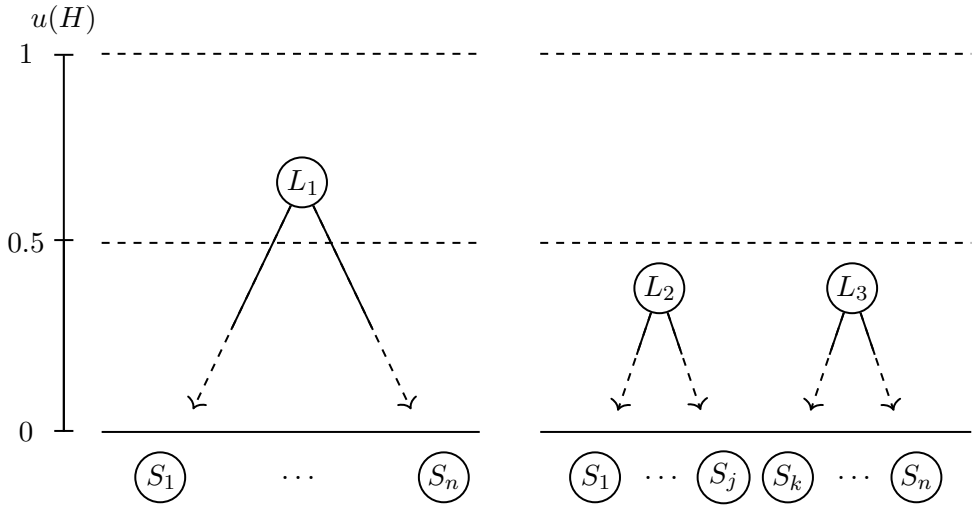

Figure 12: unified vs disunified hypotheses

While the number of subsidiary hypotheses still indicates the degree of unity, unity itself can be defined by a hypothesis's generality. Understanding unity as generality explains why Kant feels compelled to accept PPU. Reason, governed by FSU, demands that we complete the system of cognitions, which includes unifying given states under increasingly general laws. This demand, however, is only realisable if laws actually constitute such a system in nature. And so, while we cannot *know* nature to be systematic, the assumption that nature is systematic is a necessary means of realising the final end of Reason. The systemtaticity of nature is thus justified as a *doctrinal Belief*, as Gava (2018) has shown (see also Section 4.1).

But if we must believe that nature is systematically unified, then we must also believe that hypotheses representing nature as unified are more accurate. In other words, then, unified hypotheses, which derive given states from a single general law, seem to provide a more accurate depiction of nature, as Reason takes it to be. In contrast, disunified hypotheses, which derive

given states from multiple irreducible laws, contradict Reason's conception of nature. This line of reasoning would also explain why unified hypotheses seem more *natural* while disunified ones seem more *artificial*.

Natural hypotheses also seem to be more probable. For if a hypothesis corresponds to our rational understanding of nature, it clearly is more probably true than its disunified alternative. Indeed, Kant himself links artifice with improbability, and so naturalness with probability: 'The hypothesis is far more probable when all appearances can be explained by one cause than when many causes are needed to explain the events in question, because here there is already *too much artifice* [*Künstelei*]' (29:103; my emphasis). Drawing on his theory of Reason, Kant can therefore provide a compelling explanation of why unified hypotheses are more natural and so seem more probable.

This brings me to the end of this section. We have seen that the Criterion of Unity, although presented as a separate criterion, essentially introduces PPU, which, when paired with RPM, directs us to favour more unified hypotheses: more unified hypotheses are more probable, and we should favour the most probable hypothesis. By interpreting unity as generality, we have also seen that Kant has rich resources to justify PPU. But while PPU establishes *that* more unified hypotheses are more probable than less unified ones, we can still ask: *how much* more probable are unified hypotheses? How does the degree of unity, $u(H)$, affect the overall probability P of a hypothesis H, given the evidence E?

5.4 Beyond Bayes

On Kant's account, as I have reconstructed it, the probability of a hypothesis increases and decreases in proportion to (i) the number of given consequences, and (ii) the degree of unity. These two factors, however, are logically independent. One can envision a disunified hypothesis supported by numerous observations, as well as a unified hypothesis with minimal observational support. In the former case, hypothesis H has abundant evidence E but a narrow domain D; in the latter, it has a broad domain D but scant evidence E.

There will hence be potential trade-offs between the two, as Proops notes:

> [W]e may formulate the [...] demand of reason as a norm on theory construction that tells us to seek that theory among those fitting the data which maximizes the combination of explanatory power and ideological parsimony. Kant doesn't consider whether the principle in question could always identify a uniquely preferred empirically adequate theory, but since *trade-offs between explanatory power and parsimony* are plausibly possible, this question is one he might have considered. (2021: 43; my emphasis)

The goal of this and the next section will be to determine this trade-off by specifying Kant's probability function $P(H, E)$. This function must determine the probability of a hypothesis H given the evidence E, while also accounting for the degree of unity, $u(H)$. This section makes the start by considering whether Kant means for the probability of a hypothesis H *given* the evidence E, $P(H, E)$, to be interpreted as the probability of H *conditional* E, $P(H|E)$. And accordingly, whether Kant is committed to (a version of) Bayes's Law, which is entailed by the definition of conditional probabilities as a theorem.

What is Bayes's Law? Bayes's Law, formulated by Thomas Bayes, a contemporary of Kant, and published posthumously in *An Essay Toward Solving a Problem in the Doctrine of Chances* (1764), has been called 'the most important fact about conditional probabilities' (Joyce, 2019). The law calculates the posterior probability of a hypothesis H conditional the evidence E, $P(H|E)$, by multiplying the so-called 'likelihood' $P(E|H)$, i.e., the probability of E conditional H, by the prior probability of the hypothesis, $P(H)$, and then dividing the product by the probability of the evidence $P(E)$. Bayes's Law can formally be stated as follows:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Kant himself was likely unaware of Bayes's Law. Warda (1922) does not list Bayes's essay in Kant's library, and Kant never references Bayes in his writings. However, Kant was well acquainted with Jacob Bernoulli's *Ars*

Conjectandi (1713), which was part of his library, and which he mentions in his logical lectures (24:38). The broader point here is that the emerging theory of probability from the latter half of the 17th century permeated the intellectual atmosphere of the 18th century (Hacking 1975). Thus, it would not be surprising if Kant, unaware of Bayes's Law, committed himself to a set of views that align with Bayes's Law.

And indeed, Kant's claims on probability can be interpreted as highlighting various aspects of Bayes's Law. His assertion that 'the probability of the hypothesis' increases '[t]he more [it] suffices to explain the consequences' (24:439-40) may point to the fact that, within a Bayesian framework, hypotheses that better account for the evidence have a higher likelihood, $P(E|H)$, and thus a greater posterior probability, $P(H|E)$. Furthermore, as we continually update the hypothesis's probability with each new piece of evidence, its conditional probability converges towards 1, effectively resulting in an 'approximation towards certainty' (9:84).

Moreover, Kant's account of unity may be read as offering a solution to 'the problem of the priors' (Talbott 2016). This problem concerns the challenge of selecting a prior probability distribution, $P(H)$, before any evidence E is considered. Kant could address this by proposing that the prior probability is simply defined by the degree of unity of the hypothesis, such that $P(H) = u(H)$. This approach is compelling because it offers a non-subjective method for determining prior probabilities, even in situations with minimal antecedent data. By anchoring the prior in the hypothesis's unity, Kant has a systematic basis for assigning prior probabilities that avoids arbitrary or biased choices.

When viewed though this Bayesian lens, the trade-off between evidence and unity functions as follows: Unified hypotheses start with an advantage due to their higher prior probability. However, disunified hypotheses can gain ground if they increasingly account for the additional pieces of evidence, eventually surpassing the prior probability of unified hypotheses. This way of interpreting Kant's theory of probability would therefore yield a principled method for balancing the probability derived from evidence against the

probability derived from unity. This makes the Bayesian reading of Kant particularly compelling.

However, the Bayesian interpretation of Kant faces two issues. Foremost among these is that, according to Kant, the likelihood of the evidence conditional the hypotheses, P(E | H), must always either be 0 or 1 (or undefined). For as I have argued in Section 5.2, Kant defines explanation as involving logical entailment. Yet a hypothesis H either entails E or it doesn't; and so, it either explains E, i.e., $H \vdash_e E$, or it doesn't. Explanation itself doesn't seem to be a matter of degrees. But then, the probability updating turns into a game of last man standing – we just wait for the competing hypotheses to violate CC and so to be ruled out by contradicting evidence. But this account of updating seems anti-Bayesian in spirit.

The second issue is that we cannot simply equate the degree of unity, $u(H)$, with the prior probability, $P(H)$. The problem arises because the most general law, or highest genus, has a degree of unity of 1, which would imply a prior probability of 1. According to Kant's account of probability outlined in Section 5.1, a probability of 1 entails certainty, and for Kant, certainty entails to truth: 'objectively, there is no difference between truth and certainty' (R2481, 16:388).[285] Thus, merely by formulating the hypothesis of a highest genus, absent any supporting evidence, we would have proven that the hypothesis is true. But this is absurd.

We might, however, kill two birds with one stone and address both of these issues by tweaking Kant's notion of explanation. For explanation not only requires that the explanans logically entails the explanandum, but also that the explanans is more intelligible than the explanandum. However, we can not only specify that H is more intelligible than E, but also that it is more intelligible by a degree d, where d is (in part) determined by H's unity. More unified hypotheses are more general, and so respond to Reason's need for systematicity, making them more intelligible for us. But if intelligibility varies degrees, so can explanation, leading to the following two definitions:

[285] On the connection between truth and certainty, see also (2:290-1), (R2459, 16:378), (R2465, 16:382), (R2468, 16:383), (R2487, 16:391), (R3707, 17:244), (24:143-5), and (24:220-1)

Revised Definition of Explanation

A hypothesis H explains evidence E to a degree d iff for all $p \in E$:

(i) p can be logically deduced from H, and;

(ii) H is more intelligible than p by at least a degree d.

Definition of Intelligibility

A hypothesis H is more intelligible than p to a degree d iff

(i) H is really possible, and;

(ii) $u(H) - u(p) = d$.

With these definitions in place, we can try to salvage the Bayesian interpretation. The unity of the hypothesis, $u(H)$, no longer dictates the prior probability, $P(H)$, thus preventing the highest genus from being certain at the outset. Instead, the unity of the hypothesis H determines the degree d to which H can explain the evidence E. We can now use this degree d to determine the likelihood of the evidence E given the hypothesis H, $P(E|H)$. The likelihood is thereby no longer restricted to 0 and 1, but can assume any value in between, depending on H's degree of unity. Hypothesis updating would thus cease to be a game of last man standing and instead become a more nuanced process.

Indeed, this revised interpretation introduces a new dynamic into the trade-off between the evidence and unity of a hypothesis. Unified hypotheses no longer need a starting advantage because we can assign the same prior probability, $P(H)$, to all hypotheses. Instead, unified hypotheses increase the posterior probability of H conditional E more rapidly because a higher degree of unity, $u(H)$, acts as a likelihood booster. On the flip side, hypotheses with lower unity require more evidence to reach the same posterior probability as unified hypotheses. In the long run, however, both unified and disunified hypotheses approximate a probability of 1, provided they do not contradict the evidence.

As much as I like this revised Bayesian interpretation of Kant, it too faces textual challenges. As mentioned in Section 2.5, Kant persistently claims that evidence is only admissible if it is certain – my Mapping Argument

relies on this claim. For example, Kant insists that the explanandum must be 'completely certain' (24:888). He also states that 'that which is actually given', meaning the consequences that the hypothesis seeks to explain, 'is consequently certain' (A770/B798).[286] Since certainty typically entails a probability of 1, Kant is committed to think that evidence E is only admissible if $P(E) = 1$.

However, this commitment translates into an arithmetic problem. If the denominator in Bayes's Law is always 1, and the likelihood is at most 1 – satisfying the second Kolmogorov axiom – then the posterior probability can never exceed the prior probability. In fact, if the likelihood is less than 1, as is the case with all hypotheses other than those of the highest genus, the posterior probability will be less than the prior probability. Consequently, as more evidence accumulates, the hypothesis becomes increasingly less probable, approximating 0 rather than 1. This contradicts Kant's claim that 'the probability of the hypothesis' increases '[t]he more [it] suffices to explain the consequences' (24:439-40).

Let's run through an example. Consider a hypothesis H that is really possible and has a high degree of unity, $u(H) = 0.8$. Prior to any evidence, the hypothesis is equally probable as not, so $P(H) = 0.5$. Now, suppose the first piece of evidence E supports H. Since E is a singular state, $u(E) = 0$, giving an intelligibility delta of 0.8. Thus, H explains E to a degree of 0.8, making $P(E|H) = 0.8$. However, the given evidence E must be certain, and so $P(E) = 1$. Plugging these values into Bayes's Law, we find H's posterior probability to be only 0.4. As we gather more evidence in favour of H, the probability decreases further, first to 0.32, then to 0.26, and continues approaching 0 rather than 1.

In summary, while Kant appears to endorse several views that align with Bayesian principles, he ultimately cannot be fully committed to a version of Bayes's Law. Therefore, we should not model the probability P of a

[286] See also (A822/B850), (5:465-6), (R2680, 16:466-9), (R2690, 16:471), (24:647), and (24:746-7). As noted in Section 2.5, Kant likely adopts this commitment from Arnauld's *Logique de Port-Royal* (1662/1854: pt. 4, chs. 15 and 16).

hypothesis H given evidence E as a conditional probability, $P(H|E)$, since the definition of conditional probabilities entails Bayes's Law as a theorem. Instead, I will treat Kant's probability function as having two argument places, one for the hypothesis and one for the evidence, $P(H, E)$. In the next section, my goal is to reverse-engineer this function based on the insights we have gained so far.

5.5 Making Probabilities Precise

In Section 5.1, we have seen that Kant defines '[p]robability [as] a *fraction* where the sufficient ground of truth is the denominator, but the insufficient grounds of assent which I have are the numerator' (24:196). This definition gives us Kant's probability function for finite contexts, but it fails to generalise to infinite contexts – such as laws – where the sufficient objective ground is infinite. For no matter how many instances of a law we have observed, Kant claims that a law 'has *infinitely* many consequences' (24:220). Yet the probability of a finite numerator over an infinite denominator would always equal zero.

The aim of this section, then, is to reconstruct Kant's probability function for infinite contexts. Although Kant never outlines such a function, we can reverse-engineer it from his general commitments in probability theory. Specifically, the function must ensure (i) that hypotheses contradicting the evidence have a probability of 0; (ii) that hypotheses become more probable as evidence accumulates, converging on probability 1; and (iii) that unified hypotheses are more probable than disunified alternatives given the same evidence. While various functions fit this textual data, the following one does so effectively:

$$P(H, E) = \begin{cases} 0 & \text{if } \exists p(p \in E \wedge H \vdash_e \neg p) \\ \dfrac{n}{n + u(H)^{-1}} & \text{where } n := |\{p \mid p \in E \wedge H \vdash_e p\}| \end{cases}$$

The piece-wise function has two parts. The first part specifies that the probability P of a hypothesis H given the evidence E, $P(H, E)$, drops to 0 if the hypothesis contradicts the evidence – i.e., if there is a proposition $p \in E$ but $H \vdash_e \neg p$. The second, more complex, part indicates that $P(H, E)$ asymptotically approximates 1 as the number of consequences of H given by E increases. The number n of H's consequences is defined by the cardinality of the set of propositions $p \in E$ where $H \vdash_e p$. Additionally, the rate of at which $P(H, E)$ approximates 1 depends on the unity $u(H)$. To better understand this second part, consider the function's graphs for two different degrees of unity (Figure 13).

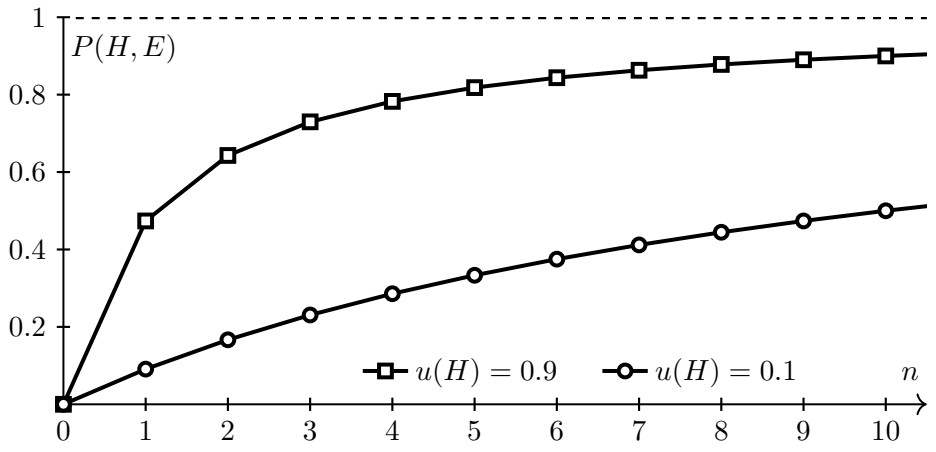

Figure 13: Kant's probability function

The reconstructed function preserves the trade-off between evidence-derived and unity-derived probability, similar to the revised Bayesian interpretation discussed earlier. The degree of unity acts as a probability booster, meaning that the probability of unified hypotheses increases more rapidly as the evidence accrues. However, even a unified hypothesis can be outweighed by a disunified one with substantial evidential support. Thus, while Kant cannot be strictly considered a Bayesian – since his views on probability jointly contradict Bayes's Law –, this version of the probability function positions

him as a quasi-Bayesian.[287]

The function $P(H, E)$ allows Kant to determine the probability of a hypothesis H, given the evidence E. But this is only one part of the full story. We must also consider how Kant determines the *joint probability* of two hypotheses, call them A and B, both being true simultaneously. This matters because disunified hypotheses come with one or more subsidiary hypotheses to explain the same evidence, covering the same domain as a unified hypothesis. So to compare a unified hypothesis with a *conjunction of* disunified hypotheses, we need another function that determines the probability of *A and B* given the evidence E, $P(A \wedge B, E)$.

Modern probability theory divides joint probabilities into two cases: those where A and B are *independent* and those where A and B are *dependent*. A and B are said to be independent iff the occurrence of A does not affect the probability of B, and vice versa. Conversely, A and B are said to be dependent iff the occurrence of A does affect the probability of B, or vice versa. This distinction matters because, arguably, two hypothesised laws, A and B, would be independent only if they don't stand in a genus species relation, but would be dependent if either A is the genus of B or B is the genus of A. We must therefore distinguish the two cases.

