Citation for published version (APA):
1. Introduction

Kant’s philosophical project brings to a close one strand in the history of philosophy and inaugurates another. Kant himself presented the philosophical trends of the Early Modern period as culminating in his own Critical project. Similarly, the narrative of the subsequent development of philosophy in the 19th Century is one whereby various elements of Kant’s philosophy were either explicitly rejected or recovered. A notable feature of this succeeding trend is that Kant’s particular philosophical position of *transcendental idealism* has been a less popular target for recovery than other broadly ‘Kantian’ or ‘Critical’ aspects of his thinking. It is arguably these latter aspects that constitute Kant’s more lasting influence. For this reason it is a worthwhile endeavour both to distinguish and connect the general tenor of Kant’s revolutionary approach to philosophical thinking and the particular philosophical conclusions he formed as a result of that approach.

That task is fruitfully approached through a consideration of the question of the status of metaphysics as first outlined in the *Critique of Pure Reason*. How to consider Kant’s attitude to metaphysics is itself a somewhat vexed question however. Kant’s particular

---

1 This is not to say that this is the only way to approach the task. Kant’s ‘Critical’ philosophical inquiries (those beginning with the publication of the first edition of the *Critique of Pure Reason* in 1781) concerned epistemology, metaphysics, the philosophy of mind, ethics, aesthetics, political philosophy, and the philosophy of religion. My aim here is neither to summarize his contribution to each of these disciplines nor to consider the shape of transcendental idealism as it evolved in a range of works throughout the Critical period. Instead, my focus here is upon his particular doctrine of transcendental idealism as it was introduced in the 1780s. Although my focus therefore will be primarily to the *Critique of Pure Reason* (the ‘First Critique’) I will make some reference to the role that transcendental idealism plays with regard to the formulation of Kant’s ethics in the *Groundwork of the Metaphysics of Morals* (1784, the ‘Groundwork’). The first edition of the First *Critique*— the ‘A’ edition – was revised in 1787 in a second ‘B’ edition. A fuller presentation of these themes would require regular reference to the *Critique of Practical Reason* (1788 – the ‘Second Critique’) and especially to the *Critique of the Power of Judgment* (1790 – the ‘Third Critique’). Of lesser but still notable importance, which must go undiscussed here, are *Religion within the Boundaries of Reason Alone* (1793), *Metaphysics of Morals* (1797), *Perpetual Peace* (1795), or *Anthropology from a Pragmatic Point of View* (1798). All references and translations are from the Cambridge Edition of the Work of Immanuel Kant. All references to the First *Critique* are from (Kant 1998) and will employ the standard A/B referencing; all other references will employ the standard *Akademie* referencing convention.

2 For some discussions see (Allais 2009; Allison 2004; Ameriks 2003; Adams 1997; Beiser 1992; Hogan 2010; Moore 2011; Strawson 1966). The secondary literature for each individual aspect of
philosophical position manages to be somehow both distinctive and elusive. A central contention of his approach however is the claim that while metaphysics must surely concern things in general, and that this must surely include things independent of the minds of human beings, one must still take account of the human share even when considering such things. This is for the reason that while the subject matter of metaphysics can concern things independent of human minds, metaphysics considered as a practice is not independent of human minds. It is not a radical thought to maintain that metaphysics considered as a practice is shaped and restricted by the nature of the representational capacities of its practitioners. A more radical claim though is that how we conceive of the subject matter itself is determined by how we antecedently conceive of the practice. Many had previously claimed metaphysical results as a consequence of a recommendation regarding new metaphysical methodology. Such manoeuvres are the explicit strategies of Bacon, Descartes, Locke, Hume, Leibniz, and others. Perhaps more groundbreaking is Kant’s contention that once his particular methodological reorientation is made, then certain philosophical problems themselves receive answers by virtue of a reconceptualization of the questions.

What is transcendental idealism? There is a simple definition of transcendental idealism that captures one aspect of Kant’s philosophical project. The simple definition involves a metaphysical thesis regarding the status of space and time and the distinction between ‘appearances’ and ‘things in themselves’. Transcendental idealism claims that space and time are products of minds and not features of things in themselves (A369, A490-1/B518-19). Most would agree however that this definition does not capture some of the most important aspects of Kant’s philosophical ambitions in the First Critique. Some other recognizably Kantian claims might be added. If one were to add a second radical metaphysical claim, namely that the laws of nature are themselves also the products of the human mind (A127-8), arguably one would have yet to touch on what is characteristic of the Kantian vision. One might persist by adding to these putatively metaphysical claims two further epistemological ones: firstly, the positive claim that objective representations of the world are possible with regard to the world of appearances, i.e. the world insofar as it is constituted by the products of the human mind (A39/B56); secondly, the notorious negative thesis that similarly objective representations of things as they are in themselves, i.e. of things as they are anyway, is impossible (A30/B45, A42/B59).³

Even after this one would still have grounds for complaining that the core of Kant’s thought is inadequately characterized by this tetrad of claims. What characterises the core of Kant’s way of thinking is not his mere adherence to these claims but in the method by which he argues for them. That method involves reflections on the nature of metaphysical inquiry itself, reflections that involve radical claims regarding the character of the concepts with which one conducts such inquiry. Kant’s metaphysical results involve claims regarding space, time, mathematics, extension, causation, substance, laws of nature, the self, God, the origin, age and size of the universe, the scientific method (and more); Kant’s metaphysical methodology involves higher-order reflections on the

³ The allusion here is to Williams’s characterization of the Cartesian ambition for absolute knowledge as knowledge of ‘what there is anyway’ (Williams 2005, 48). Kant fundamentally differs from Descartes in that he rejects the very ambition to – and not merely the possibility of – knowledge of this sort.
interconnections between the concepts of truth and reality, self and world, and subjectivity and objectivity. He argues for a specific understanding of these pairs of concepts, whereby one member of the pair cannot be understood in isolation from the other, and that this fact constrains what we might understand metaphysical inquiry as capable of achieving. One way of pursuing the topic of the meaning of Kant’s ‘transcendental idealism’ then is as a distinctive way of making metaphysical sense of things that is not identical to any of the particular results that he secured through the exercise of that methodology.