To illustrate both cases, recall that a law can be represented by its domain D, which is the set of all possible states the law aims to explain. In the first case, where laws A and B are not genera of each other, their domains are

[287] In my version of the probability function, hypotheses start with a prior probability of 0. This is will be textually contentious because Kant asserts that (i) 'hypotheses are opinion' (24:733); (ii) opinion is probable simpliciter: 'someone who adheres to an opinion holds the opinion to be something probable' (24:825); and (iii) an assent is probable simpliciter if it is more probable than its opposite, implying a probability greater than 0.5: 'Probability is an assent on insufficient grounds [...], which, however, bear a greater relation to the sufficient grounds than the grounds of the opposite' (9:81). See also (9:82), (8:396n), (R2583, 16:427), (R2600, 16:435), (R2602, 16:436), (20:299), (24:143-4), (24:194), (24:427), (24:433-6), (24:555), (24:742), (24:883), as well as Pasternack (2014b: 53) and Cooper (ms). However, the opposite of a hypothesis that p needn't be $\neg p$, but any of multiple competing hypothesis. Given that there are infinitely many hypotheses that can explain finite data points, the prior probability plummets to 0 – assuming that all hypotheses have the same prior probability. Therefore, I side with Chignell (2021: 119) against Pasternack (2014a: 63n) in arguing that, for Kant, an opinion can have a probability lower than 0.5. See also my Benzenberg (forthcoming-a). That said, my argument does not depend on this; the function can easily be adjusted to start at probability 0.5.

disjoint: $D_A \cap D_B = \varnothing$, and the same applies to their evidence, which is a subset of the respective domain: $E_A \cap E_B = \varnothing$. In the second case, where, say, A is the genus of B, the domain of B is a proper subset of A: $D_B \subsetneq D_A$; and likewise for the evidence, $E_B \subsetneq E_A$. These are the only scenarios because the domains of hypothesised laws cannot partially overlap; species must divide their genus exclusively and exhaustively (see Section 4.1). Figures 14 and 15 below illustrate these two possible cases, with 'Σ' representing the space of all possible states.

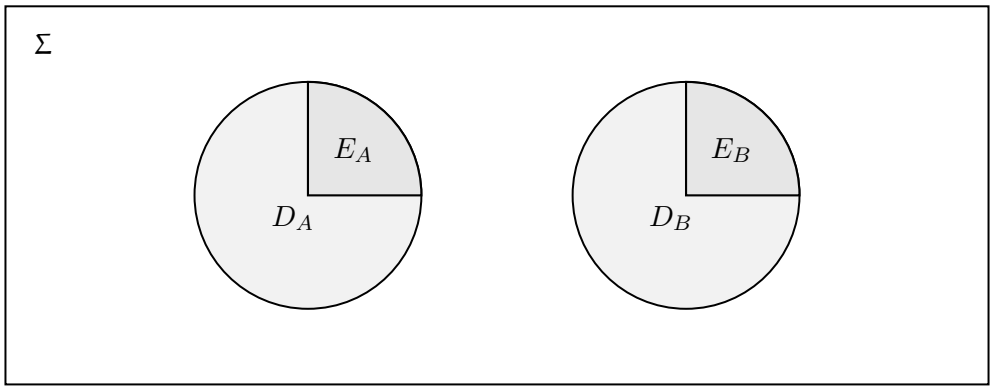

Figure 14: two independent laws

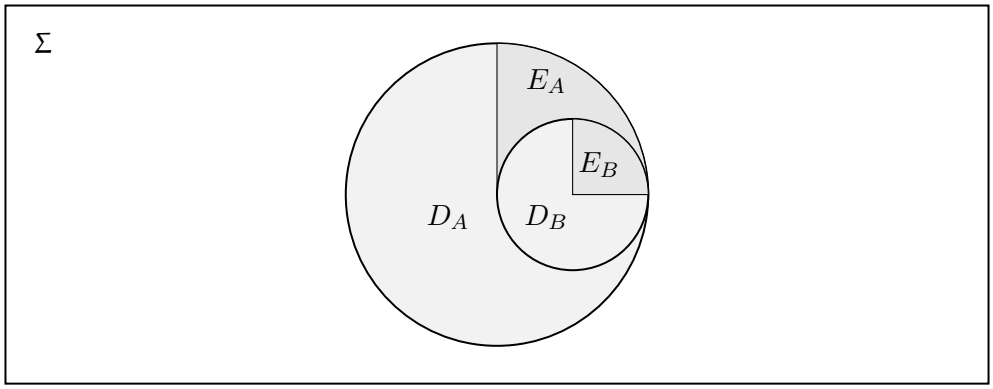

Figure 15: two dependent laws

Let's begin with the case where A and B are independent. We typically find this with theories that require multiple disunified hypotheses to cover the same domain and explain the same evidence as a more unified hypothesis. For to add explanatory power, subsidiary hypotheses cannot be species of the original hypothesis. And while Kant does not specify exactly how we are to calculate their joint probability, he notes that '[t]he more subsidiary hypotheses are necessary, the less probable the hypothesis is' (24:647). Since this follow directly from the standard definition, which determines the joint probability via the product, I suggest: if A and B are independent, then $P(A \wedge B, E) = P(A, E) \times P(B, E)$.

The second case, where A and B are dependent, is more difficult. The first thing to note here is that the hypothesis of the genus, say A, necessarily has a higher probability than the hypothesis of its species, say B, i.e., $P(A, E) > P(B, E)$ (as long as $E \neq \varnothing$). That's because all of the observations that support the species B also support the genus A; this is why the evidence for B, E_B, is a subset of the evidence for A, E_A, i.e., $E_B \subseteq E_A$. And even in the limit where all the evidence for A is also evidence for B, i.e., $E_B = E_A$, the probability of the genus A is still greater than the probability of the species B, because A, qua being more general, is more unified than B, i.e. $u(B) < u(A)$.

But this seems to create a problem because the lower probability of B would seem to drag down the joint probability of A and B. Here is why: (i) we have seen that the domain of B, D_B, is a (proper) subset of the domain of A, D_A, i.e., $D_B \subsetneq D_A$. Yet (ii) the present day definitions of joint probabilities entail: if $B \subsetneq A$, then $P(A \wedge B) = P(B)$. But, then, we are faced with the problem that any theory that hypothesises both a high-unity, high-probability genus and a low-unity, low-probability species of that genus – as we must do when we follow PS and look for lower species – will only ever be as probable as the low-probability species. However, that would be a strange result.

Let's look at an example. Suppose there are two sets of hypotheses: $\mathcal{H}_1$ consisting of a unified genus A, where $u(A) = 0.9$, and a its species B, where $u(B) = 0.1$ (as in Figure 13); and $\mathcal{H}_2$ consisting of two disunified hypotheses

about independent laws C and D, where $u(C) = u(D) = 0.5$. Suppose also that each law was supported by only one observation so far, meaning $n = 1$. If we crunch the numbers, we find that $P(A, E) \approx 0.47$, $P(B, E) \approx 0.09$ as well as $P(C, E) = P(D, E) = 1/3$. However, the joint probabilities would seem to be $P(A \wedge B, E) \approx 0.09$ and $P(C \wedge D, E) \approx 0.11$. So here the set of disunified hypotheses, $\mathcal{H}_2$ would be more probable than the unified set, $\mathcal{H}_1$. But that can't be.

Where is the mistake? The main confusion, I suggest, consists in mistaking the space of all possible states, Σ, for a sample space, Ω, defined as the set of all possible outcomes. By way of illustration, the sample space of rolling a six-sided dice is the set of rolling a 1,2,3,4,5, or 6, $\Omega = \{1, 2, 3, 4, 5, 6\}$. If A is the event of rolling an even number, $A = \{2, 4, 6\}$, and B is the event of rolling a 4 or 6, $B = \{4, 6\}$, then we have the situation where $B \subsetneq A$ – as was the case with laws. And in this case, it is clear that $P(A \wedge B) = P(B)$ because every outcome in B is also in A, making the event 'A and B' identical to B itself.

However, the space of all possible states is not a set of all possible outcomes, but instead the set of all mutually consistent states. Accordingly, the domain of a law doesn't denote the set of all states that would *each* make the law true, but instead denotes the set of states that would *jointly* make the law true. Yet all those states that jointly make the genus A true, also make its species B true, precisely because $D_B \subsetneq D_A$. It is therefore impossible that the genus A is true, but the species B is false yet possible that the species is true and the genus false – just as it is impossible that I roll a 4 and 6 but not an even number yet possible that I roll an even number but not a 4 and a 64.

The probability-affecting entailment thus runs in the opposite direction: the joint probability of the genus A and the species, given evidence E, is determined by the probability of A not B: if A is the genus of B, then $P(A \wedge B, E) = P(A, E)$. In fact, this interpretation receives independent support from Kant's discussion of hypothesis confirmation. In Section 5.1, I noted that for Kant, a hypothesis is confirmed if all of its infinitely possible consequences are given in observation, which is impossible in finite time.

However, he also acknowledges that a hypothesis can be confirmed by other means, providing the following example:

> The hypothesis must first be probable as inferences, but must also be able to be confirmed at the same time by other grounds. E.g. he who wants to explain the thunderstorm assumes an electric kind of matter in the air, and afterwards shows on other grounds that there must really be [...] electric matter in the air (29:103; my emphasis).[288]

Kant doesn't specify what these 'other grounds' are, but we can reasonably interpret the example as follows: to explain the occurrence of a 'thunderstorm' (E), we propose a moderately specific law (B) regarding 'an electric kind of matter in the air'. Suppose that further observations lead us to a more general hypothesis (A) about the laws of electromagnetism, which includes B as its species. In this scenario, B would be confirmed 'on other grounds' because it is entailed by the more general hypothesis A. Clearly, this confirmation cannot mean that B becomes certain, as A itself only has a degree of probability. Instead, B would seem to inherit the higher probability of A in that $P(B \wedge A, E) = P(A, E)$.

In this way, we also arrive at the correct verdicts for the example involving the unified set of hypotheses $\mathcal{H}_1$ and the disunified set $\mathcal{H}_1$. If the joint probability of two dependent laws is determined by their genus, then $P(A \wedge B, E) \approx 0.47$, while $P(C \wedge D, E) \approx 0.11$. Thus, the unified set $\mathcal{H}_1$ is more probable than the disunified set $\mathcal{H}_2$ and should be preferred according to RPM. More generally, theories with many independent subsidiary hypotheses will, given the same evidence, be much less probable than those with a unified law – even if we add further hypotheses about its species. Combining these insights, we can now define the function for joint probabilities more precisely as follows:

[288] Also consider this passage from the *Blomberg Logic*: 'A main reason for the truth of the hypothesis is also when one shows that the ground one has devised for the sake of the sufficiency of the consequences also deserves to be accepted for other ground. Here, then, one shows from other grounds that what has been devised ought to be accepted; one thus confirms the truth of the hypothesis' (24:223).

$$P(A \wedge B, E) = \begin{cases} P(A, E) & \text{if A is the genus of B} \\ P(B, E) & \text{if B is the genus of A} \\ P(A, E) \times P(B, E) & \text{otherwise} \end{cases}$$

This concludes my reconstruction of Kant's probability function(s). We see that Kant has the resources to precisely determine the probability of one or more hypotheses given the evidence. However, one might objective that my reconstruction overshoots the mark. For Kant repeatedly rejects the entire project of a logic of probability. For example, he writes in the *Jäsche Logic*: 'There has been much talk of a logic of probability (*logica probabilium*). But this is not possible; for [...] general rules of probability [cannot] be given anywhere' (9:82).[289] But if there cannot be any general rules of probability, then surely Kant also cannot think that there are probability functions, like the ones I have just reconstructed.

Two points in repones, the first point being that Kant confines his rejection of a logica probabilium to philosophy. In 'philosophical cognition', he argues, the objective grounds of assent are 'heterogenous', making it difficult for us to aggregate them to determine the sufficient objective ground; however, we only get to determine the probability of our assent by comparing the 'relation of insufficient grounds' we have 'to the sufficient ground' (9:83). Therefore, in philosophy, we lack strict rules for determining the probability of an assent. We can only 'pounder' but not 'number' our grounds of probability (24:555).[290]

However, the rejection does not generalise to mathematics or empirical inquiry. For here 'the moments of probability are homogeneous' (R2598, 16:435), allowing objective grounds to be combined into a sufficient one. By

[289] For Kant's rejection of a logica probabilium, see also (24:38), (24:555), and (24:879). For a further discussion, see Butts (1962: 194), Madonna (1992: 32-38), Gelfert (2006: 631), Chignell (2007a: 326-7), (2007b: 40-3), (2021: 122), and Pasternack (2014a: 70), (2014b: 75).

[290] See also (9:82), (R2591, 16:432), (R2598, 16:435), (24:38-9),(24:82), (24:196-7), (24:433), (24:555), (24:742), and (24:883).

comparing our insufficient grounds with a sufficient one, we can determine the probability of our assent according to 'rules of probability' (24:38). These rules are not just limited to mathematics, but we can 'mathematically calculate the degree of probability [...] *of various empirical things*' (24:196; my emphasis). Among these empirical things, I suggest, are empirical laws of nature, whose probability we can mathematically calculate with the functions I have reconstructed.

The second point is that I am not bound to the specific form of the probability function $P(H, E)$ introduced earlier. In fact, I do not need to assert that the exact function is known. Suppose there exists an additional variable a that scales the original probability function, either by stretching or compressing it. Even if the precise value of a is unknown, and therefore the exact probability of a hypothesis given the evidence cannot be calculated, the relative probability between two hypotheses, given the same evidence, remains invariant under this transformation. Consequently, even without knowing the exact probability function, we retain sufficient information to apply RPM effectively.

This concludes my discussion in this chapter. To find further L conditions, Reason draws probabilistic inferences from given states to hypothesised laws. These inferences, I have argued, are governed by RPM, which tells us to always choose the most probable hypothesis. All other criteria for hypothesis selections are reducible to this rule. I have also reconstructed Kant's probability functions for determining which hypothesis is most probable given the evidence. Notably, a hypothesis about a law increases in probability as more of its consequences are confirmed by experience and the greater its degree of unity. Overall, we find that Kant has a fairly coherent theory of probabilistic inferences.

6

Mental Chemistry

Chapters 4 and 5 have addressed the Material Question by specifying the precise content of the Principle of Reason. I have argued that the principle manifests in two distinct formulations: the Formula of the Totality of Conditions (FTC) and the Formula of Systematic Unity (FSU). FTC directs us to find the totality conditions for our given cognitions, while FSU requires that we organise these cognitions into a hierarchical system. To fulfil these demands, we draw probabilistic inferences from particular cognitions to general ones, including inferences from particular state to general laws – laws that constitute powers of nature.

Building on the foundations laid in Chapters 4 and 5, this chapter returns to the Epistemic Question. I argue that we can come to know that we have Reason in empirical psychology by probabilistically inferring the Principle of Reason from the states we observe in inner sense. Here is the plan: after textually motivating Epistemic Symmetry (§ 6.1), I address the two outstanding problems: the Problem of Perfection (§ 6.2) and the Problem of Aprioricity (§ 6.3). In concluding, I tackle a few remaining, if less pressing, challenges (§ 6.4) before discussing the metacritical merits of the account I have sought to defend throughout this thesis (§ 6.5).

6.1 Epistemic Symmetry

In Chapters 2 and 3, I have justified and defended Ontological Symmetry, which is the claim that mental powers – including Reason – are powers of inner nature, just as physical powers are powers of outer nature. Crucially, both types of powers are governed by causal laws that determine their activity. Ontological Symmetry motivates Epistemic Symmetry: if both mental and physical powers share the same ontological status, we can come to know mental powers – including Reason – via empirical psychology, just as we can come to know empirical powers via empirical physics.

In Chapter 5, I have reconstructed Kant's general account of how we come to know natural powers: we infer their laws probabilistically from observed states. This inference, moreover, is demanded by the Principle of Reason. For it is in inferring general laws form particular states that we seek further L conditions and find higher genera, which, as I have shown in Chapter 4, is required by FTC and FSU. Since there is no indication that Kant restricts FTC and FSU to cognition from outer sense, we should expect both to apply equally to cognition from inner sense. Epistemic Symmetry thus becomes an exegetical default: we come to know both mental and physical powers via probabilistic inferences.

In addition to serving as an exegetical default, however, Epistemic Symmetry can also be directly supported by Kant's text. While I hinted at some of this support in section 4.1, I will substantiate the view more fully in this section. Specifically, I will show that Kant intends FSU to apply to cognitions from inner sense: we must seek to unify inner states under more general laws, by which we come to know the mental powers governed by these laws. As my main witness for this reading, I will cite a remarkable passage from the Appendix to the Transcendental Dialectic, which I will call the 'Primary Passage'.[291] Here is the full passage with numbered sentences for easy reference:

[291] By way of a brief biographical note, the Primary Passage motivated me to write this thesis. Much of what I have argued in Chapters 2 through 5 aims to articulate the commitments Kant must adopt for the passage to hold meaning.

> [1] We will illustrate this by an example of the use of reason. [2] Among the various types of unity according to the concepts of understanding is also that of the *causality of a substance*, which is called *power*. [3] The different appearances of the same substance show at first sight so much dissimilarity that one must therefore initially assume almost as many different powers of the same substance as there are effects that stand out, as in the *human mind*, sensation, consciousness, imagination, memory, wit, power of discrimination, pleasure, desire, etc. [4] Initially, a *logical maxim commands that this apparent diversity be reduced as much as possible* by discovering the hidden identity through comparison, and by seeing whether imagination is not connected with consciousness. [5] Memory, wit, the power of discrimination, perhaps even understanding and *reason*. [6] The idea of a basic power, of which logic does not even determine whether such a thing exists, is at least the problem of a systematic representation of the manifold powers. [7] *The logical principle of reason requires that this unity be brought about as far as possible*, and the more the appearances of one and the other power are found to be identical among themselves, *the more probable* it becomes that they are nothing but different expressions of one and the same power, which (comparatively) can be called its basic power. [8] One proceeds in the same way with the others. [9] The comparative basic powers must in turn be compared with each other in order to bring them closer to a single radical, i.e., absolute basic power by discovering their uniformity. [10] However, this rational unity is merely hypothetical. [11] It is not asserted that such a unity must actually be found, but that *it must be sought for the sake of reason*, namely for the establishment of certain principles, for which various rules, *which experience may provide*, must be sought, and, where it is possible, *systematic unity* must be brought into cognition in this way. (A648-50/B676-8; my emphasis)

The Primary Passage encapsulates many of the core claims of this thesis, so let's unpack it carefully. First, note that the passage appears in the midst of Kant's discussion of FSU, specifically PH. Referring to PH, Sentence (4) states that 'the logical maxim commands that [the] apparent diversity be reduced as much as possible', while Sentence (7) reiterates that '[t]he logical principle of reason requires that this unity be brought about as far as possible'. In Sentence (11), Kant further specifies that, 'where it is possible, systematic unity must be brought into cognition in this way' and moreover that this 'must be sought for the sake of reason'.

Moreover – and this is the main point –, Sentence (3) applies Reason's demand for systematic unity to the *powers* '*in the human mind*'. These powers are defined in Sentence (2) as 'the causality of a substance,' specifically the mind, whose activities Sentence (3) identifies as 'effects,' reinforcing Ontological Symmetry. To achieve a 'systematic representation of the manifold powers,' as stated in Sentence (6), we must first seek a 'basic power,' as stated in Sentence (7), and then attempt to reduce these 'comparative basic powers' to an 'absolute basic power,' as stated in Sentence (9). The laws governing these basic powers would be the genera and L conditions of their more specific counterparts.

As I had noted in Section 4.1, Kant's account of systematic unity is meant to reconcile the positions of Wolff and Crusius. Wolff argued that all mental powers could be reduced to, what Sentence (9) calls, one 'absolutely basic power,' which Wolff asserts to be the power of representation. Crusius, on the other hand, maintained that there are, as Sentence (3) puts it, 'as many different powers [...] as there are effects' in the mind. Kant attempts to bridge these views by claiming in Sentence (11) that the 'rational unity' provided by an absolute basic power 'is merely hypothetical,' meaning, as clarified in Sentence (12), that '[i]t is not asserted that such a unity must actually be found, but that it *must be sought*'.[292]

Kant further clarifies that the search for these powers involves probabilistic inferences. Sentence (7) states that 'the more the appearances of one and the other power are found to be identical among themselves, *the more probable* it becomes that they are nothing but different expressions of one and the same power'. This fits with my reconstruction in Chapter 5, where I argue that more general hypotheses are more unified and, therefore, more probable. This idea is also echoed here: 'We [...] have primitive and derivative powers in every substance: it is assumed, with a *great probability*, that there must be a primitive power from which all others derive' (29:770; my emphasis).

[292] Other relevant passages where Kant attempts to find a middle ground between Wolff and Crusius include (2:59), (28:29), (28:145), (28:261-2), (28:431), (28:512), (29:770), and (29:882). For further discussion on this dynamic, see also Heßbrüggen-Walter (2004: 146-9) and Section 4.1.

What makes the Primary Passage so remarkable is that it also illustrates the process of seeking increasingly general powers through specific examples. Sentence (3) suggests that we start with the 'effects that stand out', presumably in the inner sense, and infer low-generality powers like 'sensation, consciousness, imagination, memory, wit, power of discrimination, pleasure, desire, etc.'. By uncovering their 'hidden identity', however, we ascend to more basic powers with higher-generality. Sentence (5) offers examples, such as '[m]emory, wit, the power of discrimination, perhaps even understanding *and reason*' – implying that by systematising the powers of the mind, we may ultimately discover Reason.[293]

Thus, the Primary Passage directly attests to Epistemic Symmetry. Reason's demand for systematicity, as outlined by FSU, applies not only to physical powers but also – and paradigmatically! – to mental powers. Or as Frierson notes: 'Theoretical reason places on human beings the demand to investigate the diversity of natural phenomena and systematically explain these. And this demand of reason is not limited to purely physical phenomena but includes psychological ones' (2014: 50).[294] Insofar, then, as we discover powers through systematising them, we come to know mental powers just as we come to know physical powers. And Kant explicitly extends this symmetry to include Reason.

One aspect, however, that isn't clear in the Primary Passage is its intent to outline the procedure of *empirical* psychology. Although Sentence (11) alludes to the 'various rules, which *experience* may provide', interpreters in the secondary literature often view the Primary Passage as describing a procedure within rational psychology. For example, Proops argues that the

[293] In *Metaphysics* L_1, Kant similarly references 'reason and the understanding' as examples of 'basic powers we must assume' to 'explain the phenomena of the soul' (28:262). See also Heßbrüggen-Walter (2004: ch. 7.3).

[294] Kant elsewhere concludes that '[t]here is thus revealed a system of the powers of mind' (20:247). On the system of mental powers, see also (5:198), (6:211-4), (28:564), and (29:773-82). I thus agree with Frierson, who notes that the principles of homogeneity and specification also apply to mental powers (2014: 6-7), and so that we can 'investigate the human mind in terms of natural causality' (2014: 8). Moreover, Kraus correctly highlights that Reason's demand for systematicity also represents a 'demand for *inner* systematicity' (2020: 270; my emphasis). On these issues, thee also Heßbrüggen-Walter (2004: 48-9, 161-7) and Dyck (2014: 205-7).

Primary Passage addresses the '*rational psychologist's* belief', and specifically Wolff's belief, 'that all the powers of the soul – imagination, memory, wit, and so forth – are reducible to a single fundamental power' (2021: 45; my emphasis).

Dyck even suggests that the Primary Passage regresses into Kant's pre-Critical rational psychology, falling short of the Critical standard established in the Paralogisms. Referring to the Primary Passage, he asserts that 'Kant's general endorsement in the Appendix to the Transcendental Dialectic of the method and the substance of his *pre-Critical rational psychology* considered as a doctrine of inner appearances stands in stark contrast to his treatment in the Paralogisms of rational psychology considered as a doctrine of the nature of the soul as such' (2014: 224; my emphasis). Dyck's interpretation thus forces us to accept a patch-work reading of the first *Critique*.

In opposition to Proops and Dyck, I argue that the Primary Passage is about empirical, not rational, psychology. Kant persistently emphasises throughout the Appendix that he is concerned with the empirical use of reason. He writes that 'reason [...] [in] its own *empirical use* [...] orders' our cognition (A643/B672; my emphasis). He further clarifies that the principles of systematicity govern only 'the empirical use of reason' (A663/B691), and so, for example, PS demands only 'empirical specification' (A657/B685). Indeed, Reason systematises 'all possible empirical concepts' (A652/B680) and all 'empirical laws' of nature (A664/B692). And this is just the tip of the textual iceberg.[295]

Indeed, in the second part of the Appendix we find a passage, which I have already quoted in full in Section 2.5, which is similar to the Primary Passage. There, Kant clarifies that Reason follows 'principles of systematic unity in explaining the appearances of the soul, namely: to regard all determinations as in one single subject, all powers as derived, as far as possible, from one single basic power [...]', whereby Reason seeks to complete

[295] Kant specifies that he is concerned with the 'the greatest systematic unit in the empirical use of our reason' (A670/B698). See also (A671/B699), (A674/B702), (A675/B703), (A676/B704), (A677/B705), (A678/B706), (A680/B708), (A682/B710), (A685/B713), (A688/B716), (A692/B720), (A693/B721), (A698/B726), and (A701-2/B729-30).

the 'systematic unity of all appearances of *inner sense*' (A682-3/B710-1; my emphasis). Reason, guided by FSU, thus operates on phenomena of inner sense. Yet empirical psychology is nothing but a 'physiology of inner sense' (A347/B405) that studies the 'sum total of inner perceptions under natural laws' (7:141).

Kant gets even more explicit in a near-identical counterpart to the Primary Passage found in the *Metaphysics* L_1 of the late 1770s. This counterpart leaves no doubt that Kant is concerned with empirical psychology: 'The question is still raised to conclude *empirical psychology*: whether all the powers of the soul are united and can be derived from one basic power, or whether different basic powers are to be assumed in order to explain all the actions of the soul from them' (28:261; my emphasis). And a few lines further down: 'The sentence [...] that all different actions of humans must be derived from different powers of the soul, serves to treat *empirical psychology* all the more systematically' (28:262; my emphasis).[296]

Indeed, even the rational psychologists who Kant critiques recognise the crucial role of empirical psychology. Wolff observes in his *Anmerkungen*: 'We find in experience that the soul is constantly at work, but its thoughts are not always of the same type. Sometimes it has perceptions, sometimes imaginations, sometimes clear concepts, sometimes rational conclusions [...], and so on' (1740/1983: § 265, 434-5). Similarly, Crusius notes in his *Entwurf*: 'From the manifold interconnection of the powers and activities of a living substance, certain physical laws of the activities can be abstracted, which [...] are learned from experience [...]. These can be called the empirical laws of pneumatology' (1745/1964: §459, 902-3).[297]

While there is more to be said about the subtle interplay between empirical and rational psychology in Section 6.3, it is textually evident that the Primary Passage is intended to specify the epistemic procedure of empirical

[296] Kant also writes that 'psychology amounts to this: deriving diverse powers, which we *know only through observation*, as much as possible from basic powers' (28:564; my emphasis). See also (2:60).

[297] Meier similarly notes in his *Auszug* 'that I think is a sensation and an immediate experience, but that I have a faculty to think is an indirect experience' (1752: § 201, 57).

psychology. Empirical psychology begins with the effects or states observed in inner sense and then probabilistically infers increasingly general mental powers and their laws, striving to complete the system of all such powers and laws; just as empirical physics begins with states observed in outer sense and probabilistically infers the laws of physical powers. The two disciplines are procedurally symmetrical.

This symmetry between (empirical) psychology and (empirical) physics is further supported by numerous passages from all over the corpus. For example, Kant writes in the *Inaugural Dissertation* that '[p]henomena are reviewed and set out, first, in the case if the phenomena of outer sense, in physics, and secondly, in the case of the phenomena of inner sense, in empirical psychology' (2:397). Likewise in *Metaphysics* L_1: 'Psychology is thus a physiology of inner sense or of thinking beings, just as physics is a physiology of outer sense or corporeal beings' (28:224; see also 28:656). And there are more such passages.[298]

Coloured by this additional text, the Primary Passage firmly supports Epistemic Symmetry. Just as we establish physical powers in empirical physics by systematising their laws through probabilistic inferences from the states observed in outer sense, so we establish mental powers in empirical psychology by systematising their laws through probabilistic inferences from the states observed in inner sense. Since Kant explicitly extends this argument to Reason, the Primary Passage directly supports the Main Claim of this thesis – my answer to the Epistemic Question – namely that, according to Kant, we can come to know that we have Reason via empirical psychology.

More specifically, we come to know that we have Reason in empirical psychology by following the Principle of Reason, which dictates that we seek a complete and systematic explanation of all phenomena in the inner sense.

[298] See also (4:295), (4:467). (7:143), 20:308), (28:222-4), (28:656), (28:670), (29:754-6), and (29:954). In the *Metaphysics Dohna* from 1792/3, Kant even introduces psychology as a species of physics: 'All powers are divided into primitive or basic powers and derivative or derived powers. We seek to reduce the vires derivativae to the primitive powers. All *physics*, both of the body and *of the mind*, the latter *being called psychology*, comes down to this: to derive the various powers, which we know only through observation, as far as possible from basic powers' (28:664; my emphasis).

For it is in seeking such an explanation that we hypothesise mental powers and their laws, including the power of Reason and its law, the Principle of Reason. Thus, Reason is the power that has the power to establish its own existence, which makes it unique among all other powers of nature – inner and outer. Empirical psychology provides the context for this 'process of self-isolating reason', as Kant puts it in a letter to Marcus Herz (10:144).[299]

Although the primary passage demonstrates *that* Kant must think that Reason can isolate itself within empirical psychology, it remains unclear *how* exactly this is done. Specifically, there are two major issues still to be addressed: the Problem of Perfection and the Problem of Aprioricity. How can we infer the Principle of Reason from states in inner sense, if these states either do not exhibit perfect reasoning or if the Principle of Reason is a priori? These are crucial questions, and I will address them in detail in the following two sections, Sections 6.2 and 6.3, respectively.

6.2 Reasoning About Reason

In this section, I address the Problem of Perfection. The problem questions how we can come to know that we possess a faculty governed by the Principle of Reason, simply by observing our actual reasoning in in inner sense. As discussed in Chapter 3, the Principle of Reason describes how a perfect reasoner – or indeed, the faculty of Reason itself – in fact operates. However, as finite beings, we are not perfect reasoners, nor is Reason our only faculty, which is why the Principle of Reason is normative rather than descriptive for us. Thus, it would seem that by observing our actual, imperfect reasoning in inner sense does not suffice to establish that we have a faculty governed by the Principle of Reason.

Our finitude introduces two distinct challenges. The first is that we have only existed for a finite time (even if we turn out to be immortal),

[299] This process – which Rohs might describe as 'self-illumination of reason' (1987: 381) – then answers Heßbrüggen-Walter's question: 'How can a faculty grasp [...] itself?' (2004: 262). Reason grasps its own existence by systematically explaining phenomena of inner sense. As Pinker puts it: '[T]hat is the power of reason: it can reason about itself' (2021: 71).

and so we can never have observed in inner sense that we have actually sought *the totality* of L conditions for a given cognition; or that we sought to complete the system of all cognitions. Instead, all we observe in inner sense is that we sought this-or-that L condition for this-or-that cognition; or that we sought this-or-that genus for this-or-that species. This limitation suggests that empirical psychology replicates the problem of the act-awareness solution discussed in Section 1.2 – the problem that we can never be aware of Reason's infinite act over a finite time.

The second challenge is that we, as finite beings, are subject to the corrupting influence of sensibility. As I have argued in Chapter 3, this leads us to *assert* rather than *seek* the totality of L conditions for a given conditioned; or assert that the system is complete, rather than seeking to complete it. What we observe in inner sense, then, is that we – i.e., 'we' the metaphysically-minded people – assert that there is a totality of L conditions, for example, by claiming to know that we have an immortal soul, that the world has a first cause, that God exists, and so on. This corrupting influence of sensibility is also why Reason becomes what I had called a zombie power.

Though our finitude paints a grim epistemic picture, our rational nature still offers a glimmer of hope. While we may not directly observe perfect reasoning through inner sense, the hypothesis that we possess a perfect power of Reason might nonetheless be the most probable explanation for our imperfect reasoning. Admittedly, there is no direct, logical link between our power of Reason and the imperfect reasoning we observe – 'from power, the effect does not always follow (logically)' (28:26).[300] However, Kant's theory of probability and account of abduction, reconstructed in Chapter 5, could help us overcome this epistemic limitation.

Indeed, this will be my central claim in this section. The hypothesis that we possess Reason is the most probable explanation for the imperfect reasoning we observe in inner sense, and should therefore be adopted based

[300] On this point, I concur with Heßbrüggen-Walter (2004: 135), who observes that the real opposition of mental powers means that they can cancel each other out, thereby undermining any analytic link between powers and perceived activities, since perceived activity can be the result of multiple powers.

on RPM. But I want to set realistic expectations: this section presents an argument within first-order empirical psychology, so my aim is not to offer a conclusive case. Rather, my aim is to provide the overall flavour for the sort of argument that Kant might invoke to justify the existence of Reason. A conclusive case would require a more comprehensive analysis of inner states and consider a wider range of potential hypotheses.

In fact, this section considers and compares only three hypotheses: (i) the hypothesis H_R that we have Reason, defined as the faculty which, for any given cognition C, seeks the totality of L conditions of C; (ii) the hypothesis H_I that we have a more moderate faculty of inference which, for any given cognition C, seeks n-many L conditions of C (for some $n \in \mathbb{N}$); and (iii) the hypothesis H_U that we have Unreason, stipulated to be the faculty which, for any given cognition C, asserts rather than seeks the totality of L conditions of C.[301] My claim is that the hypothesis that we have Reason, H_R, is the most probable of the three.

Let's start by comparing the hypothesis that we have Reason, H_R, with the hypothesis that we have a moderate power of inference, H_I. To do this, we must first consider the evidence E that each hypothesis seeks to explain. While we can only ever have observed ourselves in inner sense to have sought some, not all, L conditions, we can still distinguish two cases: either we have at least once observed ourselves to have sought more than n-many L conditions for a given cognition (let us call this set of observations '$E_{>n}$'), or we have always observed ourselves to have sought n-many or fewer L conditions for our given cognitions (let us call this set of observations '$E_{\leq n}$').

I will consider both cases in turn. The first case is straightforward, because here the hypothesis which posits a moderate power of inference fails to explain the observed evidence and so violates CC, i.e., there is some $p \in E_{>n}$ such that $H_I \nvdash_e p$. In contrast, the hypothesis that we have Reason explains the observed evidence, as any number of L conditions we have

[301] This definition of Unreason should not be conflated with Kant's own concept of 'Unreason' ('*Unvernunft*') as 'rulelessness' (9:139). Refer also to (16:94-5), (16:780), as well as Hutter (2015: 2419). After all, Unreason as I define it follow a clear principle: for any given cognition *C*, if *C* is L conditioned, then the totality of L conditions for *C* must be asserted.

sought can be seen as part of a process driven by a power that ultimately aims to seek all L conditions, and so for all $p \in E_{>n}$, $H_R \vdash_e p$. Based on my reconstruction of Kant's probability function in Chapter 5, it follows that, given the evidence $E_{>n}$, it is more probable that we have Reason than just a moderate power of inference, i.e., $P(H_R, E_{>n}) > P(H_I, E_{>n})$.

The second case is slightly more subtle. For here both hypotheses can explain the observed evidence, i.e., for all p $E_{\leq n}$, $H_I \vdash_e p$ and $H_R \vdash_e p$. However, I argue that the hypothesis that we have Reason is still more probable because it exhibits a greater degree of unity, i.e., $u(H_R) > u(H_I)$. After all, H_R accounts for a greater domain of phenomena because it not only covers all possible cases where we have sought n-many of fewer L conditions, but also all possible cases where we would have sought more than n-many L conditions. According to Kant's theory of probability, more unified hypotheses are more probable given the same evidence, making $P(H_R, E_{\leq n}) > P(H_I, E_{\leq n})$.

Thus, for both sets of evidence, $E_{>n}$ and $E_{\leq n}$, the hypothesis that we have Reason is more probable than the hypothesis of a moderate faculty of inference. This conclusion is striking because the Criterion of Unity, in particular, supports the hypothesis that we possess a faculty seeking *all* L conditions, rather than just *some*, due to the greater degree of unity. Remarkably, this holds true even though our observations in inner sense have only shown us seeking some L conditions. The Criterion of Unity, therefore, enables us to establish an infinitely demanding faculty on finite evidence. Empirical psychology overcomes the metacritical challenges that burden theories like the act-awareness account (more on this in Section 6.5).

But is Reason also more probable than Unreason? Before answering this question, it's important to acknowledge a potential objection: one might argue that Unreason is really impossible, thus violating both CP and CC. After all, Unreason is stipulated to be a faculty that asserts totalities, which, when exercised correctly, leads to dialectical errors and false judgments. A faculty that, by its nature, produces error might be considered really impossible because it would not 'agree with the formal conditions of experience'

(A218/B265), specifically it would conflict with the truth-directed principles of understanding.

While this objection has merit, I will charitably assume that Unreason is really possible.[302] To see whether Reason or Unreason is more probable, we must first determine the relevant evidence. Suppose we only consider our totality-asserting behaviour, meaning we observe ourselves in inner sense to assert an immortal soul, that the world has a first cause, that God exist, and so on. Let's call this set of evidence 'E_a' ('a' for asserting). The Unreason hypothesis easily explains E_a because a faculty that asserts totalities would naturally explain why we make such assertions. So for all $p \in E_a$, $H_U \vdash_e p$. Yet the Reason hypothesis cannot explain E_a: if we only ever seek totalities, we don't assert them. And so, for some $p \in E_a$, $H_R \nvdash_e p$

However, this is not the whole story. As we explored in Chapter 3, Kant suggests that we can explain our totality-asserting behaviour as the combined result of Reason and the corrupting influence of sensibility – illustrated by the power parallelogram in Figure 2. The idea is that our practical sensibility instils a desire for knowledge, leading us to assert totalities to satisfy this desire. If we introduce sensibility as an auxiliary hypothesis, H_S, we find that together, H_R and H_S can indeed explain our totality-asserting behaviour. Thus, for all $p \in E_a$, $\{HR, HS\} \vdash_e p$, meaning that the combined hypotheses of Reason and sensibility can account for the evidence just as well as Unreason.

While H_R and H_S together satisfy the CC, it seems they lack the unity of H_U. Since we need two hypotheses to explain the same evidence, E_a, rather than just one unified hypothesis, it appears that $u(H_R) < u(H_U)$ and $u(H_S) < u(H_U)$. Given that less unified hypotheses are less probable when considering the same evidence, and given that the joint probability of two

[302] Indeed, one might question whether the hypothesis that we have Reason is really possible. Kant introduces Reason as an infinitely demanding power, directing us to seek the totality of conditions. However, it is uncertain whether such an infinite power could be part of the natural order, for Kant also denies the real possibility of an infinitely powerful being, such as God, as being part of nature (see Section 5.2). Consequently, the hypothesis of Reason might face a similar challenge as the God-hypothesis, potentially violating CP. I will not pursue this question further here.

independent hypotheses is the product of their individual probabilities – with Reason and sensibility clearly being independent –, it would seem that the joint probability of H_R and H_S is still lower than that of H_U, meaning that $P(H_R \wedge H_S, E_a) < P(H_U, E_a)$.