This chapter will be structured as follows. In the following section I will outline Kant’s so-called ‘Copernican Turn’, which is key to the methodological shift that makes transcendental idealism possible. In §3 I discuss some of the key terminology of the Kantian project in the First Critique. In §4 I detail how these concepts are put to positive use in validating certain traditional metaphysical concepts. In §5 I turn to the negative task of Kant’s project, which shows that certain other traditional metaphysical concepts are in his view incapable of either vindication or refutation. In §6 I discuss how these negative restrictions of Kant’s project are themselves geared towards a return to a positive register. This is that of making a conceptual space possible – that of the practical sphere – wherein some of these latter metaphysical concepts can be thought of as having a genuinely possible application, albeit one we cannot fully comprehend. Transcendental idealism secures the peculiar result of a providing a theoretical justification for the claim that certain concepts themselves cannot be proven (see A479/B507 – note). More surprising still is his ultimate conclusion that such concepts emerge as possessing an importance within the human being’s existential orientation just on account of that justificatory lacuna. For Kant, the strategy of demonstrating certain concepts’ lack of comprehensibility is part of the very task of showing how and why they must nevertheless be preserved as a perennial part of human beings’ conceptual repertoire. I conclude in §7 with some reflections on the limitations and appeal of Kant’s project.

Kant’s overarching ambition is to provide a vindication of an extensive set of traditional metaphysical concepts whilst acknowledging that this set cannot find a place within a single rationally tractable ontological order. In Kant’s view, previous attempts at metaphysics have suffered from misguided ambitions either to claim that all the notions most characteristic to the human being’s worldview – from space, time, substance and causation to those of the self, freedom and God – have a genuine application within a single order (which he thinks cannot be done), or to eliminate the notions that lack a valid application in a single order (which he thinks does too much violence to human beings’ existential condition, rendering some of their most intimate experiences fundamentally misconstrued). Rather than gerrymander the boundaries of the rational order or slim down our conceptual scheme, Kant formulates transcendental idealism as the only system that can explain how there might be two rational orders, each bearing the correct ontological and epistemological relations to human beings needed to account for all the core notions. According to the aspirations of the Enlightenment project, we humans are capable of a rational understanding of a scientific reality made up of

---

4 These themes are particularly brought to the fore in both (Allison 2004; Strawson 1966), interpreters who nevertheless read Kant’s attitude to metaphysics very differently.

5 For variations on this theme see (Moore 2011).

6 This way of characterizing matters, i.e. of making sense of things in terms of clarifying both the required number of explanatory orders and the relations between them, is indebted to (Price 1964).
2. The Copernican Turn

Kant wrote the first edition of the *Critique* in 1781 – and thereby inaugurated the so-called ‘Critical’ period of his philosophy that was to occupy him for another quarter of a century – when he was already 57 years of age. The production of that work was the result of the second act of his adult life, the first act having been taking up in large part with exercises in scientifically-informed rationalist metaphysics, inquiries that he himself came to regard as inconclusive at best by the time he reached middle age. In the 1760s Kant adopted a significantly pessimistic outlook regarding both the possibility of successful metaphysical inquiry and the value of the sciences generally. Reflecting on his own past arguments and under the influence of the anti-rationalist and anti-scientific rhetoric of Rousseau in particular, he publicly declared that while metaphysics was still a splendid ambition, ‘no metaphysics has yet been written’. Metaphysics teemed with unsolved problems, such as the validation of the basic concepts of substance and causation and their connection to the natural sciences. Kant had come to believe that standard proofs for the existence of God were thoroughly hopeless, that no satisfactory proof of the external world had yet been provided, and so on.

Worse still perhaps, Kant accepted the force of Pierre Bayle’s contention that reason could be deployed to equal effect for and against certain propositions. To take a single example, the apparent availability of sound arguments both for the existence of simple substances and the infinite divisibility of the space within which those substances reside had nagged at Kant throughout his life. The intuitiveness of both the existence of freedom and the truth of determinism seemed a similarly galling antinomy. Progress was only made insofar as philosophers helped themselves to unwarranted concepts with magical problem-solving properties: the materialist explained the nature of consciousness by positing the notion of ‘thinking matter’ (leading Rousseau to quip that they would

---

7 For biographical information see (Kuehn 2001). For some accounts of Kant’s intellectual development from the Pre-Critical to Critical Periods, see (Ameriks 2012; Carson 2004; Cassirer 1983; Watkins 2005; Schönfeld 2000; R. L. Velkley 1989); for background on the German philosophical climate see (Beck 1969).
8 See (Zammito 2012) and (R. L. Velkley 2013) for discussion.
9 *Inquiry* (2: 283, in (Kant 1992)).
10 Kant first reached these conclusions regarding God in *The Only Possible Argument in Support of a Demonstration of the Existence of God* in 1763, in (Kant 1992). He complains of the ‘scandal’ of the lack of a proof of the external world at Bxxxix.
12 He addresses the issue in *Physical Monadology* in 1756 (1: 478, in Kant 1992), the *Inquiry* in 1764 (2: 279) and in the Second Antinomy in the First *Critique* in 1781 (A434/B462ff).
rather grant feeling to a stone than a soul to a human being); Leibniz explained the possibility of a substance possessing only obscure representations by characterizing it in terms of the composite concept of a ‘slumbering monad’. So long as such undisciplined concept-mongering was accepted, philosophy resembled less a science and more the supernaturalism of the contemporaneous popular mystic Swedenborg.