Where do we go wrong? The mistake lies in focusing solely on our totality-*asserting* behaviour as evidence. Instead, the evidence should also include our totality-*seeking* behaviour. While it is true that we can observe ourselves in inner sense asserting various totalities when we engage in dogmatic metaphysics, it is far more common for us to observe ourselves seeking further L conditions in various forms of empirical inquiry. For example, in astronomy we seek to explain the movement of celestial bodies; in medicine we seek to explain the causes of cancer; in law we seek to explain who committed the murder; and so on. Let's call this set of evidence, where we observe ourselves seeking L conditions, 'E_s' ('s' for seeking).

So, the goal must be to explain both our totality-asserting and totality-seeking behaviours, or $E_a \cup E_s$. Why do we sometimes assert, and other times seek a series of L conditions? We've already established that Reason alone can account for all totality-seeking behaviour, and that, combined with sensibility, it can also explain all totality-asserting behaviour. Thus, for any $p \in E_a \cup E_s$, $\{H_R, H_S\} \vdash_e p$. In contrast, the hypothesis of Unreason seems to fall short, as it cannot explain our totality-seeking behaviour. If we only ever assert totalities, we would never seek them. Therefore, there are some $p \in E_a \cup E_s$, where $H_U \nvdash_e p$.

However, just as the Reason-hypothesis is supported by the addition of sensibility to explain our totality-asserting behaviour, one might argue that the Unreason hypothesis could similarly be strengthened by sensibility to explain our totality-seeking behaviour. For instance, one could suggest that we sometimes seek rather than assert the totality of L conditions because we want to cherish all the intricate details that nature presents. Just as totality-asserting behaviour can be seen as the combined result of both Reason and sensibility, totality-seeking behaviour might be understood as the combined effect of Unreason and sensibility. Figures 16 and 17 illustrate this idea.

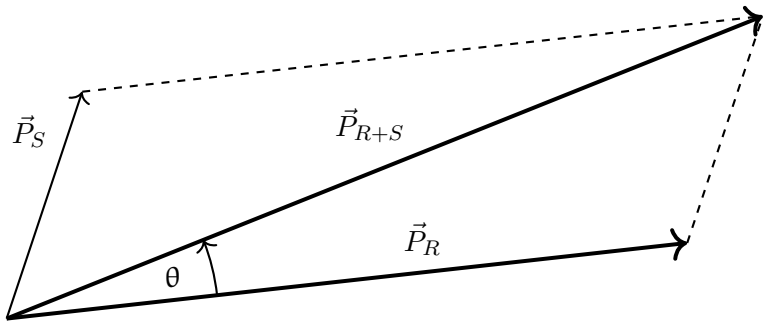

Figure 16: power parallelogram, v1

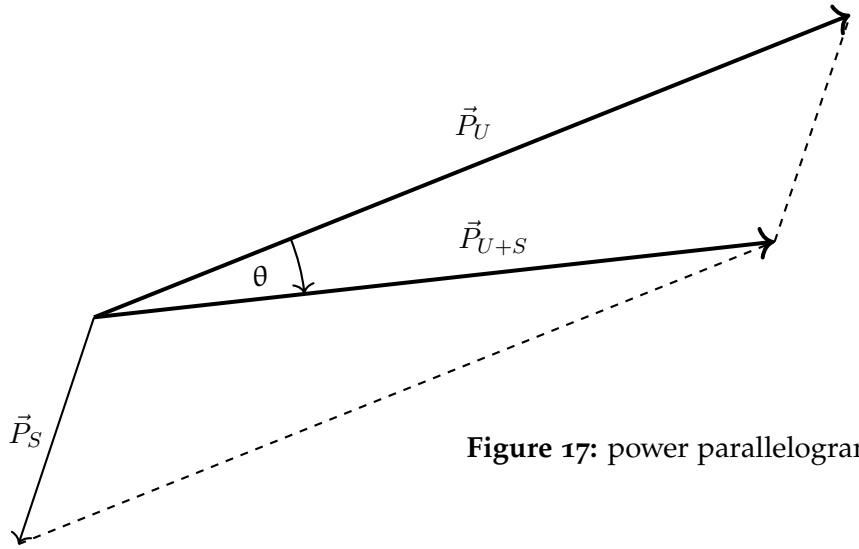

Figure 17: power parallelogram, v2

Let me briefly explain both figures. Figure 16 is a version of the power parallelogram from Figure 2. In this diagram, vector $\vec{P}_R$ represents the activity of Reason, vector $\vec{P}_S$ represents the activity of sensibility, and the diagonal vector $\vec{P}_{R+S}$ represents the aggregate activity of both Reason and sensibility. We can observe in inner sense both the combined vector $\vec{P}_{R+S}$ ($= E_a$) and the vector $\vec{P}_R$ ($= E_s$) – we can see ourselves asserting and seeking totalities –, which is why I highlighted the vectors in bold. Yet we only hypothesise a power corresponding to $\vec{P}_R$. The totality-asserting activity of $\vec{P}_{R+S}$ is interpreted as an aggregate activity, deviating from the totality-seeking behaviour by an angle θ.

Figure 17, on the other hand, presents an alternative scenario. Here, vector $\vec{P}_U$ represents the activity of Unreason ($= E_a$), which was the combined vector in the original account. Vector $\vec{P}_S$ again represents sensibility, which now points in a different direction, reflecting a different practical desire. In this alternative account, our totality-seeking behaviour is represented as the combined vector $\vec{P}_{U+S}$ ($= E_s$), which was previously attributed to Reason. In this scenario, the angle θ represents the degree to which we deviate from the path of Unreason when we seek rather than assert totalities.

We can now ask: which set of hypotheses is more probable given the total evidence of both E_a and E_s? Is it the combination of Reason and sensibility, H_R and H_S, or the combination of Unreason and sensibility, H_U and H_S? To answer this, let's assume that the probability of H_S is equal in both scenarios. With this assumption, we must determine whether Reason or Unreason is more probable initially. I propose that we initially assume Reason to explain our totality-seeking behaviour, E_s, and Unreason to explain our totality-asserting behaviour, E_a. Therefore, the crucial question is whether H_R is more probable given E_s or H_U is more probable given E_a, i.e., whether $P(H_R, E_s) > P(H_U, E_a)$.

I argue that H_R is more probable given E_s. First, consider that the domain of H_R is likely broader than that of H_U because totality-seeking behaviour occurs across various fields – such as astronomy, medicine, and law – whereas totality-asserting behaviour is mostly confined to dogmatic metaphysics. In line with this, there are likely more observed instances of totality-seeking than totality-asserting behaviour. This implies not only that the hypothesis of Reason has a higher degree of unity than that of Unreason, $u(H_R) > u(H_U)$, but also that Reason likely has a broader evidential base, meaning that $|\{p|p \in E_s \wedge H_R \vdash_e p\}| > |\{p|p \in E_a \wedge H_U \vdash_e p\}|$.

Drawing on the probability function outlined in Section 5.5, we can conclude that Reason is more probable given the evidence of our totality-seeking behaviour than Unreason is given the evidence of our totality-asserting behaviour – $P(H_R, E_s) > P(H_U, E_a)$. This difference is decisive if – as we've assumed – H_S has the same probability in both scenarios, and if both H_R

and H_U can explain the total evidence when combined with H_S. Thus, the joint probability of H_R and H_S given the total evidence surpasses that of H_U and H_S. In other words, $P(H_R \wedge H_S, E_a \cup Es) > P(H_U \wedge H_S, E_a \cup Es)$.

In light of Kant's own theory of probability, we find that the hypothesis that we possess Reason – a power that seeks the totality of conditions – is more probable than both the hypothesis that we have a moderate faculty of inference, which seeks only some conditions, and the hypothesis that we have Unreason, which asserts totalities without seeking them. While there may be further hypotheses to consider,[303] I hope to have shown that Kant has the resources to resolve the Problem of Perfection. For even though we cannot directly observe perfect reasoning in inner sense, we can still infer Reason as the most probable explanation for our imperfect reasoning.

In a sense, then, empirical psychology is more like chemistry than physics. Just as chemists rarely find pure elements, but rather discover them mixed with other substances, so psychologists do not observe perfect reasoning, but rather its imperfect shadow, corrupted by sensibility. But this does not dissuade either the chemist or the psychologist – chemists hypothesise pure elements to explain the observed mixtures, just as psychologists hypothesise ideal faculties to explain flawed reasoning.[304] Indeed, Kant himself alludes to this parallel when he introduces the following example from chemistry just two pages before the Primary Passage:

> One admits that *pure earth*, *pure water*, *pure air*, etc. are difficult to find. Nevertheless, one still needs the concepts of these (which, as far as complete purity is concerned, only have their origin in reason) in order to properly determine the share that each of these natural causes has in the appearance, and thus one brings all materials to the earth [...], salts and combustible substances [...], and

[303] For example, Andrew Chignell suggested that we might consider a hypothesis according to which we lack a faculty that seeks the totality of conditions, and instead have only a faculty that seeks additional conditions *indefinitely*. Although this faculty would probably be weaker than reason – as defined in (A307/B364) –, it could still account for our inner observations.

[304] On the parallel between empirical psychology and chemistry, see Kant's discussion of 'inner experiments' (7:98). Contrary to Kitcher (1990: 39-40), I argue that Kant's method of isolation extends beyond transcendental philosophy. For a comprehensive discussion of Kant's theory of chemistry, consult McNulty (2014), (2017).

> finally water and air as vehicles [...] , in order to explain the chemical effects of the materials on each other according to the idea of a mechanism. (A646/B674)

In light of this passage, I propose that empirical psychology isolates pure Reason in the same manner that chemistry isolates pure elements – such as pure earth, pure water, or pure air –, as a faculty untainted by other influences, specifically unaffected by sensibility. In essence, then, we abductively infer from our flawed human reasoning to an idealised notion of pure Reason. However, Kant also defines 'pure reason' as that faculty 'which contains the principles of cognising something a priori' (A11/B24). This suggests that pure Reason might only be known a priori, not through empirical psychology, which now leads me to the Problem of Aprioricity.

6.3 From Hypothesis to Theory

A brief reminder: what is the Problem of Aprioricity? The problem is that we cannot come to know that we have Reason through empirical psychology, because the Principle of Reason must surely be necessary, and so a priori. But we cannot come to know a priori principles on the basis of inner sense. All we can learn from inner experience are contingent empirical generalisations, or as Martens puts it: 'from mere empirical observation of how we do think, we could only ever get contingent natural laws' (ms).[305] The goal in this section is to resolve the Problem of Aprioricity.

Underlying the Problem of Aprioricity is a more general worry about the modality of empirical laws. The worry starts from the observation that, on Kant's account, all *and only* a priori cognition is necessary and strictly universal. In the B Introduction, Kant explains that 'a proposition that is thought to be necessary [...] is a judgement a priori', and moreover that 'a judgement thought in strict universality [...] is not derived from experience, but is valid a priori' (B3-4). Empirical cognition, by contrast, is contingent and

[305] Or in Kitcher's words: 'psychologism would make the laws of logic merely probable and contingent, rather than necessary' (1990: 8).

has only 'comparative universality (*via induction*), so that it should actually be said: as far as we have been able to perceive, there is no exception to this or that rule' (ibid.; my emphasis).

But this raises a key question: how are empirical laws of nature even possible for Kant?[306] In the *Prolegomena*, Kant suggests that empirical 'laws of nature' qua being *empirical*, can be 'known by means of experience' (4:318-9). Yet at the same time, Kant also maintains that, 'in nature, law-governedness rests on the necessary connections of appearances in experience' (4:320), and so 'particular causal laws', qua *laws*, must be 'necessary and strictly universal' (Friedman 1992b: 165). But if empirical laws are necessary, they would need to be a priori. We can articulate this 'Modal Worry' as an inconsistent triad:

Modal Worry

(1) What is cognised via experience, is merely contingent.
(2) Empirical laws, qua being empirical, can be cognised via experience.
(3) Empirical laws, qua being laws, are necessary.

If we take (3) as set in stone, there are two main strategies for addressing the Modal Worry. The first one, which has become increasingly popular, is to resolve the worry by challenging (1). The main idea is to distinguish between the metaphysical or nomic necessity of laws and the logical or epistemic necessity entailed by a priori cognition.[307] Metaphysical necessity is grounded directly in causal powers (Watkins 2005: 224),[308] dispositions (Engelhard 2018: 8),[309] or essences (Stang 2016: 229). It is, then, suggested that we can come to know or cognise this metaphysical necessity through experience (Cooper 2023),[310] implying that empirical cognition need not

[306] This question is not to be read in a sceptical way, because Kant clearly thinks *that* there are particular empirical laws of nature. For example, see (A766/B794), (4:318), (4:468), (4:534), (5:180-1), and (20:203-5).

[307] Recent scholarship has increasingly sought to disentangle Kant's concept of the a priori from his notion of necessity. Stang, for instance, contends that Kant remained 'agnostic as to whether necessity and a priority are co-extensive' (2011: 443). Stephenson (forthcoming) develops a similar perspective.

[308] See also Kreines (2009) and Messina (2017).

[309] See also Massimi (2017).

[310] Kreines (2009: 536) and Messina (2017: 138) contend that separating metaphysical

always be contingent.[311]

The second, more traditional, strategy – pioneered by Friedmann (1992a, 1992b) – solves the Modal Worry by rejecting (2). While empirical laws can be cognised through experience, some of them can *also* be cognised a priori. That's because 'particular laws of nature are somehow grounded in or made possible by the transcendental principles of the understanding', which is why they 'must have a more than inductive status' (1992b: 172). Indeed, Kant acknowledges that some a priori cognition can also be learned from experience (B2-3). On Friedman's account, then, the necessity of empirical laws of nature is grounded in a priori principles, from which they can be derived.

Both strategies effectively alleviate the tension at the heart of the Problem of Aprioricity. If necessary laws of nature can be known through experience, the need for the Principle of Reason to be a priori diminishes – the principle can be necessary even if we come to know it via inner sense.[312] Conversely, if the Principle of Reason can be derived a priori, there is no problem either. Inner sense provides empirical access to the principle, without precluding an additional a priori access. While I remain neutral on which strategy best grounds the necessity of causal laws, I aim to show that Kant has the resources to claim – and for some time *did* claim – that we can *also* derive the Principle of Reason a priori.

To demonstrate this, I will examine Kant's often-overlooked account of

---

necessity from epistemic necessity introduces raising doubts about how we could know the laws of nature. Yet these doubts have been defused by Engelhard (2018: 9), Massimi (2017: 169), Breitenbach (2018: 114), and Cooper & Jones (2023). Moreover, much of what I have outlined in Chapters 4 and 5 provides a robust framework for understanding how we can come to know empirical laws, even if they cannot be derived a priori.

[311] Kitcher (1986: 209) asserts that specific causal laws derive their necessity from being part of a uniquely optimal system of laws. Although Kitcher likewise maintains that the necessity of laws is independent of their a priori status, Messina (2017: 136), Engelhard (2018: 30), and Jones & Cooper (2023: 340) have convincingly challenged this view, arguing that the mechanism by which the system imparts necessity to the laws remains unclear.

[312] This would sit well with Kant's claim in the second *Critique* – which I had already discussed at the end of Section 2.2 – that 'all human insight comes to an end as soon as we reach the basic powers or basic faculties; for their possibility cannot be understood by anything, but equally it cannot be arbitrarily invented and assumed. Therefore, in the theoretical use of reason, *only experience can entitle us to assume them*' (5:47; my emphasis).

hypothesis confirmation in detail.[313] We have already explored two methods by which Kant suggests a hypothesis might be confirmed. In Section 5.2, I noted that a hypothesis about a law would be confirmed if all its possible consequences were given in experience – which, however, is impossible since there are infinitely many such consequences. Additionally, in Section 5.5, I proposed that the hypothesis about a law could also be confirmed if said law is derived as a species from a genus. While the species law doesn't become certain in this way, it still inherits the higher probability of its genus.

Both of these are ways to confirm a hypothesis empirically. In a hand-full of passages, however, Kant indicates that hypotheses can also be confirmed a priori. For example, he writes: 'If, in addition to the *a posteriori proofs*, there are also *a priori grounds* for a hypothesis, then it is certain. It is a main duty in natural science that one also proves a priori what one assumes' (24:440; my emphasis).[314] So not only are we able to give *additional* a priori grounds for the hypothesis that was only empirically justified, but doing so is a scientific duty. Once confirmed a priori, the hypothesis 'ceases to be a hypothesis' – i.e., something that is abductively justified – and instead becomes a known 'theory' (24:220-2).[315]

The mechanics of this a priori confirmation are murky, and my aim is not to clarify them. But it is generally thought that a hypothesis is confirmed a priori when it is derived from transcendental principles of the understanding (excluding Reason). As Friedman (1992a, 1992b, 2013) points out, a prime example is Kant's a priori validation of Newton's Laws in the *Metaphysical Foundations* (1786).[316] Another example may be the Copernican system. As late as the 1770s, Kant considered this system to be 'a hypothesis' (24:221), but one that could somehow be deduced from Newton's Laws. Following the

[313] Though see Cooper (2023), (ms), as well as my Benzenberg (ms-d).

[314] Kant repeats the same point in this Reflexion: 'If the same thing that is shown by its consequences can also be proved independently a priori, then the hypothesis is confirmed' (16:465, R2678). See also (24:220-1).

[315] In the *New Tone* essay, Kant asserts that theories are certain (8:402), with (logical) certainty being characteristic of the grounds of knowledge. Further discussion of the type of assent typical of hypotheses can be found in Section 6.4.

[316] In addition to from the *Metaphysical Foundations*, Friedman (1992b) also points to (A159/B198), (A216/B263), as well as (R5414 18:176).

confirmation of these laws in 1786, Kant proudly proclaims in the B Preface of 1787 that Newton's Laws 'provided what Copernicus at first assumed to be only a hypothesis with *outright* [*ausgemachte*] *certainty*' (Bxxiin; my emphasis).