The Enlightenment project of establishing a rational metaphysics to supplement the success of the sciences had evidently overreached dangerously. The threat was not merely that the constructed metaphysical structures were vulnerable to sceptical exposure as ‘castles in the sky’ or a mere playing with words; the spirit behind the ambition was also at risk of being discredited as a result. For Kant, the rapid oscillation between dogmatic construction and sceptical exposure threatened to generate a cultural indifference to the very idea of the weight of rational warrant for or against any given proposition (Axiom, A3/B7). The Enlightenment project more broadly construed had hoped to organize human life around the establishment of the authority of reason over appeals to blind feeling, uncomprehending faith, or arbitrary power. The fear was that the parlous state of metaphysics might lead to throwing the baby out with the bathwater and de-legitimating the very ideal of a ‘culture of reason’ (Bxxx, A850-1/B879).

Kant had though acquired an insight regarding antinomial problems at some point in the late 1760s that fortuitously coincided with the availability of a chair in metaphysics in Königsberg. A hastily sketched new method of metaphysics was presented in an inaugural dissertation in 1770. This was followed by a decade of quiet perseverance working out the systematic expansion of that dissertation’s strengths and correction of its weaknesses. The product of that perseverance was the Critique of Pure Reason and the doctrine of transcendental idealism. Kant bemoans how metaphysics had once been the ‘Queen of the Sciences’ only to be subsequently ‘despised by all sides’ (A8ii) but who could now be restored to the throne. He was to some degree reporting the trajectory of his own personal relationship with metaphysics – the Critique represents Kant’s own belief that he had discovered the key to preserving much that was worth preserving in traditional rationalist metaphysics while retaining a place within this new system to accommodate, rather than dismiss, the forceful counter-arguments that motivated his previous pessimism.

In the preface to the second edition of the Critique Kant made claim to a kind of revolution for metaphysics. The rhetorical goal of the preface is to convince the reader that the required new method of metaphysics is in fact nothing but the scientific method as it had been practiced for millennia. The revolution that advocates for involves the

---

13 (Rousseau 1762/1979, 279).
14 Inquiry, 2: 277.
15 Kant first makes these claims in Dreams of a Spirit Seer (1766) in (Kant 1992).
16 The reference to ‘castles in the sky’ is made in Dreams, 2: 342. Locke’s Essay warned of succumbing to theories that construct ‘castles in the air’, lacking any possibility of falsification (Locke 1975, 4.4.1). Both Kant and Locke are possibly taking the phrase from Montaigne (Montaigne 2003, 426 II: 6, “On Practice”). By the time of the Critique Kant was well aware of the threat that metaphysical system-building must prove to be nothing but an elaborate form of self-indulgent intellectual fantasizing.
17 See (Deligiorgi 2005) for discussion of similar themes.
19 Inaugural Dissertation in (Kant 1992).
demand that objects conform to the structure of human rationality so as to secure the possibility that human beings are capable of objective representations of them (Bxxvi). The claim immediately looks suspect: our everyday experience of the world confirms objects as the kind of thing we have to adapt ourselves to in order to know them and not vice versa. Given the plainly counterintuitive character of the claim, the repackaging of the scientific method as having proceeded along similar assumptions is of vital suasive importance for the entire project.

Kant’s famous invocation of canonical scientific achievements – the first demonstrations of geometrical proofs, Galileo’s refutation of the Aristotelian claim that speed is relative to weight, and Copernicus’s proposal of a heliocentric astronomical model (Bxxvii), are all carefully chosen cases for his rhetorical purposes. In this context it is worth noting that Hume’s sceptical conclusions in Book I of his Treatise of Human Nature had been reached as a result of his initial self-avowed attempt to inaugurate a ‘science of man’, one that began by examining the nature and limits of human reasoning in accordance with the methods and standards of the Newtonian paragon. Thus Kant had to contend not just with Rousseau’s thought (infamous from the First Discourse) that the scientific endeavour was less worthy of esteem than Enlightenment thinkers maintained but also with the thought from Hume that the emulation of the sciences in philosophy would inevitably undermine rather than establish our rational grasp of the fundamental concepts that govern the natural world. Kant’s examples are chosen for the purpose of suggesting that not only was the emulation of the sciences in metaphysics still a live option it was in fact a hitherto unexplored one, since to date no philosopher had sufficiently attended to the actual commitments of the scientific method when properly conceived.