While Friedman (1992a, 1992b, 2013) limits his account to the a priori derivation of physical laws of outer nature, I suggest that it extends to psychological laws of inner nature. As a first piece of textual evidence consider Kant's taxonomy of 'physiology' in *Metaphysics* L_1 from the late 1770s. After contrasting 'physics' and 'psychology' as the physiologies 'of the objects of outer [...] and of inner sense' (28:222), Kant divides psychology into empirical psychology, as 'the cognition of the objects of inner sense insofar as it is obtained from experience', and rational psychology, as 'the cognition of the objects of inner sense insofar as it is borrowed from pure reason' (ibid.). He then does the same for rational and empirical physics (28:223). Figure 18 depicts the taxonomy.[317]

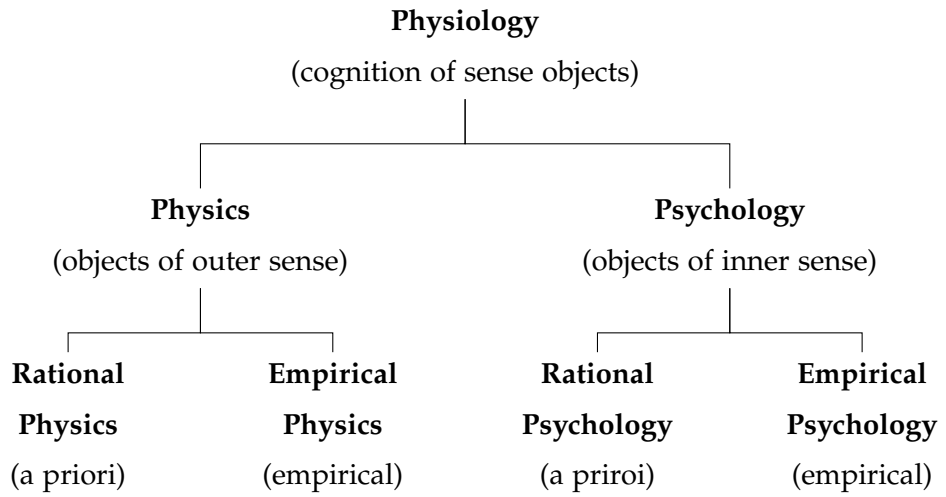

Figure 18: taxonomy of physiology

[317] For a further discussion of this taxonomy, see Proops (2021: 73). The 'immanent physiology' Kant discusses in the Architectonic chapter of the first *Critique*, is only a 'physiology of *pure reason*' and so excludes empirical physics and empirical psychology (A845-6/B873-4; my emphasis)

On Kant's account, empirical psychology stands to rational psychology as empirical physics stands to rational physics. In the Architectonic chapter, Kant further explains the parallel between rational physics and psychology: 'The metaphysics of corporeal nature is called [...] rational physics. The metaphysics of thinking nature is called [...] rational [psychology]' (A846/B875). And both disciplines 'take from experience nothing more than what is necessary to give us an object, either of the outer or of the inner sense. The former is done by the *mere concept of matter* (impenetrable lifeless extension), the latter by *the concept of a thinking being* (in the empirical inner representation: I think).' (A848/B876; my emphasis).

This passage is remarkable when viewed in light of Kant's taxonomy of physiology. After all, the 'mere concept of matter' is the linchpin of Kant's a priori derivation of Newton's Laws in the *Metaphysical Foundations* (Friedman 1992a, 2013). So it is in rational physics that we can a priori confirm hypotheses formulated in empirical physics. In the same way, then, we should expect that the 'concept of a thinking being' enables rational psychology to a priori derive all sorts of psychological laws, including the Principle of Reason; and more generally, to a priori confirm the hypotheses formulated in empirical psychology.

Indeed, there is explicit evidence supporting this expectation. As Gava (2018: 4n) points out, Kant references projects for both a '*metaphysical foundations* of the doctrine of bodies as well as [...] that *of the doctrine of the soul*' in a 1785 letter to Gottfried Schütz (10:406; my emphasis). This indicates that, up until a year before publishing the *Metaphysical Foundations*, Kant believed in the necessity of a Critically-enlightened rational psychology – distinct from the rational psychology critiqued in the Paralogisms[318] – that would allow us to a priori confirm the hypotheses formulated by empirical psychology about the mind, including, I suggest, the probable hypothesis that we possess Reason.

[318] After all, the Critically-enlightened rational psychology is part of 'immanent physiology', which 'refers to nature, insofar as its cognition can be applied in experience (*in concreto*)' (A873/B845).

Moreover, this intricate interplay between rational and empirical psychology has a prominent predecessor in Wolff. Without being able to reconstruct the nuances of Wolff's psychology here, Dyck sums it up well when he states that 'Wolffian rational psychology relies upon empirical psychology. [...] Wolff even compares empirical and rational psychology to empirical and theoretical parts of astronomy' (2014: 13).[319] Dyck (2014: chs. 1-2) further demonstrates that this dual model of psychology became the standard for post-Wolffian thinkers such as Baumgarten, Crusius, Lambert, Meier, Tetens, and notably Kant.[320]

In the *Metaphysical Foundations*, Kant famously revises his stance. He argues that there can be no a priori foundation for empirical psychology, which he now groups with chemistry, because 'mathematics is not applicable to the phenomena of the inner sense and its laws' (4:471). Kant's idea is that a science can only have an a priori foundation if its concepts can be constructed in intuition, which is possible only in mathematics. However, mathematics requires the three-dimensional form of space; yet the objects of psychology exist solely in the one-dimensional form of time (4:470-1).[321] The specifics aren't crucial; what's key is that by 1786, Kant no longer believes that rational psychology can a priori confirm the hypothesis of empirical psychology.

But this demise of rational psychology need not undermine the general solution I've offered to the Problem of Aprioricity. For even if rational psychology cannot confirm a priori that we have Reason, this confirmation might still be achievable by other means. As noted in Sections 1.2 and 1.3, my Main Claim is that empirical psychology is *one* way of knowing that we have Reason; not that it is the *only* way. I remain open to the possibility that a future reading may answer the Epistemic Question on synthetic a priori grounds. Such a reading, in turn, would provide an a priori confirmation

[319] For a further discussion of the relation between empirical and rational psychology in Wolff, See also Richards (1980), Arndt (1983), Kitcher (1990: 12), Heßbrüggen-Walter (2004: 84), and Rudolph (2004).

[320] In Dyck's words: 'Kant's own rational psychology (Critical and pre-Critical) remains broadly continuous with that of his immediate predecessors and in fact even serves to vindicate a variety of claims' (2014: 237).

[321] For a rigorous discussion of this argument, see Friedman (1992a), (1992b: 189), (2013), and Kraus (2016).

of the hypothesis that we have Reason – a hypothesis that was originally formulated in empirical psychology.

So even if we cannot confirm the hypotheses of *empirical* psychology through *rational* psychology, we might still be able to do so through *transcendental* psychology. In this case, I would concur with Kitcher that there exists a 'commonality of subject matter' between 'transcendental and empirical psychology' (1990: 25).[322] But either way, Kant must think that we can a priori confirm the empirical hypotheses that we have Reason. This addresses the Problem of Aprioricity: just as Kant was able to a priori confirm the empirical hypotheses of Newton and Copernicus, so too can he a priori confirm the Principle of Reason, even if it was initially advanced as hypothesis based on inner sense.[323]

6.4 Remaining Reservations

Having addressed the four primary problems, I now turn to some remaining reservations about my interpretation. A major reason, I believe, why Kant scholarship has historically undervalued the epistemic utility of empirical psychology is that it has overvalued Kant's own concerns with empirical psychology, as expressed in the *Metaphysical Foundations* (4:471) and the *Anthropology* (7:133, 7:161).[324] In this section, I will examine these concerns and argue that, while they pose significant challenges to empirical psychology, they do not constitute insurmountable obstacles; in particular, the concerns do not prevent empirical psychology from being a powerful metacritical tool for Kant.

[322] However, I believe Kitcher misconstrues the direction of dependence. She argues that empirical psychology relies on transcendental psychology for its subject matter, claiming that '[t]ranscendental psychology offers an idealization of cognitive functioning; as such, it can provide guidance about the sorts of mental equipment that empirical research needs to look for' (1990: 206). However, as I will argue in Section 6.5, the dependence is precisely the opposite: empirical psychology proposes potential mental laws, which transcendental psychology then seeks to confirm.

[323] Indeed, in the B Preface, Kant parallels his transcendental psychology with his a priori confirmation of the Copernican hypothesis (Bxxiin).

[324] For a comprehensive overview and discussion of Kant's concerns regarding empirical psychology in the *Metaphysical Foundations* and *Anthropology*, see Sturm (2001), (2009).

The first of these concerns is that empirical psychology cannot be a science, strictly speaking. As previously noted, Kant argues in the *Metaphysical Foundations* that mathematics is not applicable to inner sense. Yet without mathematics, he argued there can be no true science: '[the] empirical doctrine of the soul must always remain removed from the rank of a natural science in the proper sense, [...] because mathematics is not applicable to the phenomena of the inner sense and their laws' (4:471). Kraus notes that this argument is often (mis)interpreted as saying that science requires quantification, yet without mathematics, 'Kant's theory does not allow for quantification in psychology' (2016: 332).[325]

This demotion from the status of a proper science has cast empirical psychology in a negative light, leading many to doubt its epistemic authority. Kraus observes that 'Kant's apparent denial of proper scientific status to empirical psychology in the *Metaphyiscal Foundations of Natural Science* [...] has led many commentators to think that Kant must reject the very possibility of any theoretical knowledge of psychological phenomena and hence of inner experience as empirical cognition' (2020: 6). Indeed, a long list of scholars remain sceptical about the epistemic prospects of empirical psychology, given its exclusion from the ranks of proper science.[326]

This scepticism about empirical psychology can be countered in two ways. First there are interpreters that push back against, or limit the scope, of Kant's argument that empirical psychology cannot be a science. For example, Sturm suggests that 'Kant does not claim that psychological phenomena cannot be mathematized at all' but only that, 'a specific conception of empirical psychology – one shared by many of his contemporaries – [...] cannot be mathematized appropriately' (2001: 164). In the same spirit, Nayak & Sotnak (1995: 148) suggest that some version of the law of continuity in the flux of

[325] Brook summarises the argument as follows: '[Kant] tells us that science proper must be apodictic, therefore in part a priori, therefore mathematical, because expressions in a mathematical model is the only way for a science to achieve necessity and certainty' (1994: 9). For a further discussion, of the argument see Schönrich (1991), Friedman (1992b: 189), Nayak & Sotnak (1995), Sturm (2001), and Kraus (2016).

[326] On this point, see Mischel (1967), Gouaux (1972), Washburn (1976), Leary (1978), Schönrich (1991), Hudson (1994), Klemme (1996), Makkreel (2003), and Friedman (2013), Dyck (2014: 222), and Gomes (2017: 15).

inner changes can be mathematised and does apply to inner sense, after all.

But if we accept that Kant's rejection of empirical psychology as a proper science holds without restriction, we can still ask: so what? As we've already seen, empirical psychology behaves much like a science: it is grounded in empirical observation; it adheres to rigorous standards of probabilistic reasoning;[327] it seeks to explain inner phenomena through causal laws; it strives to unify these laws systematically; and its hypotheses can be confirmed a priori. For all intents and purposes, empirical psychology behaves like a science, which is also why Kant calls it a 'science in academia' (25:472).[328] The label of 'proper science' then turns out to be a mere honourific, and its absence does not imply a deep epistemic deficiency.[329]

That being said, one might worry that without being a science (*Wissenschaft*), in the strict sense, empirical psychology also doesn't produce knowledge (*Wissen*), in the strict sense. As Brook notes, 'Kant thought that the products of empirical psychology fell short of being genuine knowledge' (Brook 1994: 4). This worry has a textual basis. For as we have seen, empirical psychology hypothesises laws of the mind – including the Principle of Reason. Yet hypotheses for Kant aren't knowledge, but a species of opinion (*Meinung*): 'Hypotheses are opinion' (24:733)[330] – a textual fact that has been widely recognised in the literature.[331]

And indeed, while Kant defines knowledge as assent based on objective grounds that are sufficient (see Section 2.4), opinion is assent that rests on

[327] While mental powers themselves may resist direct quantification in empirical psychology, the probabilities of these powers do not. For as I demonstrated in Chapter 5, Kant utilises a rigorous probability function that allows him to accurately determine the probability of a hypothesis given the evidence.

[328] See also (5:219) as well as Frierson (2014: 27) for a discussion.

[329] After all, Kant also claimed that chemistry could 'never become a proper science' (4:471), yet no one would argue that chemistry lacks epistemic standing.

[330] For the claim that all hypothesis is opinion, see (A769-70/B797-8), (A774-5/B802-3), (5:142), (R2678, 16:465), (R2679, 16:465-6), (R2682, 16:469) (R2690, 16:471), (R2693, 16:472) (24:220), (24:230), (24:541), (24:733), (24:746-7), (28:596-7), and (28:1291). For the (implied) claim that some opinion is hypothesis, see (5:470), (8:311-2), (10:311), and (24:541). And for the claim that not all opinion is hypothesis, see (24:220-2), (24:439-40), (24:466), and (24:541).

[331] Butts (1962: 189, 195) notes that all scientific hypotheses are opinion. Chignell suggests that working hypotheses are opinion (2007a: 332; see also 2007b: 35-7). Pasternack (2014b: 67) identifies hypotheses with philosophical opinions. Sturm (2015: 1059-60) notes that hypotheses are opinion and not knowledge. See also Willaschek (2010: 184).

objective grounds that are insufficient: 'Opinion or the assent from a ground of cognition that is [...] [not] objectively sufficient' (9:66).[332] As I suggested in Section 5.1, the sufficient objective ground of a hypothesis is the set of all its possible consequences – which is why a hypothesis would be known if all its consequences were given in experience. But since that's impossible, we always have only part of the sufficient objective ground, which is insufficient. That's why, for Kant, all hypotheses, including those of empirical psychology, are fallible opinions.[333]

While this interpretation is textually accurate, it is crucial to clarify that opinion for Kant isn't epistemically irrational. Some authors have mistakenly claimed that opinion equates to unjustified assent, even likening it to persuasion (*Überredung*).[334] But they are wrong. Kant explicitly states that opinion, like knowledge, rests on *some* objective grounds: 'The assent from objective grounds is either opinion or knowledge (insufficient and sufficient)' (R2488, 16:391).[335] Stevenson aptly captures this distinction, noting that 'Kant sees meinen and wissen as different grades on the same scale of theoretical justification' (2003: 86).[336]

Indeed, as more and more consequences of a hypothesis are confirmed through experience, its probability increases, approaching an asymptotic 'approximation towards certainty' (9:84) – a point Kant reiterates in several

[332] See also (A822/B850), (R2450, 16:374), (R2458, 16:378), (R2473, 16:385), (R2477, 16:387), (R2486, 16:389), (R2488, 16:391), (R2492, 16:392), and (24:440, 541, 637, 732, 850, 853). At times, Kant also defines opinion as problematic judgement: 'Opinion is a problematic [...] judgement' (9:66). See also (R2474, 16:385) and (24:541, 732). For a further discussion, see Mattey (1986: 432-9), van der Schaar (2003: 6, 8), Höwing (2016: 207), (2017: 116-7), and Insole (2019: 517).

[333] The fallibility of empirical psychology is also mentioned by Frierson (2014: 50).

[334] Stevenson turns persuasion into a species of opinion (2003: 80-2). On the same point, see also Butts (1962: 155), Mattey (1986: 433), and van der Schaar (2003: 9).

[335] Elsewhere Kant remarks that 'opinion is compared with knowledge' (24:422), and that by adding objective grounds, 'opinion can eventually become knowledge' (8:141). He even calls opinion 'partial knowledge' (8:396n). See also (R2451, 16:374), (R2487, 16:390), (24:219), (24:421), (24:638-9), and (24:850-1). Kant also contrasts opinion with persuasion. See (9:73) and (24:218, 439). The contrast between opinion and persuasion is discussed by Heimsoeth (1971: 778) and Gava (2023).

[336] Insole states, 'if more evidence becomes available, the state of opining can become the state of knowing' (2019: 517). See also Chignell (2007a: 332), Spoerhase (2009: 269), Trullinger (2013: 381), Pasternack (2014a: 58, 61, 67-8), (2014b: 48n, 50), (2014c: 82), Kitcher (2015: 2668-9), and Thielke (2015: 1502).

places.[337] While opinion never achieves a probability of 1, it can come arbitrarily close. And as we edge nearer to full certainty, Kant notes in the *Vienna Logic* that hypotheses can reach an 'analogue of certainty', where one can 'surrender to it, as to complete certainty' (24:887).[338] Thus, even if hypotheses do not meet the infallible standards of knowledge,[339] they can become practically indistinguishable from complete certainty.[340]

What is more, even hypotheses that have not yet become an analogue of complete certainty still possess what Kant refers to as empirical certainty: 'A hypothesis [...] never attains complete certainty [...], empirical [certainty] it may well attain' (24:746).[341] They also come with hypothetical certainty – 'Hypothetical [certainty] [...] is opinion' (28:597)[342] – as well as comparative certainty, which Kante defines as follows: 'When the grounds of approval surpass all the grounds of opposite, this is then a comparative certainty' (24:220-1).[343] All these types of certainty underscore that hypotheses, including those in empirical psychology, rest on a firm epistemic foundation, capable of reaching near-infallible justification.

Kant makes two more disparaging remarks about empirical psychology that have resonated widely in the secondary literature. The first, found in the *Metaphysical Foundations*, is his assertion that empirical psychology is built on epistemic quicksand, as the very act of observing our inner states inevitably alters them: 'Even the observation itself alters and distorts the state of the

---

[337] See (A647/B675), (9:81-2), (21:61), and (24:886-9). Kant notes that all approximation (*Annäherung*) is asymptotic (*asmptotisch*) (A663/B692).

[338] Likewise, in the *Jäsche Logic*: 'the probability of a hypothesis can grow and become an analogue of certainty' (9:85). See also Butts (1962: 196), Demarest & van den Berg (2022: 12), Cooper (ms), and my Benzenberg (ms-d).

[339] Note that fallibilist readings, like those of Chignell (2007a, 2007b, 2021), can even go so far to say that high-probability hypotheses cease to be opinion and become knowledge. For a rejection of these readings, see my Benzenberg (ms-d).

[340] As Kohl puts it: 'Thus, some well-founded empirical judgments – e.g. medical or chemical hypotheses – can be treated as if they did yield knowledge, even though our assents here are fallible and lack necessity' (forthcoming).

[341] This conclusion independently follows from Kant's claims about induction. First, Kant claims that hypotheses are inferred by induction. See (A646-7/B674-5), (9:84-5), (24:558), as well as Madonna (1993: 40), Vanzo (2012: 82), Proops (2021: 53), and my Benzenberg (ms-d). Second, Kant claims that '[i]nduction and analogy [...] give empirical certainty' (9:133; see also 24:594, 679). Therefore, hypotheses can gain empirical certainty.

[342] On hypothetical certainty, see also (28:596-70, 605) and my Benzenberg (ms-d).

[343] On comparative certainty, see (24:225), (24:888), Benzenberg (ms-d), and Cooper (ms).

object observed' (4:471). When we focus too intently on our inner lives, those states can shift and slip through our grasp. As Kraus aptly puts it, Kant 'warns that in paying too close attention to oneself, one may involuntarily change the course of one's thoughts and feelings' (2020: 244).[344]

This remark introduces challenges, none of which, however, are fatal to empirical psychology. For even if introspection locally altered some inner states, such alteration cannot be global; otherwise, we wouldn't be able to tell that the inner states were altered in the first place; we can only tell by comparing the altered stated to the original. Furthermore, while we may have some limited control over how we carve up inner states, this control is not without limits. For example, mere introspection cannot convert a cognition of p into a desire for p.[345] It would seem that we can sidestep these issues by not excessively fixating on our inner states for long periods of time.

The second remark is taken from the *Anthropology*. There, Kant makes the following claim: 'The soul is the organ of the inner sense, which is now said to be subject to deception, in that humans take the appearances of the soul for external appearances, i.e., imaginations for sensations' (7:161). As a consequence, Kant also argues that excessive introspection in inner sense can lead to 'head confusions [*Kopfverwirrung*]' (7:133). So not only can introspection alter the observed state, but it can even cognitively damage the observer. Empirical psychology thus seems to pose great a hazard.