The scientific method, Kant held, was not best captured in crudely inductive terms, as the accumulation of empirical data followed by a hypothetical formulation of explanatory generalizations. The paradigmatic cases pointed instead to the need for subject-legislated demands for rational coherence prior to that accumulation of empirical data. Observational data is always interrogated in accordance with theory and the latter is always formulated in accordance with internal criteria of rational coherence. The geometrical case, for example, points to the simple fact that fundamental Euclidean propositions (such that the internal angles of a triangle necessarily are equal to the sum of two right angles) were not derived by a process of examining triangular objects found in nature and the subsequent inductive abstraction of their common features. Rather, necessary truths regarding geometrical types were stipulated prior to empirical inquiry and established by way of proof. Only then was nature examined in accordance with mathematical laws and found to conform to them. Similarly, Galileo’s ‘falling bodies’ thought experiment is invoked as a way of revealing the background contradictions in

---

20 When Kant speaks of human rationality he means it to have a broader scope to include all finite rational cognizers in general, so that the conclusions can claim objectivity and distinguish itself from mere anthropocentric generalizations.
22 (Rousseau 1750/2008).
23 Kant was intellectually well-positioned to appreciate the enormous explanatory resources of that scientific tradition. That appreciation would be worked out in fuller detail in 1786 in the Metaphysical Foundations of Natural Science (in (Kant 2012)). See (Friedman 2013) for an in-depth examination.
physical theories that can nevertheless be exposed through sound reasoning. More obviously, the Copernican shift of perspective is one that allowed for the full coherence of the theory to be properly expressed. There are some plausible realist intuitions suggested by ordinary perceptual experience (that heavier things might fall more quickly than lighter things; that the sun is in motion through the sky, etc.) that are nevertheless false and moreover proved to be impediments to the access of the objective law-governed facts regarding these differently-sized bodies. It is only when one granted a counter-intuitive assumption that progress could in fact be made and the advancement of humankind achieved.

Kant’s claim then is that the same manoeuvres must be made with regard to metaphysical methodology if metaphysics is to catch up with the progress of the sciences. In this context, the required shift of perspective is that objects must conform to human cognition. What could this mean however? It might connote a straightforwardly idealist claim, whereby we understand that empirical objects are creations of the human mind. Kant however explicitly denies that the existence of empirical objects is in any way mind-dependent. In this way Kant distinguishes transcendental idealism from more radical forms of idealism in a straightforward manner since it is not true, as it is for more radical idealists, that there are no minds the world itself would not exist. A lesser but still quite radical idealism would be one where a range of features other than its existence are nevertheless mind-dependent. There are more grounds for attributing this kind of idealism to Kant: while the empirical world’s very existence is not mind-dependent, the empirical world’s spatiotemporality, substantiality, causal structure, and so on are all merely mind-dependent features.

It is worth mentioning another meaning to the Copernican claim however, one that is purely methodological in character. This claim is not primarily about whether the existence of objects or their properties are mind-dependent, but rather is a quite general claim about how one ought to proceed when doing metaphysics. The claim is that one should inquire into the nature of objects by starting with examples of objects that are suitable to be targets of human cognition. This might sound uncontroversial at first but the real claim is that one should initially take the concept of objects to be well-understood in terms of the cognitive achievements of human beings. This is more controversial: one might have supposed that the objects we can cognize (if there are any) might represent a mere subset of the set of objects in general, and that it would be fallacious to infer the character of the latter from the character of the former. The historical cases are meant to suggest that successful science has always proceeded without such anxieties. Rather, scientists assumed that the objects it investigates are properly described in terms of rationally formulated lawlike generalizations just because those objects are supposed as being inherently apt for human reason’s investigations. Kant’s Copernican Turn involves the claim that metaphysical inquiry should make an analogous epistemological commitment explicit, and in doing so assume the hypothesis that the only well-understood characterization of what it is to be an object is that one that relates to the objects that we can cognize. Kant claims that if one begins with this methodological

24 For discussion see (Gendler 1998).
25 The Copernican analogy is examined at greater length in (Schulting 2010).
26 I use here Kant’s technical term ‘cognition’, which shall be discussed in more detail in the following section.
27 Prolegomena to any Future Metaphysics, 4: 288-9, in (Kant 2002, 84).
28 For hereon I designate mention of a concept by the use of italics and angle brackets.
constraint, one can give a successful account of causation, substance, space, time, etc. and thereby can vindicate the Copernican hypothesis initially assumed.

There is one yet further sense in which Kant’s Copernican turn aims to bring about a revolution in metaphysics. This concerns Kant’s epistemological focus upon investigating how a certain kind of judgment is made. Kant is interested in the possibility of *a priori* judgments: true judgments about universal and necessary aspects of empirical reality. Kant’s approach is to identify putatively secure epistemic achievements, such as those of mathematics, and to inquire how that type of *a priori* judgment is possible (Axi – note, Bx, B385-6, A4/B8, B4-5, B14-15, B20). Since for Kant an *a priori* judgment must involve the subject having some epistemic contact with the grounds that explain the truth of the judgment, mathematical judgments can offer an insight into how it could be that we can be in epistemic contact with the very things that make such judgments true. Moreover, Kant thinks that the grounds upon which such judgments are made need not be assumed to be exclusive to the subject of mathematics. Thus, if one can show that there are at least some cases of genuine *a priori* judgments, one can then investigate whether there might not be more non-mathematical *a priori* judgments that also qualify as genuine members of that set. The portion of the First *Critique* entitled the ‘Transcendental Analytic’ is largely devoted to showing that traditional metaphysical principles, such as that all events necessarily have a preceding cause, or that all changes are necessarily changes in an unchanging underlying substance, are indeed also genuine *a priori* judgments.

The core of Kant’s rhetoric then might be summarized as follows: metaphysics has tried and failed to justify our cognition of objects just because the notion of ‘object’ has been left uninterpreted. One should instead seek to stipulate from the beginning the meaning of that into which one is inquiring and moreover stipulate it in terms of a restriction, i.e. metaphysics is concerned with ‘objects’ insofar as that term connotes targets of possible cognition by finite rational beings. This is turn generates a notion of ‘objectivity’: a representation is objective if and only if its content matches the invariant outputs of the higher cognitive faculties of finite rational beings. Kant claims that in this newly-conceived context one can at last secure a genuine metaphysics of reality, one that will provide the grounding for the core metaphysical concepts that underpin both scientific investigation and common experience of the empirical world.

A reasonable metaphysical assumption in philosophical inquiry is that what is meant by ‘reality’ is reality *par excellence*: the ‘really real’, reality at the ‘fundamental’ or ‘deepest’ ontological level, reality as it expresses the ‘inner essences’ or ‘intrinsic properties’ of things, reality ‘as it is in itself’, or – to put it in epistemological terms – reality as it would be seen from a ‘God’s eye-point of view’. A moment’s reflection reveals that this sets the bar rather high for us humans. Many – perhaps most – of the things we take ourselves to cognize plausibly concern the less-than-fundamental aspects of things. Similarly, the things we take ourselves to cognize are cognized in ways that are plausibly non-identical with (and perhaps even incommensurable with) the way an omniscient being might be

29 I examine this topic and other Kantian terminology in further detail in the following section.
30 For some discussion of the notion of a priori as knowledge ‘from the grounds’ see (Hogan 2009; Smit 2009).
31 In this way Kant thinks he can secure important dialectical leverage over Hume, who had maintained scepticism over causal judgments though not mathematical ones (B19-20). According to Kant however, these judgments all belong to the same class and stand or fall together.
acquainted with them. The more demanding a definition of knowledge offered, the greater a threat to the possibility of humans actually achieving it is raised. Conceiving of the ‘reality’ that is the proper target of human beings’ objective representation as the domain of things as they are in themselves, where the latter is understood in terms of how a perfect and infinitely powerful knower would represent them, threatens to undermine the very possibility of such objective representation for us.

Kant’s sketch of the history of science was meant to suggest that scientists never tied their hands with such a demanding working definition of objective representation. It is this insight that Kant attempts to rework within a metaphysical context. Arguments for and against the possibility of cognition of the external world, of substance and causation, of laws of nature, of the self and other minds, of God, etc. have all foundered on the shared assumption of an implausible demand for objective representation to be representation of things in themselves. Dogmatists on the one side over-egg the power of our cognitive capacities so as to make it possible; sceptics on the other side draw pessimistic conclusions based on their adherence to that same demanding definition. Only once the definition is abandoned, and the relationship between objectivity and reality is redrawn, can both the achievements and limitations of human cognition be identified.

On this new picture the epistemological and metaphysical conceptions that determine what is at stake in philosophical inquiry are stripped of their theological connotations and instead are re-oriented relative to the capacities of finite beings.32 ‘Transcendental Idealism’ then has a very general connotation of an approach to the notion of objectivity that relativizes that concept to the particular cognitive needs and interests of finite beings. This also generates a general meaning of a general opposition position. ‘Transcendental Realism’ is Kant’s name for a genus of philosophical positions, all of which make a certain commitment regarding what kind of thing fills the role of ‘reality’ and what kind of epistemic subject fills the role of ideal possessor of an ‘objective’ representation. According to Transcendental Realism reality is made up of things in themselves, and the epistemic subject in question is that of an infinitely powerful intelligence (A369ff, A252/B308). According to this view one represents things objectively only when one represents the fundamental ontological level in the way one with a God’s-eye point of view would represent it.

Transcendental Idealism seems then to offer what seem like two lower-grade alternative candidates for these roles. Reality is now conceived of as the realm of mere possible and actual appearances of things (rather than the things in themselves) (Bxxix); the epistemic agent in question is that of a mere finite cognitive agent (rather than an infinite one). There is a sense in which Kant’s Critical philosophy can be thought of as making a ‘subjective turn’ but this too can prove more misleading than enlightening.33 It is undoubtedly true that it requires a focus upon the subject, but this is hardly at the cost of an inquiry into the nature of objects. Rather Kant’s ambition is to find a concept of ‘object’ commensurate with the minimal condition that it might be possible for finite rational subjects to cognize them. The transcendental idealist alternatives might appear lower-grade at first; however, part of the ambition of the first Critique is to re-present the

32 For an important corrective impression though see (Watkins 2013).
33 With regard to the latter claim, one could cite the obvious example of Descartes’s Meditations as a subject-centred methodology, and the prioritization of human standards as measures of inquiry is a commonplace humanist trope, advocated by Bacon, Montaigne, and many others.
competing methodological positions to the reader in way such that Transcendental Idealism loses its implausibility and Transcendental Realism loses its credibility. Both Transcendental Idealism and Transcendental Realism posit a correlation between the concepts <objectivity> and <reality>. Transcendental Realism, Kant claims, posits a correlation that would render a priori cognition impossible. Transcendental Realism is refuted then by the actuality of a priori cognition by human beings, most notably in the case of mathematics (B40-1, A38-9/B55-6). A crucial aim of the Critique is to show that the Transcendental Realist perspective is unmotivated, generates contradictions, and forbids us from explaining how it is that we might actually know that 2+2=4. In the face of these challenges the 'subjective’ Transcendental Idealist alternative is supposed to look less like a lower-grade alternative and more like the only game in town.