Two points in response. First, while some individuals might struggle to differentiate between object-based sensations and subject-based imagination – possibly even due to excessive self-observation – this is not the typical case. There may also be astronomers who have damaged their eyesight by staring at the sun; but such cases don't discredit the field of astronomy. Similarly, just because some people confused their heads by prolonged self-observation, this doesn't invalidate empirical psychology; the claims of empirical psychology

[344] Kitcher concludes from this that empirical '[p]sychology cannot really be a good experimental science either, because [...] observation through inner sense or introspection alters the state of the object observed' (1990: 11). See also Friedman (1992b).

[345] This goes against Brook, who argues that 'the introspective observer distinguishes the items from one another; there are no real distinctions among them' (1994: 10; see also 27).

remain justified when made by mentally healthy individuals. The principle is simple: avoid excessive introspection, just as one should avoid staring at the sun for too long.

Second, some interpret Kant's remarks in the *Anthropology* as not just an epistemic but an ethico-practical objection: empirical psychology is unhealthy. Kitcher, for instance, notes that Kant 'criticizes introspection as unreliable, unstable, unnatural, and a potential *route to lunacy*!' (1990: 6; my emphasis). Brook echoes this, stating that 'introspection is even bad for the health' (1994: 10).[346] But as previously noted, empirical psychology poses a risk to mental health only when taken to extremes. In moderation, it can actually enhance our mental well-being. As I argued in Section 3.5, empirical psychology can diagnose the madness that drives speculative metaphysics, potentially serving as a *medicina mentis*.

Where does this leave us? While it is true that Kant is occasionally dismissive of empirical psychology, contributing to its poor reputation in Kantian scholarship, a closer examination reveals that his concerns do not render empirical psychology impossible as an epistemic endeavour. Empirical psychology is at most 'very difficult' (7:121), a point also acknowledged by Frierson (2014: 37-38). The challenges facing empirical psychology can either be overcome or are less severe than they appear. Therefore, I propose that empirical psychology can serve as a robust metacritical tool, enabling us to establish the existence of Reason. In the final section, I will highlight the distinct metacritical benefits of empirical psychology.

6.5 Metacritical Scaffold

In this thesis, I have argued that, according to Kant, we can come to know that we possess Reason through empirical psychology. Specifically, I have suggested that the hypothesis of Reason – a faculty governed by the Principle

[346] Kant raises similar objections to empirical psychology in (9:41-3), (7:132-3), (25:252), (25:477-8), (25:863), and (25:868). For an in-depth analysis of the health risks associated with empirical psychology, refer to Frierson's discussion (2014: 40-43).

of Reason – is the most probable explanation for the (imperfect) explanation-seeking behaviour we can observe in inner sense. This use of empirical psychology as a metacritical resource marks a significant departure from the longstanding tradition in Kant scholarship, which has largely denied the epistemic, let alone metacritical, utility of empirical psychology.

In this concluding section, I will highlight the metacritical utility of empirical psychology by contrasting it with its a priori counterparts – for remember that the hypotheses of empirical psychology can also be confirmed a priori (see section 6.3). This coexistence of empirical and a priori approaches prompts a fundamental question: why should we rely on empirical psychology when Kant also holds that faculties, including Reason, can be derived a priori? What distinct advantages does empirical psychology offer?[347] To answer this question, I will first outline empirical psychology's general metacritical benefits, and then address those specific to Reason.

In December 1759, Kant received a letter from Johann Georg Haman, which could just as well have been written by Kant, for it nails the relation between a priori and empirical reasoning. The letter says: 'An axiom is preferable to a hypothesis; but the latter is not to be rejected; it must be used not as a *foundation* [*Grundstein*], but as a *scaffold* [*Gerüst*]' (10:26; my emphasis). So even if we can establish our faculties a priori as mental axioms, there is still an epistemic value in hypothesising them in empirical psychology – not as a foundation, but as a scaffold. But what does this mean? Let us unpack the nuances.

The first point to recognise is that a foundation cannot be built without first erecting the scaffold. This suggests that empirical psychology serves as a propaedeutic to any a priori endeavour – whether in rational or transcendental psychology. Even if the a priori enterprise is logically prior, empirical psychology historically precedes it. We always start by gauging a probable structure of the mind, hypothesising basic and derivative powers to explain the states we observe in inner sense. Only once these hypotheses are in place can we then try to also confirm or derive them a priori.

[347] Thanks to Yazmeen Martens for pushing me on this question.

Consider an analogy with physics: it would have been utterly bizarre if Kant had written the *Metaphysical Foundations* before Newton published his *Principia*. Only after Newton had hypothesised his laws did Kant have a clear object for his a priori derivation. Without any prior empirical support, Kant's derivation would have seemed not only highly unlikely, but completely unconvincing – much like Aristotle's attempt to justify a priori that women have fewer teeth than men. In the same way, then, empirical psychology introduces and empirically motivates the object for any a priori derivation.

Beyond serving as a propaedeutic, empirical psychology has the additional metacritical virtue of demystifying Kant's mental inventory. When we first encounter Kant's Critical philosophy, we find the mind already laden with numerous powers and faculties, many of which Kant presupposes without proper introduction. On the interpretation I have proposed, this need not be an immediate cause for concern: for although Kant's ambition must be to eventually justify these faculties a priori, he can initially presuppose them as empirically justified – the various faculties are simply probable explanations for making sense of our inner experience.

By the same token, empirical psychology has the added advantage of being able to introduce mental faculties while relying solely on the resources provided by Kant's first-order epistemology – which includes not only his account of empirical cognition, but also his theory of assent, and especially his theory of hypotheses. By contrast, most attempts to justify mental faculties a priori face the difficulty of having to introduce a special second-order epistemology. For example, act-awareness accounts typically claim that we can epistemically access our faculties and their laws without relying on intuitions or concepts (see Section 1.2).

Moreover, empirical psychology presents a metacritical perspective marked by epistemic humility – which may be seen as a virtue depending on one's philosophical inclinations. Kitcher, for example, worries that 'Kant's claims to certainty are considerably more problematic. Certain claims are immune from refutation. Beginning with assumptions about cognitive capacities, it may seem that he could not hope for certain results. [...] [C]ertainty seems

beyond his reach' (1990: 24). Empirical psychology lowers the bar, but in a way that may be more appealing precisely because it is, at least in principle, subject to revision, while also being, no doubt, more attainable than any a priori certainties about the mind.

To clarify, a scaffold in Kant's sense is not the same as Wittgenstein's ladder. While it is true that a priori grounds will eventually supersede any empirical justification by providing a more robust and infallible support, this does not invalidate the empirical grounds on which our initial understanding of faculties was based. Rather, once the a priori justification is put in place, the empirical grounds no longer carry the primary philosophical burden. In other words, the empirical scaffold can be left standing even after we have established the foundation of our mental faculties through a priori means.

It is important to recognise that empirical psychology as a metacritical tool may work better for some faculties than for others. In this thesis, I have argued that it works particularly well for Reason. For, as we have seen in Section 1.2, Reason poses unique metacritical challenges as it is the only faculty whose characteristic activity is infinite, and which therefore makes an infinite demand on us – to find the totality of L conditions for all conditioned cognition. Standard metacritical approaches falter here because they cannot easily establish an infinitely demanding faculty. For example, Reason cannot be grounded in apperceptive act of awareness, because its activity is never complete, and thus never fully present to us.

Empirical psychology, however, surprisingly overcomes these metacritical challenges. Kant's theory of probabilistic inference enables us to infer an infinitely demanding power from finite evidence. For indeed, every hypothesis about a law infers from what is directly given, particular states, to what is not directly given, a universal law. And so, although no infinite act is present in inner sense, we can still infer an infinite power – in large part because the hypothesis of an *all*-demanding power is more unified, and therefore more probable, than a *some*-demanding power. Remarkably, then, finite beings like us can come to know that we possess an infinitely demanding faculty through self-observation.

Finally, what makes my account distinctly Kantian is that, within empirical psychology, it is Reason that comes to hypothesise its own existence. For Reason is defined by its principle as the power that seeks complete explanations of all given phenomena, including phenomena in inner sense. In inner sense, we can observe our tendency to ask why-questions and to seek explanations for what is given to us. This tendency, Kant thinks, is best explained by hypothesising a power that seeks complete explanations, i.e., Reason. So, Reason is the power by which we come to know that we have the power of Reason.

Bibliography

Abela, Paul (2006): 'The Demands of Systematicity: Rational Judgement and the Structure of Nature', in: Graham Bird (ed.), *A Companion to Kant*, Oxford: Blackwell, 408-422.

Allison, Henry E. (2004): *Kant's Transcendental Idealism, An Interpretation and Defense*, New Haven: Yale University Press.

Ameriks, Karl (2000): *Kant's Theory of Mind, An Analysis of the Paralogisms of Pure Reason*, Oxford: Oxford University Press.

Anderson, Abraham (2013): 'The Dreams of a Spirit Seer and the Method of Hypotheses', in: Stefano Bacin et al. (eds.), *Kant und die Philosophie in weltbürgerlicher Absicht*, Berlin: de Gruyter, 423-428.

Anderson, R. Lanier (2005): 'Neo-Kantianism and the Roots of Anti-Psychologism', in *British Journal for the History of Philosophy*, 13, 287-323.

Anderson, R. Lanier (2015): *The Poverty of Conceptual Truth, Kant's Analytic/Synthetic Distinction and the Limits of Metaphysics*, Oxford: Oxford University Press.

Arias-Albisu, Martín (2017): 'The Methodological Prescriptions of the "Appendix to the Transcendental Dialectic" of Kant's Critique of Pure Reason and the Foundations of Improper Science', in *Studia Kantiana*, 15, 5-26.

Armstrong, David M. (1996): *A World of States of Affairs*, Cambridge: Cambridge University Press.

Arnauld, Antoine (1662/1854): *Logique de Port-Royal*, Paris: Librairie de L. Hachette.

Arndt, Hans W. (1983): 'Rationalismus und Empirismus in der Erkenntnislehre Christian Wolffs', in: Werner Schneiders (ed.), *Christian Wolff, 1679–1754*, Hamburg: Felix Meiner Verlag, 31-47.

Bacon, Francis (1620/2004): 'Novum Organon', in: Graham Rees & Maria Wakley (eds.), *The Oxford Francis Bacon*, vol. 11, Oxford: Oxford University Press, 48-448.

Bader, Ralf M. (2010): *The Transcendental Structure of the World*, PhD thesis, St Andrews: University of St Andrews.

Bader, Ralf M. (2017a): 'Inner Sense and Time', in: Anil Gomes & Andrew Stephenson (eds.), *Kant and the Philosophy of Mind*, *Perception, Reason, and the Self*, Oxford: Oxford University Press, 124-137.

Bader, Ralf M. (2017b): 'The Refutation of Idealism', in: James O'Shea (ed.), *Kant's Critique of Pure Reason*, *A Critical Guide*, Cambridge: Cambridge University Press, 205-222.

Bader, Ralf M. (2020): 'Kantian Meta-Ontology', in: Ricki Bliss & James T.M. Miller (eds.), *The Routledge Handbook of Metametaphysics*, Milton Park: Routledge, 23-31.

Baiasu, Sorin (2016): 'Constitutivism and Transcendental Practical Philosophy: How to Pull the Rabbit Out of the Hat', in *Philosophia*, 44, 1185-1208.

Banham, Gary (2013): 'Regulative Principles and Regulative Ideas', in: Stefano Bacin et al. (eds.), *Kant und die Philosophie in weltbürgerlicher Absicht*, Berlin: de Gruyter, 15-24.

Baumgarten, Alexander G. (1750): *Metaphysica*, Halle: Herman Hemmerde.

Bayes, Thomas (1763): 'An Essay Toward Solving a Problem in the Doctrine of Chances', in *Philosophical Transactions of the Royal Society of London*, 53, 370-418.

Bennett, Jonathan (1974): *Kant's Dialectic*, Cambridge: Cambridge University Press.

Benossi, Lisa (2021): 'In Defence of Inner Intuitions', in: Camilla Serck-Hanssen & Beatrix Himmelmann (eds.), *The Court of Reason*, Berlin: de Gruyter, 677-686.

Benzenberg, Christopher (forthcoming-a): 'Room for Responsibility: Kant on Direct Doxastic Voluntarism', in *Inquiry*.

Benzenberg, Christopher (forthcoming-b): 'Was Kant a Kantian about Doxastic States?', in: Paul Silva Jr. (ed.), *On Believing and Being Convinced*, Cambridge: Cambridge University Press, 51-61.

Benzenberg, Christopher (ms-a): 'Belief Without Bounds: Kant on the Supreme Principle of Reason' (*manuscript*).

Benzenberg, Christopher (ms-b): 'Conceptual Continuity: Kant's Third and Fourth Argument of Space Reconsidered' (*manuscript*).

Benzenberg, Christopher (ms-c): 'Epistemic Means: Kant's Notion of a Hypothesis in General' (*manuscript*).

Benzenberg, Christopher (ms-d): 'From Opinion to Knowledge: Why Kant Could Not Have Been a Fallibilist' (*manuscript*).

Benzenberg, Christopher (ms-e): 'How Many Possible Empirical Concepts Are There for Kant?' (*manuscript*).

Benzenberg, Christopher (ms-f): 'Out of this World: Kant's Belief in Extraterrestrial Life' (*manuscript*).

Benzenberg, Christopher (ms-g): 'Probably Immortal: Kant's Empirical Argument for the Afterlife' (*manuscript*).

Benzenberg, Christopher (ms-h): 'The Menace of Mysticism: Kant's Defence of Discursive Cognition' (*manuscript*).

Benzenberg, Christopher & Chignell, Andrew (2024): 'Ian Proops, The Fiery Test of Critique: A Reading of Kant's Dialectic', in *Philosophical Review*, 133, 197-202.

Beck, Lewis W. (1960): *A Commentary on Kant's Critique of Practical Reason*, Chicago: University of Chicago Press.

Bernoulli, Jacob (1713): *Ars conjectandi, opus posthumum*, *Accredit tractatus de seriebus infinitis, et epistola gallice scripta de ludo pilae reticularis*, Basel: Johann Rudolph & Emanuel Thurneisen.

Bird, Graham (2006): *The Revolutionary Kant*, *A Commentary on the Critique of Pure Reason*, Chicago: Open Court.

Boyle, Matthew (2020): 'Kant on Logic and the Laws of the Understanding', in: Sofia Miguens (ed.), *The Logical Alien*, *Conant and His Critics*, Cambridge, MA: Harvard University Press, 117-144.

Boyle, Robert (1662/1999): 'Defense against Linus", in: Michael C.W. Hunter & Edward B. Davis (eds.), *The Works of Robert Boyle*, vol. 3, London: Pickering and Chatto, 3-107.

Boyle, Robert (1666/2008): *The Text of Robert Boyle's 'Designe about Natural History'*, Michael C.W. Hunter & Peter Ansteay (eds.), London: Robert Boyle Project.

Breitenbach, Angela (2018): 'Laws and Ideal Unity', in: Walter R. Ott & Lydia Patton (eds.), *Laws of Nature*, Oxford: Oxford University Press, 108-122.

Breitenbach, Angela (2021): 'Regulative Principles and Kinds of the Unconditioned', in *Kantian Review*, 26, 287-297.

Brennan, Haley (ms): 'Kant on Consciousness of an Invisible Self' (*manuscript*).

Brink, Claudi (ms): 'The Distinction Between Absolute and Relative Spontaneity Reconsidered' (*manuscript*).

Brook, Andrew (1994): *Kant and the Mind*, Cambridge: Cambridge University Press.

Butts, Robert E. (1961): 'Hypothesis and Explanation in Kant's Philosophy of Science', in *Archiv für Geschichte der Philosophie*, 43, 153-170.

Butts, Robert E. (1962): 'Kant on Hypotheses in the "Doctrine of Method" and the Logik', in *Archiv für Geschichte der Philosophie*, 44, 185-203.

Butts, Robert (1997): 'Kant's Dialectic and the Logic of Illusion', in: Patricia A. Easton (ed.), *Logic and the Workings of the Mind*, *The Logic of Ideas and Faculty Psychology in Early Modern Philosophy*, Atascadero: Ridgeview, 307-317.

Caimi, Mario (1995): 'Über die wenig beachtete Deduktion der regulativen Ideen', in *Kant-Studien*, 86, 308-320.

Carl, Wolfgang (2014): *The First-Person Point of View, Berlin*, Berlin: de Gruyter.

Cassam, Quassim (1997): *Self and World*, Oxford: Oxford University Press.

Chaplin, Rosalind (ms): 'Is Kant's Supreme Principle of Pure Reason the Principle of Sufficient Reason?' (*manuscript*).

Chignell, Andrew (2007a): 'Kant's Concepts of Justification', in *Noûs* , 41, 33-63.

Chignell, Andrew (2007b): 'Belief in Kant', in *Philosophical Review*, 116, 323-360.

Chignell, Andrew (2009a): 'Kant, Modality, and the Most Real Being', in *Archiv für Geschichte der Philosophie*, 91, 157-192.

Chignell, Andrew (2009b): 'Real Repugnance and Belief about Things-in-Themselves: A Problem and Kant's Three Solutions', in: James Krueger & Benjamin J. Bruxvoort Lipscomb (eds.), *Kant's Moral Metaphysics*, *God, Freedom, and Immortality*, Berlin: de Gruyter, 177-209.

Chignell, Andrew (2010): 'Causal Refutations of Idealism', in *Philosophical Quarterly*, 60, 487-507.

Chignell, Andrew (2017): 'Can't Kant Cognize Himself? Or, a Problem for (Almost) Every Interpretation of the Refutation of Idealism', in: Anil Gomes & Andrew Stephenson (eds.), *Kant and the Philosophy of Mind*, *Perception, Reason, and the Self*, Oxford: Oxford University Press, 139-157.

Chignell, Andrew (2021): 'Kantian Fallibilism: Knowledge, Certainty, Doubt', in *Midwest Studies in Philosophy*, 45, 99-128.

Chignell, Andrew (2023): 'Kant's Panentheism: The Possibility Proof of 1763 and its fate in the critical period' in: Ina Goy (ed.), *Kant on Proofs of God's Existence*, Berlin: de Gruyter, 41-68.

Chulanon, Pirachula (2022): *Kant and the Origin of the Concept of Mind*, PhD thesis, Chicago: University of Chicago.

Cohen, Alix (2021): 'Kant on Epistemic Autonomy', in: Camilla Serck-Hanssen & Beatrix Himmelmann (eds.), *The Court of Reason*, Berlin: de Gruyter, 687-696.

Conant, James (1992): 'The search for Logically Alien Thought: Descartes, Kant, Frege, and the Tractatus', in *Philosophical Topics*, 20, 115-180.

Cooper, Andrew (2023): 'Hypotheses in Kant's Philosophy of Science', in *Studies in History and Philosophy of Science*, 99, 97-105.

Cooper, Andrew (ms): 'From Hypothesis to Theory: Kant on Experimental Reasoning' (*manuscript*).

Cooper, Andrew & Jones, Andrew (2023): 'Kant's Ongoing Relevance for Philosophy of Science', in *Kantian Review*, 28, 337-354.