3. The Critical Apparatus

The wholesale re-evaluation of philosophical methodology requires both a basis upon which that re-evaluation might take place, and a new philosophical vocabulary with which to make it. The basis upon which philosophical theories are evaluated is that of 'critique'. The term is difficult to define though there is no doubt that it involves subjecting metaphysical claims to what Kant calls ‘the court of justice’ of reason (Axi, cf. A669/B697). A methodology that purports to criticize the nature, scope and authority of rationality's operation itself faces what might seem to us to be an obvious anxiety, namely how such a critique is itself supposed to be legitimated. Were it something other than reason performing the critique then one might have valid grounds for wondering why the critique ought to be viewed as rationally compelling; were the critique to proceed from rational grounds, one might wonder whether the exercise could be anything other than a circular one, a mere rationalization of rationality. Kant thinks it relatively unproblematic that it is reason itself that will be judging the validity of its own outputs. It is important to note again though that for Kant the exercise of critique does not proceed in the form of Cartesian meditating, whereby all possible prior commitments are put under a sceptical quarantine so that philosophy can proceed untethered from any first-order commitments. Instead Kant’s method is to take for granted the actuality of certain first-order rational achievements from the sciences that might serve as an anchor for the project.

Since Kant holds that metaphysics must be investigated in terms of the representational and epistemic powers required to make true metaphysical judgments, he introduces or re-introduces various pairs of notions to explain such achievements: a priori/a posteriori, concept/intuition, cognition/thought, receptive/spontaneous, synthetic/analytic, appearances/things in themselves.

A Priori/A Posteriori

‘A priori’ can function both adverbially and adjectivally, to describe a particular way in which we can cognize but also the outputs of such cognition respectively. Kant uses the term in a variety of ways that might involve some or all of the following features on different occasions. One might assume from the confident way in which Kant wrote however that he thought that some of use of the term was semantically continuous with the preceding tradition. Many of those connotations involve the idea that x is a priori if x concerns a necessary truth that can be determined to be true independently of any
particular experience (B3), and that our cognition of \( x \) is certain (Axi). Some perhaps more idiosyncratic and controversial aspects of Kant’s use of term are that \( x \) is a cognized a priori if \( x \) concerns a lawlike generalization (B4). A more contentious connotation still is that \( x \) is cognized a priori if \( x \) is cognized ‘from the grounds’, i.e. if one’s epistemic basis for accessing \( x \) involves the metaphysical basis that makes \( x \) true.\(^{34}\) It is unclear which (if any) of these are supposed to be individually sufficient for a cognition to count as an a priori one, or whether they all constitute individually necessary and jointly sufficient conditions.

A remarkable feature of Kant’s investigation into a priori cognition is its starting point, which is the claim that a priori cognition is a very common phenomenon indeed. Kant does his best to draw attention to this shift with the title to the second subsection of the B-Edition’s Introduction: ‘[w]e are in possession of certain a priori cognitions, and even the common understanding is never without them’ (B3). Rather than being a rarefied achievement of the reflective understanding of philosophical experts, Kant renders a priori cognition implicitly present within mundane cognitive achievements.\(^{35}\) By shifting the achievement of a priori cognition to ordinary perception, Kant could now re-present the question of its possibility in terms of a different one, that of ‘Plato’s Problem’, the question of how we can know so many necessary truths about reality given the paucity of our input from experience.\(^{36}\) Kant’s solution to Plato’s Problem is that the necessary truths that we can learn and articulate propositionally turn out to be generated from within the basic cognitive structure of ordinary human agents, a structure that in turn makes it possible for them to have any coherent experience of the world. The First Critique is less an exercise in the procurement of new a priori cognition and more a rational reconstruction of the grounds of the a priori cognition ordinary agents already possess yet also ordinarily lack the capacity to articulate explicitly.

This might make it seem that Kant is proposing a straightforward innatist theory of our accessing of necessary truths. However, his model of cognition is different from an innate ideas theory in important ways. Kant does not hold that reason possesses articulated propositional knowledge ‘within it’ in some sense, such that experience serves as an enabling condition for accessing it. Instead, Kant claims that the relevant propositional content is properly generated only upon the occasion of innate structures in the mind interacting with particular types of received input. Only when the mind is stimulated into action by the receipt of sensory input (which is itself processed as having a particular spatiotemporal form) does the mind first generate its a priori contents. What this means is that the contents of these necessary propositions are crucially tied to the context of the possible representation of spatiotemporal particulars.\(^{37}\) We can only come to cognize these truths because of the context of possible experience and for that very reason we come to see that those truths are properly restricted in their application to that context also.

---

\(^{34}\) The precise meaning here is a matter of some contention. For further discussion see (Kitcher 1980, 2006; Stang 2011).

\(^{35}\) It is obvious then that the ‘independence of experience’ criterion does not entail the absurd demand that subject only cognize a priori when they are not having any particular sensory experience. On the contrary, Kant thinks that the ‘independence’ in question concerns the explanatory or justificatory grounds for elements contained within every coherent experience.

\(^{36}\) The explanatory problem is expressed thus in (Russell 1948, v). Russell’s formulation is entitled ‘Plato’s Problem’ in (Chomsky 1986, xxv).

\(^{37}\) For discussion of Kant’s subtle take on innatism see (Callanan 2013).
A more fundamental way in which *a priori* representations differ from innate ones concerns the grounds upon which the *truth* of the knowledge acquired from those representations can be determined. There is an obvious sense in which some belief might have an innate source and yet might also be false (B167-8). Whether or not a subject can justify that belief depends on what now seems like a very different question, i.e. whether or not one is justified in thinking that claims that are based upon innate representations are true. Similar considerations can be brought to bear in considering the question of Kant’s continuity with Early Modern faculty psychology. While such continuity seems at one level to be present, there is a recognition that mere appeal to faculties is not sufficient for the explanation of the truth of the content conveyed by the representational outputs of those faculties. Therefore, different considerations than what Kant calls the question of fact (*quaestio facti*) – detailing the originating faculty for a representation – are needed to answer the question of right (*quaestio juris*) (A84/B116), i.e. as to whether that representation plays a necessary or sufficient explanatory role for our putative epistemic achievements.38

*A priori* cognition on the other hand aims to meet these requirements. To give a crude and simplified example: upon seeing four balls arranged in a pair of rows, one row on top of the other, one might come to cognize that 1+1=2. One’s cognition that this is the case, indeed that it is necessarily the case, is explained by appeal to the spatiotemporal representation that shows one pair aligned such that they must manifest a one-to-one relation with the other pair. When asked what justifies our claim that 1+1=2, an appeal to this representation alone would suffice to show that it is necessarily true. Moreover, according to Kant, arithmetical truths are fundamentally themselves truths about the most general features of spatiotemporal representation itself. Thus by appealing to properties of spatial particulars in this case one is also appealing to the very things that explain why it is true – one’s cognition is ‘from the grounds’ in the appropriately traditional sense, while also being a commonplace human epistemic achievement.

*A posteriori* cognition is relatively straightforward in that it can be characterized in contrast to *a priori* cognition. *A posteriori* cognition concerns the sensory content of experience directly, concerns cognition of contingent truths that can at best attain the status of universal generalizations, and is a form of cognition that lacks certainty.

**Concept & Intuition**

Intuition is a radical innovation of Kant’s own, since it in effect adds a new type of representational content to the repertoire of human beings than had been previously countenanced. ‘Intuition’ is in some respect a Kantian term of art.39 An intuition is fundamentally a representation of a particular by virtue of that particular itself being presented to consciousness (A19/B33). An intuition of something is therefore a representation of that particular thing by virtue of one’s having it actually there before one’s mind. In this way intuition depends on the presence of the thing represented (unlike thinking of, imagining, remembering or abstract reasoning about that object). A

---

38 See (Hench 1989; Proops 2003).
further but related feature is that of immediacy, in that an intuition of something involves the object being directly presented to consciousness, and not through any intermediary. It is hard to see how an intuition can manage this. Many interpretations are available, but two obvious candidates are to claim either that the intuitive representation itself suffices to play the role of object (such that Kant’s model of perception involves a kind of phenomenalism, reducing objects to collections of possible and actual representations) or to claim that intuition does not constitute a distinct entity but rather a mental state, i.e. the mental state of being in an immediate conscious relation to the thing represented.  

Whichever interpretation one pursues, it is clear that Kant thinks of human intuitions as sensible intuitions, i.e. intuitions that are activated in combination with the activation of our sensory modalities. The crucial point is that in Kant’s view, sensation alone (which Kant thinks constitutes the ‘matter’ of experience – A20/B34, A50/B74) cannot generate crucial representational and epistemic components: firstly, the features whereby a representation at least purports to pick out a particular and thus has the intentional content of being about a particular; secondly, that the representation does in fact manage to represent that very particular, i.e. it secures reference. Sensations for Kant are mere modifications in the consciousness of a subject – on a rocking boat one may represent both the movement of the boat and the dizziness one is experiencing as a result. The former purports to be about an object, the latter does not. When we have an ordinary perceptual experience – of the greenness of an apple, say – a sensation of greenness is necessary but far from sufficient to explain that experience. The mere modification of consciousness involved with that sensory component is sufficient neither for it to manifest as being about something nor does it in fact succeed in referring to that thing.

We require for some further non-sensory capacities to work upon our sensory input and process them to a degree whereby they can manifest in our conscious minds as being about a particular chunk of the world. ‘Intuition’ is Kant’s term for the capacity to harness sensory input in the aid of reference to given things. The candidate functions that he identifies are those capacities that might allow us to tie those sensations to a particular. Those capacities provide an ‘x marks the spot’ for our sensory input by providing it with spatiotemporal co-ordinates. By locating a cluster of sensations as being at somewhere and sometime we in effect pick out a point in space and time that is the proper target of our consciousness, and by doing so we represent a particular. Perhaps reflections like these led Kant to deny that space and time serve as part of the ‘matter’ of perceptual experience but rather in fact serve a special role of structuring perceptual experience by providing its ‘form’ (A20-1/B34-5). The cognitive hylomorphism here allows for an understanding of how sensible intuition might always be a cognitive achievement involving the presence of sensation without being reduced to that sensory contribution. We require something to have some matter as an enabling condition for perceiving its form but this kind of dependence claim does not entail the reductive claim.