Crucius, Christian A. (1745/1964): *Die philosophischen Hauptwerke*, *Entwurf der notwendigen Vernunft-Wahrheiten wiefern sie den zufälligen entgegen gesetzet werden*, vol 2., Giorgio Tonelli (ed.), Hildesheim: Olms.

Currie, Nicholas (2022): *A New Interpretation of Kant's Formal Logic*, PhD thesis, London: University College London.

de Freitas Araujo, Saulo & Faggion, Andrea (2015a): 'Hypothese, metaphysiche', in: Marcus Willaschek et al. (eds.), *Kant-Lexikon*, Berlin: de Gruyter, 1061-1062.

de Freitas Araujo, Saulo & Faggion, Andrea (2015b): 'Hypothese, transcendentale', in: Marcus Willaschek et al. (eds.), *Kant-Lexikon*, Berlin: de Gruyter, 1062-1064.

de Jong, Willem R. (1995): 'Kant's Analytic Judgments and the Traditional Theory of Concepts', in *Journal of the History of Philosophy*, 33, 613-641.

de Jong, Willem R. (2010): 'The Analytic-Synthetic Distinction and the Classical Model of Science: Kant, Bolzano and Frege', in *Synthese*, 174, 237-261.

Deleuze, Gilles (1983/2008): *Kant's Critical Philosophy*, *The Doctrine of the Faculties*, Hugh Tomlinson & Barabara Habberjam (trans.), London: Continuum.

Demarest, Boris & van den Berg, Hein (2022): 'Kant's Theory of Scientific Hypothesis in its Historical Context', in *Studies in History and Philosophy of Science*, 92, 12-19.

Descartes, René (1644): *Principia Philosophiae*, *Meditationes de prima philosophiae*, Amsterdam: Louis Elzevir.

Dicker, Georges (2008): 'Kant's Refutation of Idealism', in *Noûs*, 42, 80-108.

Dicker, Georges (2011): 'Kant's Refutation of Idealism: A Reply to Chignell', in *Philosophical Quarterly*, 61, 175-183.

Dicker, Georges (2012): 'Kant's Refutation of Idealism: Once More Unto the Breach', in *Kantian Review*, 17, 191-195.

Dyck, Corey W. (2014): *Kant and Rational Psychology*, Oxford: Oxford University Press.

Emundts, Dina (2006): 'Die Paralogismen und die Widerlegung des Idealismus in Kants "Kritik der reinen Vernunft"', in *Deutsche Zeitschrift für Philosophie*, 54, 295-309.

Engelhard, Kristina (2018): 'The Problem of Grounding Natural Modality in Kant's Account of Empirical Laws of Nature', in *Studies in History and Philosophy of Science*, 71, 24–34.

Englert, Alexander T. (2023): 'Kant's Favorite Argument for Our Immortality: The Teleological Argument', in *Res Philosophica*, 100, 357-388.

Falduto, Antonio (2014): *The Faculties of the Human Mind and the Case of Moral Feeling in Kant's Philosophy*, Berlin: de Gruyter.

Falkenburg, Brigitte (2000): *Kants Kosmologie*, *Die wissenschaftliche Revolution der Naturphilosophie im 18. Jahrhundert*, Frankfurt a.M.: Vittorio Klostermann.

Ferrarin, Alfredo (2015): *The Powers of Pure Reason*, *Kant and the Idea of Cosmic Philosophy*, Chicago: The University of Chicago Press.

Frankena, William K. (1958): 'Obligation and Motivation in Recent Moral Philosophy', in: Abraham I. Melden (ed.), *Essays in Moral Philosophy*, Seattle: University of Washington Press, 40-81.

Frege, Gottlob (1879/2020): 'Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens', in: Ignacio Angelelli (ed.), *Begriffsschrift und andere Aufsätze*, Hildesheim: Georg Olms Verlag.

Friedman, Michael (1992a): 'Causal Laws and the Foundations of Natural Science, in: Paul Guyer (ed.), *The Cambridge Companion to Kant*, Cambridge: Cambridge University Press, 161-199.

Friedman, Michael (1992b): *Kant and the Exact Sciences*, Cambridge, MA.: Harvard University Press.

Friedman, Michael (2013): *Kant's Construction of Nature*, *A Reading of the Metaphysical Foundations of Natural Science*, Cambridge: Cambridge University Press.

Frierson, Patrick R. (2014): *Kant's Empirical Psychology*, Cambridge: Cambridge University Press.

Funaki, Shuku (2002): *Kants Unterscheidung zwischen Scheinbarkeit und Wahrscheinlichkeit, Ihre historischen Vorlagen und ihre allmähliche Entwicklung*, Lausanne: Peter Lang.

Gava, Gabriele (2018): 'Sind die regulativen Ideen ein doktrinaler Glaube? Über die Rechtfertigung des regulativen Gebrauchs derIdeen im Anhang zur transzendentalen Dialektik', in: Violetta Waibel et al. (eds.), *Natur und Freiheit*, Berlin: de Gruyter, 1207-1216.

Gava, Gabriele (2019): 'Kant and Crucius on Belief and Practical Justification', in *Kantian Review*, 24, 53-71.

Gava, Gabriele (2022): 'Kant on the Status of Ideas and Principles of Reason', in *Open Philosophy*, 5, 296-307.

Gava, Gabriele (2023): *Kant's Critique of Pure Reason and the Method of Metaphysics*, Cambridge: Cambridge University Press.

Geiger, Ido (2003): 'Is the Assumption of a Systematic Whole of Empirical Concepts a Necessary Condition of Knowledge?", in *Kant Studien*, 94, 273-298.

Geiger, Ido (2015): 'How Are the Different Formulas of the Categorical Imperative Related?', in *Kantian Review*, 20, 395-419.

Gelfert, Axel (2006): 'Kant on Testimony', in *British Journal for the History of Philosophy*, 14, 627-652.

Gerhardt, Volker (1986): 'Handlung als Verhältnis von Ursache und Wirkung: Zur Entwicklung des Handlunsgbegriffs bei Kant', in: Gerold Prauss (ed.), *Handlungstheorie und Transcendentalphilosophie*, Frankfurt a.M.: Klostermann, 98-131.

Goldberg, Nathaniel J. (2004): 'Do Principles of Reason Have 'Objective bur Indeterminate Validity'?', in *Kant-Studien*, 95, 405-425.

Goldman, Avery (2012). *Kant and the Subject of Critique, On the Regulative Role of the Psychological Idea*, Bloomington: Indiana University Press.

Gomes, Anil (2017): 'Kant, the Philosophy of Mind, and Twentieth-Century Analytic Philosophy', in: Anil Gomes & Andrew Stephenson (eds.), *Kant and the Philosophy of Mind*, *Perception, Reason, and the Self*, Oxford: Oxford University Press, 5-24.

Gomes, Anil (2024): *The Practical Self*, Oxford: Oxford University Press.

Gomes, Anil, Moore, Adrian W. & Stephenson, Andrew (2022): 'On the Necessity of Categories' in *Philosophical Review*, 131, 129–168.

Gouaux, Charles (1972). 'Kant's View on the Nature of Empirical Psychology', in *Journal of the History of the Behavioral Sciences*, 8, 237-42.

Green, Garth W. (2010): *The Aporia of Inner Sense*, *The Self-Knowledge of Reason and the Critique of Metaphysics in Kant*, Leiden: Brill.

Grier, Michelle (1997): 'Kant on the Illusion of a Systematic Unity of Knowledge', in *History of Philosophy Quarterly*, 14, 1-28.

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 Siginifcance of Systematicity', in *Nous*, 24, 17-43.

Guyer, Paul (2003): 'Kant on the Systematicity of Nature: Two Puzzles', in *History of Philosophy Quarterly*, 20, 277-295.

Guyer, Paul (2019): *Kant on the Rationality of Morality*, Cambridge: Cambridge University Press.

Hacking, Ian (1975): *The Emergence of Probability*, *A Philosophical Study of Early Ideas About Probability Induction and Statistical Inference*, Cambridge: Cambridge University Press.

Haji, Ishtiyaque (2012): *Reason's Debt to Freedom*, *Normative Appraisals, Reasons, and Free Will*, Oxford: Oxford University Press.

Hatfield, Gary (1990): *The Natural and the Normative*, *Theories of Spatial Perception from Kant to Helmholtz*, Cambridge, MA: MIT Press.

Hatfield, Gary (1992): 'Empirical, Rational, and Transcendental Psychology: Psychology as Science and as Philosophy', in: Paul Guyer (ed.), *The Cambridge Companion to Kant*, Cambridge: Cambridge University Press, 200-227.

Hatfield, Gary (2018): 'Baumgarten, Wolff, Descartes, and the Origins of Psychology', in: Courtney D. Fugate & John Hymers (eds.), *Baumgarten and Kant on Metaphysics*, Oxford: Oxford University Press, 61-77.

Hebbler, James (2022): 'Reason, Reflection, and Reliabilism: Kant and the Grounds of Rational and Empirical Knowledge', in: Camilla Serck-Hanssen & Beatrix Himmelmann (eds.), *The Court of Reason*, Berlin: de Gruyter, 735-742.

Heimsoeth, Heinz (1956): *Studien zur Philosophie Immanuel Kants, Metaphysische Ursprünge und Ontologische Grundlagen*, Köln: Kölner Universitätsverlag.

Heimsoeth, Heinz (1966): *Transzendentale Dialektik, Ein Kommentar zu Kants Kritik der reinen Vernunft, Ideenlehre und Paralogismen*, vol. 1, Berlin: de Gruyter.

Heimsoeth, Heinz (1971): *Transzendentale Dialektik, Ein Kommentar zu Kants Kritik der reinen Vernunft, Die Methodenlehre*, vol. 4, Berlin: de Gruyter.

Heßbrüggen-Walter, Stefan (2004): *Die Seele und ihre Vermögen, Kants Metaphysik des Mentalen in der "Kritik der reinen Vernunft"*, Paderborn: Mentis.

Heßbrüggen-Walter, Stefan (2015): 'Vermögen', in: Marcus Willaschek et al. (eds.), *Kant-Lexikon*, Berlin: de Gruyter, 2481-2184.

Hölder, Alfred (1873): *Darstellung der Kantischen Erkenntnistheorie mit besonderer Berücksichtigung der verschiedenen Fassungen der transscendentalen Deduction der Kategorien*, Tübingen: Laupp.

Hook, Robert (1705): 'A General Scheme, or Idea of the Present State of Natural Philosophy', in: Richard Waller (ed.), *The Posthumous Works of Robert Hooke*, London: Smith and Walford, 1-70.

Howard, Stephen (forthcoming): 'Kant's Concept of Force and its Application in Physics and Psychology' in: Luis Fellipe Garcia (ed.), *The Concept of Nature in Classical German Philosophy*, Berlin: de Gruyter.

Höwing, Thomas (2016): 'Kant on Opinion, Belief and Knowledge', in: Thomas Höwing (ed.), *The Highest Good in Kant's Philosophy*, Berlin: de Gruyter, 201-222.

Höwing, Thomas (2017): 'Kant über Wissen, Allgemeingültigkeit und Wahrheit', in: Giuseppe Motta & Udo Thiel (eds.), *Kant über die Einheit des Bewusstseins*, Berlin: de Gruyter, 114-129.

Huber, Jakob (2018): 'Pragmatic Belief and Political Agency', in *Political Studies*, 66, 651-666.

Hudson, Hud (1994): *Kant's Compatibilism*, Ithaca: Cornell University Press.

Hume, David (1740/2000): 'An Abstract of a Book Lately Published, Entitled, a Treatise of Human Nature', in: David F. Norton & Mary J. Norton (eds.), *David Hume, A Treatise of Human Nature*, Oxford: Oxford University Press.

Hutter, Axel (2015): 'Unvernunft', in: Marcus Willaschek et al. (eds.), *Kant-Lexikon*, Berlin: de Gruyter, 2419.

Hutton, James (2021): 'Kant, Causation and Laws of Nature', in *Studies in History and Philosophy of Science*, 86, 93-102.

Indregard, Jonas J. (2017): 'Self-Affection and Pure Intuition in Kant', in *Australasian Journal of Philosophy*, 95, 627-643.

Insole, Christopher J. (2019): 'Free Belief: The Medieval Heritage in Kant's Moral Faith', in *Journal of the History of Philosophy*, 57, 501-528.

Jauernig, Anja (2021): *The World According to Kant, Appearances and Things in Themselves in Critical Idealism*, Oxford: Oxford University Press.

Joyce, James (2019): 'Bayes' Theorem', in *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/bayes-theorem/>.

Kahn, Samuel (2019): *Kant, Ought Implies Can, the Principle of Alternate Possibilities, and Happiness*, Lanham: Lexington Books.

Kahneman, Daniel & Tversky, Amos (1983): 'Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment', in *Psychological Review*, 90, 293-315.

Kant, Immanuel (1900–): *Kants Gesammelte Schriften*, Königlich- Preußische und Deutsche Akademie der Wissenschaften (ed.), Berlin: Georg Reimer, de Gruyter.

Kant, Immanuel (1998): *Kritik der reinen Vernunft*, Jens Timmermann (ed.), Hamburg: Felix Meiner Verlag.

Keller, Pierre (1998): *Kant and the Demands of Self-Consciousness*, Cambridge: Cambridge University Press.

Kemp Smith, Norman D. (1918/1962): *A Commentary to Kant's 'Critique of Pure Reason'*, New York: Humanities Press.

Kern, Andrea (2021): 'Kant on Doubt and Error', in *Midwest Studies in Philosophy*, 45, 129-154.

Kitcher, Patricia (1982) 'Kant's Paralogisms', in *Philosophical Review*, 91, 515-547.

Kitcher, Patricia (1990): *Kant's Transcendental Psychology*, Oxford: Oxford University Press.

Kitcher, Patricia (2011): *Kant's Thinker*, Oxford: Oxford University Press.

Kitcher, Patrica (2015): 'Wissen', in: Marcus Willaschek et al. (eds.), *Kant-Lexikon*, Berlin: de Gruyter, 2668-2670.

Kitcher, Patricia (ms): 'Practical and Empirical Cognition of the Cognitive Self' (*manuscript*).

Kitcher, Philip (1986): 'Projecting the Order of Nature': in: Robert E. Butts (ed.), *Kant's Philosophy of Physical Science*, Dordrecht: Reidel Publishing, 201-238.

Kitcher, Philip (1994): 'The Unity of Science and the Unity of Nature', in: Paolo Parrini (ed.), *Kant and Contemporary Epistemology*, Dordrecht: Kluwer Academic, 253-272.

Kleist, Edward E. (2005): 'The Freedom to Design Nature: Kant's Strong Ought → Can Inference in 21st Century Perspective', in *Cosmos and History, Journal of Natural and Social Philosophy*, 1, 213-221.

Klemme, Heiner F. (1996): *Kants Philosophie des Subjekts, Systematische und entwicklungsgeschichtliche Untersuchungen zum Verhältnis von Selbstbewusstsein und Selbsterkenntnis*, Hamburg: Felix Meiner.

Kohl, Makus (2015): 'Kant and "Ought Implies Can"', in *Philosophical Quaterly*, 65, 690-710.

Kohl, Markus (forthcoming): 'Kant on Cognition and Knowledge', in: Andrew Stephenson & Anil Gomes (eds.), *The Oxford Handbook of Kant*, Oxford: Oxford University Press.

Korsgaard, Christine M. (1996): *The Sources of Normativity*, Cambridge: Cambridge University Press.

Kraus, Katharina (2016) 'Quantifying Inner Experience? Kant's Mathematical Principles in the Context of Empirical Psychology', in *European Journal of Philosophy*, 23, 1-27.

Kraus, Katharina (2019): 'The Parity and Disparity between Inner and Outer Experience in Kant', in *Kantian Review*, 24, 171-195.

Kraus, Katharina (2020): *Kant on Self-Knowledge and Self-Formation, The Nature of Inner Experience*, Cambridge: Cambridge University Press.

Kraus, Katharina & Freitag, Wolfgang (2022): 'An Expressivist Interpretation of Kant's I think', in *Noûs*, 56, 1-23.

Krauser, Peter (1987): 'Über den hypothetischen Vernunftgebrauch in der Kritik der reinen Vernunft', in *Archiev für Geschichte der Philosophie*, 69, 164-196.

Kreines, James (2009): 'Kant on the Laws of Nature: Laws, Necessitation, and the Limitation of our Knowledge', in *European Journal of Philosophy*, 17, 527-558.

Laudan, Larry (1981): *Science and Hypothesis, Historical essays on Scientific Methodology*, Dordrecht: Reidel.

Lavin, Douglas (2004): 'Practical Reason and the Possibility of Error', in *Ethics*, 114, 424-457.

Leary, David E. (1978). 'The Philosophical Development of the Conception of Psychology in Germany, 1780–1850', in *Journal of the History of the Behavioral Sciences*, 14, 113-121.

Leech, Jessica (2017a): 'Judging for Reasons: On Kant and the Modalities of Judgment', in: Anil Gomes & Andrew Stephenson (eds.), *Kant and the Philosophy of Mind*, *Perception, Reason, and the Self*, Oxford: Oxford University Press, 173-188.

Leech, Jessica (2017b): 'The Normativity of Kant's Logical Laws', in *History of Philosophy Quarterly*, 34, 351-370.

Lehmann, Ernst (1908): *Idee und Hypothese bei Kant*, PhD thesis, Leipzig: Universität Leipzig.

Locke, John (1689/1975): *An Essay Concerning Human*,Peter H. Nidditch (ed.), Oxford: Clarendon Press.

Longuenesse, Béatrice (2008): 'Kant's "I think" versus Descartes's "I am a thing that thinks"', in: Dan Garber & Beatrice Longuenesse (eds.), *Kant and the Early Moderns*, Princeton: Princeton University Press, 9-31.

Longuenesse, Béatrice (2017): *I, Me, Mine*, *Back to Kant, and Back Again*, Oxford: Oxford University Press.

Lu-Adler, Huaping (2012), *Kant's Conception of Logical Extension and Its Implications*, PhD thesis, Davis: University of California, Davis.

Lu-Adler, Huaping (2017): 'Kant and the Normativity of Logic', in *European Journal of Philosophy*, 25, 207-230.

Lu-Adler, Huaping (2018): *Kant and the Science of Logic*, *A Historical and Philosophical Reconstruction*, Oxford: Oxford University Press.

Lu-Adler, Huaping (2021): 'Kant and the Principle of Sufficient Reason', in *Review of Metaphysics*, 74, 301-330.

MacFarlane, John (2000): *What Does It Mean to Say That Logic is Formal?*, PhD thesis, Pittsburgh: University of Pittsburgh.

Madonna, Luigi C. (1992): 'Kant und der Probabilismus der Aufklärung', in: Norbert Hinske (ed.), *Aufklärung*, *Kant und die Aufklärung*, vol. 7, Hamburg: Felix Meiner Verlag, 25-42.

Makkreel, Rudolf A. (2003): 'The Cognition-Knowledge Distinction in Kant and Dilthey and the Implications for Psychology and Self-Understanding', in *Studies in History and Philosophy of Science*, 34, 149-164.

Malzkorn, Wolfgang (1999): *Kants Kosmologie Kritik, Eine formale Analyse der Antinomienlehre*, Berlin: de Gruyter.

Massimi, Michaela (2017) 'Grounds, Modality, and Nomic Necessity in the Critical Kant', in: Michela Massimi & Angela Breitenbach (eds.), *Kant and the Laws of Nature*, Cambridge: Cambridge University Press, 150-170.

Mattey, George J. (1986): 'Kant's Theory of Propositional Attitudes', in *Kant-Studien*, 77, 423-440.

Martens, Yazmeen (ms): 'Logical Form and Contradiction in Kant' (*manuscript*).

McLaughlin, Peter (2014): 'Transcendental Presuppositions and Ideas of Reason', in *Kant-Studien*, 105, 554-572.

McNulty, Michael B. (2014): 'Kant on Chemistry and the Application of Mathematics in Natural Science', in *Kantian Review*, 19, 393-418.

McNulty, Michael B. (2017): 'What is Chemistry, for Kant?', in *Kant Yearbook*, 9, 85-112.

Meier, Georg F. (1752): *Auszug aus der Vernunftlehre*, Halle: Johann Justinus Gebauer.

Merritt, Melissa (2015): 'Varieties of Reflection in Kant's Logic', in *British Journal for the History of Philosophy*, 23, 478-501.

Messina, James (2017) 'Kant's Necessitation Account of Laws and the Nature of Natures', in: Michela Massimi & Angela Breitenbach (eds.), *Kant and the Laws of Nature*, Cambridge: Cambridge University Press, 131-149.

Mischel, Theodore (1967): 'Kant and the Possibility of a Science of Psychology' in *Monist*, 51, 599-622.

Mohr, Georg (1991): *Das sinnliche Ich, Innerer Sinn und Bewusstsein bei Kant*, Würzburg: Könighausen & Neumann.

Mudd, Sasha (2017): 'The Demand for Systematicity and the Authority of Theoretical Reason in Kant', in *Kantian Review*, 22, 81-106.

Nayak, Abhaya C. & Sotnak, Eric (1995): 'Kant on the Impossibility of the Soft Sciences', in *Philosphy and Phenomenological Research*, 55, 133-151.

Nehring, Robert (2015): 'Witz', in: Marcus Willaschek et al. (eds.), *Kant-Lexikon*, Berlin: de Gruyter, 2675-2677.

Newton, Alexandra (2020): 'Kant on Negation', in *Australasian Journal of Philosophy*, 99, 435-454.

Newton, Isaac (1687/1999): *The Principia, Mathematical Principles of Natural Philosophy*, I. Bernard Cohen & Anne Whitman (eds.), Berkeley: University of California Press.

Newton, Isaac (1715): 'An Account of the Book Entituled Commercium Epistolicum Collinii & Aliorum, De Analysi Promota', in *Philosophical Transactions of the Royal Society of London*, 29, 173-224.

Nisenbaum, Karin (ms): 'Realizing Our Capacity to Know the Good: Kant on the Dialectic of Pure Practical Reason' (*manuscript*).

Ostaric, Lara (2009): 'Kant's Account of Nature's Systematicity and the Unity of Theoretical and Practcial Reason', in *Inquiry*, 52, 155-178.

Pasternack, Lawrence (2011a): 'Kant's Doctrinal Belief in God', in: Oliver Thorndike (ed.), *Rethinking Kant*, vol. 3, Newcastle: Cambridge Scholars Press, 200-218.

Pasternack, Lawrence (2011b): 'Regulative Principles and the Wise Author of Nature', in *Religious Studies*, 47, 411-429.

Pasternack, Lawrance (2014a): 'Kant on Knowledge, Opinion, and the Thresholds for Assent', in: Oliver Thorndike & Pablo Muchnik (eds.), *Rethinking Kant*, vol. 4, Newcastle: Cambridge Scholars Press, 55-74.

Pasternack, Lawrance (2014b): 'Kant on Opinion: Assent, Hypothesis, and the Norms of General Applied Logic', in *Kant-Studien*, 105, 41-82.

Pasternack, Lawrance (2014c): 'Kant's Touchstone of Communication and the Public Use of Reason', in *Society and Politics*, 8, 78-91.

Peirce, Charles S. (1932): *The Collected Papers of Charles Sanders Peirce*, vol. 2, Cambridge, MA.: Harvard University Press.

Pickering, Mark (2016): 'Kant's Theoretical Reasons for Belief in Things in Themselves', in *Kant-Studien*, 107, 589-616.

Pinker, Steven (2021): *Rationality, What It Is, Why It Seems Scarce, Why It Matters*, Dublin: Peguin.

Pippin, Robert B. (1987): 'Kant on the Spontaneity of Mind', in *Canadian Journal of Philosophy*, 17, 449-475.

Pollock, Lansing (1974): 'On Treating Others as Ends', in *Ethics*, 84, 260-261.

Pollok, Konstantin (2017): *Kant's Theory of Normativity*, *Exploring the Space of Reason*, Cambridge: Cambridge University Press.

Powell, C. Thomas (1990): *Kant's Theory of Self-Consciousness*, Oxford: Oxford University Press.

Proops, Ian (2021): *The Fiery Test of Critique,A Reading of Kant's Dialectic*, Oxford: Oxford University Press.

Rajiva, Suma (2006): 'Is Hypothetical Reason a Precursor to Reflective Judgement?', in *Kant-Studien*, 97, 114-126.

Rauscher, Frederick (2010): 'The Appendix to the Dialectic and the Canon of Pure Reason: The Positive Role of Reason', in: Paul Guyer (ed.), *The Cambridge Companion to Kant's Critique of Pure Reason*, Cambridge: Cambridge University Press, 290-309.

Reath, Andrews (1988): 'Two Conceptions of the Highest Good in Kant', in *Journal of the History of Philosophy*, 26, 593-619.

Reid, Thomas (1785/2002): *Essays on the Intellectual Powers of Man*, *A Critical Edition*, Knud Haakonssen & Derek R. Brookes (eds.), Cambridge: Cambridge University Press.

Reininger, Robert (1900): *Kants Lehre vom inneren Sinn und seine Theorie der Erfahrung*, Wien: Braumüller.

Richards, Robert J. (1980): 'Christian Wolff's Prolegomena to Empirical and Rational Psychology: Translations and Commentary', in *Proceedings of the American Philosophical Society*, 124, 227-239.

Rödl, Sebastian (2013): 'Why Ought Implies Can', in: Mark Timmons & Sorin Baiasu (eds.), *Kant on Practical Justification*, Oxford: Oxford University Press, 42-57.

Rohs, Peter (1987): 'Philosophie als Selbsterhellung der Vernunft', in: Forum für Philosophie Bad Homburg (ed.), *Philosophie und Begründung*, Frankfurt a.M.: Suhrkamp, 363-390.

Rosefeldt, Tobias (2017): 'Subjects of Kant's First Paralogism', in: Anil Gomes & Andrew Stephenson (eds.), *Kant and the Philosophy of Mind*, *Perception, Reason, and the Self*, Oxford: Oxford University Press, 221-244.

Rudolph, Oliver-Pierre et al. (eds.) (2004): *Die Psychologie Christian Wolffs, Systematische und historische Untersuchungen*, Berlin: de Gruyter.

Russell, Bertrand (1921): *Unpopular Essays*, London: George Allen & Unwin.

Schafer, Karl (2018a): 'A Brief History of Rationality: Reason, Reasonableness, Rationality, and Reasons', in *Manuscrito*, 41, 501-529.

Schafer, Karl (2018b): 'Constitutivism About Reasons: Autonomy and Understanding', in: Francois Schroeter & Karen Jones (eds.), *The Many Moral Rationalisms*, Oxford: Oxford University Press, 70-90.

Schafer, Karl (2019): 'Kant: Constitutivism as Capacities-First Philosophy', in *Philosophical Explorations*, 22, 177-193.

Schafer, Karl (2020): 'Transcendental Philosophy As Capacities-First Philosophy', in *Philosophy and Phenomenological Research*, 103, 661-686.

Schafer, Karl (2021): 'A System of Rational Faculties: Additive or Transformative?', in *European Journal of Philosophy*, 29, 918-936.

Schafer, Karl (2023): *Kant's Reason, The Unity of Reason and the Limits of Comprehension in Kant*, Oxford: Oxford University Press.

Schiemann, Gregor (1992): 'Totalität oder Zweckmäßigkeit: Kants Ringen mit dem Mannigfaltigen der Erfahrung im Ausgang der Vernunftkritik', in *Kant-Studien*, 83, 294-303.

Schmitz, Friederike (2015): 'On Kant's Conception of Inner Sense: Self-Affection by the Understanding', in *European Journal of Philosophy*, 23, 1044-1063.

Schönrich, Gerhard (1991): 'Kant und die vermeintliche Unmöglichkeit einer wissenschaftlichen Psychologie', in *Psychologie und Geschichte*, 2, 130-1387.

Schüssler, Rudolf (2020): 'Kant und der Probabilismus', in: Rudolf Schüssler (ed.), *Kasuistik und Theorie des Gewissens, Von Pascal bis Kant*, Berlin: de Gruyter, 193-216.

Sellars, Wilfrid (1970): '...this I or He or It (The Thing) Which Thinks', in *Proceedings and Address of the American Philosophical Association*, 44, 5-31.

Sensen, Oliver (2013) 'Kant's Constructivism', in: Carla Bagnoli (ed.), *Constructivism in Ethics*, Cambridge: Cambridge University Press, 63-81.

Serck-Hanssen, Camilla (2011): 'Der Nutzen von Illusionen: Ist die Idee der Seele unentbehrlich?', in: Bernd Dörflinger & Günter Kruck (eds.), *Über den Nutzen von Illusionen, Die regulativen Ideen in Kants theoretischer Philosophie*, Hildesheim: Olms, 59-70.

Silva Jr., Paul (forthcoming): *On Believing and Being Convinced*, Cambridge: Cambridge Univsersity Press.

Spoerhase, Carlos (2009): 'Die "mittelstrasse" zwischen Sceptizismus und Dogmatismus: Konzeptionen hermeneutischer Wahscheinlichkeit um 1750', in: Carlos Spoerhase et al. (eds.), *Unsicheres Wissen, Sceptizismus und Wahrscheinlichkeit 1550-1850*, Berlin: de Gruyter: 269-300.

Stang, Nicholas F. (2011): 'Did Kant Conflate the Necessary and the A Priori?', in *Noûs*, 45, 443-471.

Stang, Nicholas F. (2014): 'Kant, Bolzano, and the Formality of Logic', in: Sandra Lapointe & Clinton Tolley (eds.), *The New Anti-Kant*, London: Palgrave Macmillan, 192-234.

Stang, Nicholas F. (2016): *Kant's Modal Metaphysics*, Oxford: Oxford University Press.

Stang, Nicholas F. (2014): 'A Guide to Ground in Kant's Lectures on Metaphysics', in: Courtney D. Fugate (ed.), *Kant's Lectures on Metaphysics, A Critical Guide*, Cambridge: Cambridge University Press, 74-101.

Stephenson, Andrew (2017): 'Imagination and Inner Intuition', in: Anil Gomes & Andrew Stephenson (eds.), *Kant and the Philosophy of Mind, Perception, Reason, and the Self*, Oxford: Oxford University Press, 104-123.

Stephenson, Andrew (forthcoming): 'Kant and Kripke: Rethinking Necessity and the A Priori', in: James Conant & Jonas Held (eds.), T*he Palgrave Handbook of German Idealism and Analytic Philosophy*, London: Palgrave MacMillan.

Stern, Robert (2004): 'Does "Ought" Imply "Can"? And Did Kant Think It Does?', in *Utilitas*, 16, 42-61.

Stevenson, Leslie (2003): 'Opinion, Belief or Faith, and Knowledge', in *Kantian Review*, 7, 72-100.

Strawson, Peter F. (1966/2005): *The Bounds of Sense*, *An Essay on Kant's Critique of Pure Reason*, Milton Park: Routledge.

Stuhlmann-Laeisz, Rainer (1990): 'Formale und transzendentale Logik im Paralogismenkapitel von Kants Kritik der reinen Vernunft', in: Hubertus Busche et al. (eds.), *Bewußtsein und Zeitlichkeit*, *Ein Problemschnitt durch die Philosophie der Neuzeit*, Würzburg: Königshausen und Neumann, 61-74.

Sturm, Thomas (2001): 'Kant on Empirical Psychology: How Not to Investigate the Human Mind', in: Eric Watkins (ed.), *Kant and the Sciences*, Oxford: Oxford University Press, 163-184.

Sturm, Thomas (2009): *Kant und die Wissenschaften vom Menschen* Paderborn: Mentis.

Sturm, Thomas (2015): 'Hypothese', in: Marcus Willaschek et al. (eds.), *Kant-Lexikon*, Berlin: de Gruyter, 1059-1061.

Talbott, William (2016): 'Bayesian Epistemology', in *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/>.

Techert, Jakub (ms): 'Kant on the Scientific Method: The Employment of Hypotheses in 1785' (*manuscript*).

Teixeira, João (2019): *Kant, Philosophy of Mind and Cognitive Science*, College Station: Virtualbookworm.

Thielke, Peter (2015): 'Meinung, Meinen', in: Marcus Willaschek et al. (eds.), *Kant-Lexikon*, Berlin: de Gruyter, 1502-1504.

Tolley, C. (2006): 'Kant on the Nature of Logical Laws', in *Philosophical Topics*, 34, 371-407.

Trullinger, Joseph S. (2013): 'Kant's Two Touchstones for Conviction: The Incommunicable Dimension of Moral Faith', in *Review of Metaphysics*, 67, 369-403.

Turnbull, George (1740): *The Principles of Moral Philosophy*, *An Enquiry into the Wise and Good Government of the Moral World*, London: John Noon.

Valaris, Markos (2008): 'Inner Sense, Self-Affection, and Temporal Consciousness in Kant's Critique of Reason', in *Philosophers' Imprint*, 8, 1-18.

van der Schaar, Maria S. (2003): 'Opinion, Assertion and Knowledge: Kantian Epistemic Modalities', in: Timothy Childers & Ondrej Majer (eds.), *The Logica Yearbook 2002*, Prague: Filosofia, 259-268.

Vanzo, Alberto (2012): 'Kant on Experiments', in *Studies in History and Philosophy of Science*, 28, 75-95.

Vogel, Jonathan (1993): 'The Problem of Self-Knowledge in Kant's "Refutation of Idealism": Two Recent Views', in *Philosophy and Phenomenological Research*, 53, 875-887.

Walker, Ralph (2017): 'Self and Selves', in: Anil Gomes & Andrew Stephenson (eds.), *Kant and the Philosophy of Mind, Perception, Reason, and the Self*, Oxford: Oxford University Press, 204-220.

Walsh, William H. (1966): 'Philosophy and Psychology in Kant's Critique', in *Kant-Studien*, 57, 186-198.

Wang, Kuizhi L. (2023): 'Kant on Propositional Content and Knowledge', in *Kant Yearbook*, 15, 175-196.

Warda, Arthur (1922): *Immanuel Kants Bücher*, Berlin: Martin Breslauer.

Wartenberg, Thomas E. (1979): 'Order through Reason: Kant's Transcendental Justification of Science', in *Kant-Studien*, 70, 409-424.

Washburn, Michael C. (1976): 'Did Kant Have a Theory of Self-Knowledge?', in *Archiv für Geschichte der Philosophie*, 58, 40-56.

Watkins, Eric (2005): *Kant and the Metaphysics of Causality*, Cambridge: Cambridge University Press.

Watkins, Eric (2010): 'The Antinomy of Practical Reason: Reason, the Unconditioned and the Highest Good', in: Reath, Andrews & Timmermann, Jens (eds.), *Kant's Critique of Practical Reason, A Critical Guide*, Cambridge: Cambridge University Press, 145-167.

Watkins, Eric (2013): 'Kant on Infima Species', in: Stefano Bacin et al. (eds.),*Kant und die Philosophie in weltbürgerlicher Absicht*, Berlin: de Gruyter, 283-294.

Watkins, Eric (2019a): *Kant on Laws*, Cambridge: Cambridge University Press.

Watkins, Eric (2019b): 'Kant on Real Conditions', in: Violette L. Waibel et al. (eds.) *Natur und Freiheit*, Berlin: de Gruyter, 1133-1140.

Watkins, Eric (ms): 'The Systematicity of Logical and Real Conditions' (*manuscript*).

Watkins, Eric & Stratmann, Joe (forthcoming): 'Kant on the Principle of Sufficient Reason', in: Michael Della Rocca & Fatema Amijee (eds.), *The Principle of Sufficient Reason*, Oxford: Oxford University Press.

Watkins, Eric & Willaschek, Marcus (2017): 'Kant's Account of Cognition', in *Journal of the History of Philosophy*, 55, 83-112.

Watkins, Eric & Willaschek, Marcus (2020): *Kant on Cognition and Knowledge*, in *Synthese*, 197, 3195-3213.

Waxman, Wayne (1991): *Kant's Model of the Mind, A New Interpretation of Transcendental Idealism*, Oxford: Oxford University Press.

Westphal, Kenneth (2004): *Kant's Transcendental Proof of Realism*, Cambridge: Cambridge University Press.

Willaschek, Marcus (2010): 'The Primacy of Practical Reason and the Idea of a Practical Postulate', in: Andrews Reath & Jens Timmermann (eds.), *Kant's Critique of Practical Reason, A Critical Guide*, Cambridge: Cambridge University Press, 168-196.

Willaschek, Marcus (2016): 'Must We Believe in the Realizability of Our Ends? On a Premise of Kant's Argument for the Postulates of Pure Practical Reason', in: Thomas Höwing (ed.), *The Highest Good in Kant's Philosophy*, Berlin: de Gruyter, 223-243.

Willaschek, Marcus (2018): *Kant on the Sources of Metaphysics, The Dialectic of Pure Reason*, Cambridge: Cambridge University Press.

Wolff, Christian (1736/1962): *Philosophie prima sive Ontologia*, in: Jean École et al. (eds.), *Christian Wolff, Gesammelte Werke*, vol. II/3, Hildesheim: Olms.

Wolff, Christian (1740/1972): *Psychologia rationalis*, in: Jean École et al. (eds.), *Christian Wolff, Gesammelte Werke*, vol. II/6, Hildesheim: Olms.

Wolff, Christian (1740/1983): *Der vernünfftigen Gedancken von Gott, der Welt und der Seele des menschen, auch allen Dingen überhaupt, anderer Theil, bestehend in ausführlichen Anmerckungen*, in: Charles A. Corr et al. (eds.), *Christian Wolff, Gesammelte Werke*, vol. I/3, Hildesheim: Olms.

Wolff, Christian (1751/1983): *Vernünfftigen Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt*, in: Charles A. Corr et al. (eds.), *Christian Wolff, Gesammelte Werke*, vol. I/2, Hildesheim: Olms.

Wood, Allen W. (1970): *Kant's Moral Religion*, Ithaca: Conell University Press.

Wood, Allen W. (1999): *Kant's Ethical Thought*, Cambrige: Cambridge University Press.

Wuerth, Julian (2014): *Kant on Mind, Action, and Ethics*, Oxford: Oxford University Press.

Yuen, Michael (2022): *Kant on Rational Faith and Hope*, PhD thesis, Canberra: Australian National University.

Zedler, Johann H. (1731–1754): *Grosses vollständiges Universal-Lexicon Aller Wissenschafften und Künste*, Halle: Zedler.

Zeldin, Mary-Barbara (1971): 'The Summum Bonum, The Moral Law, and the Existence of God', in *Kant-Studien*, 62, 43-54.