Concepts are in one sense more straightforward in that Kant conceives of them in ways familiar to the preceding tradition, as discursive representations available for linguistic

---

40 The former interpretation can be found in (Broad 1978; Strawson 1966; Van Cleve 1999); a recent defender of the latter view is (Allais 2015).
42 See (George 1981) for discussion.
43 On Kant’s hylomorphism see (Engstrom 2006).
articulation by rational agents capable of self-consciousness. Here too though there are radical innovations. For one, Kant thinks of concepts in a new way, not in the empiricist manner, as images or pictures abstracted from sensed things, but rather as rules for the organization of sensible intuitions into coherent representations of objects (A126-7, A133/B172). It is similarly clear that Kant thinks that the task of successful reference cannot be secured by purely rationalist means, i.e. by clearly and distinctly thinking about a concept until it provides such a discursively fine-grained description that it picks out a particular individual. For Kant concept-use alone cannot present the actual individual that falls under the concept to consciousness. Secondly, Kant’s focus in the First Critique is not upon the origin, structure, and validation of ordinary empirical concepts that are usually taken as stock examples in contemporary philosophy of perception. Rather, Kant is concerned with a special set of concepts, which he calls ‘Categories’, many of which reflect traditional metaphysical notions such as those of substance and causation (A125). Kant’s cognitive hylomorphism is equally essential here, as he gives an account of these concepts as those rules that also provide a form for coherent perceptual experience. These discursive rules, he claims, generate a priori cognition when ‘schematized’, i.e. given expression in the context of possible spatiotemporal representation (A137/B176 ff.). Since these rules are interwoven into the structure of rationality per se, and (he claims) reflect universal principles that natural objects must follow, Kant argues that the sought-after goal of a priori cognition of a law-governed nature, long presumed a rarefied achievement involving considerable philosophical expertise, is in fact ubiquitous and found wherever there is rational consciousness.

Cognition & Thought

These themes help to understand Kant’s account of cognition [Erkenntnis], since it too serves a kind of term of art. It is clear that Kant intends it to involve the kind of cognitive relation a human mind can have to the world when the thing it is representing is in fact present and when it does in fact bear the features is it is represented as bearing. It serves some of the connotations of ‘knowledge’ then, since it involves having a true representation of things such that an understanding of the character of those things is coterminously made possible. Nevertheless, there are grounds to resist characterizing cognition as an act of knowing, not least considering Kant’s explicit attempt to distinguish cognition [Erkenntnis] from knowledge [Wissen], the latter which (along with belief, opinion, etc.) receives far less attention in the First Critique. Kant’s approach is to recast epistemic achievements as ones secured within what we would call the philosophy of mind, rather than (say) through explicit linguistically-articulated inference or similarly reflective practices. Such linguistic practice is what is going on in the arguments detailed on the pages of the Critique of Pure Reason, but the subject matter of those pages is an altogether different kind of epistemic achievement.

Kant’s elaboration of cognition makes it appear as an epistemic achievement exclusive to human beings. Non-human animals cannot secure it since it essentially involves concept-use, since (it is subsequently claimed) this essentially involves self-consciousness (B131-2). Cognition cannot then be understood in terms of a kind of sophisticated sensing. Conversely, cognition involves both concepts and intuitions as ineliminable elements

44 Properly considered, intuition is necessary but not sufficient for the above-mentioned task of explaining intentionality, i.e. the representation of objects as objects.
45 For a recent discussion see (Watkins and Willaschek 2017; Tolley 2012).
46 See A820/B848 ff. For discussion see (Chignell 2007).
so it is similarly misguided to think of cognition as sensorily encumbered occasions of thinking, as if sensation were an unfortunate material condition that reason must overcome in order to operate in a pure manner. Cognition is different in kind again from the type of epistemic achievement that might be secured through purely rational means, by way of clear and distinct discursive representation (or what we might today think of as analysis, as the successfully clarification of concepts through a perspicuous rendering of their contents in linguistically tractable terms). This focus upon cognition as essentially involving these elements are reflective of Kant’s general tendency towards synthesizing different traditions into a sui generis phenomenon, one that resists reduction to either.

**Receptive/Spontaneous**

What marks a finite cognitive agent is their dependence upon some distinct ontological participant to cooperate in their epistemic activities in order for that subject to secure cognition. For an infinite cognizer, the act of cognizing a thing is simultaneously the act of bringing that thing into being, and in this way there is no ontological gap between knowing and being. For a finite cognizer, on the other hand, one does not bring a piece of empirical reality into existence in the act of cognizing it. Rather the piece of reality has to come into view of its own accord in order to facilitate our cognition of it. One might summon up a representation of Tiananmen Square from one’s armchair, and that representation might correctly represent many real aspects of Tiananmen Square. Nevertheless, for Kant such a representation (a thinking or an imagining or a mental visualizing) cannot constitute a cognition of Tiananmen Square. It would only do so were I also in the very presence of that Square and if my representation were related to the Square in the right way. Obviously, I cannot bring it or any other portion of empirical reality into being just from the act of thinking of it. Conversely, I cannot think any piece of empirical reality out of existence – if it is in fact really there, thinking that it isn’t so has no efficacy on that state of affairs.

This finitude of the cognitive subject is referred to as its receptivity. We must receive outside assistance if we are to convert our representations of things into cognitions of them. Kant thinks that human receptivity is shaped by two different but deeply connected ‘forms’: those of space and time. To continue with the previous example, what is missing from my representation of Tiananmen Square from my armchair is the presence of Tiananmen Square itself. But this presence is captured by a subject’s being spatially and temporally related to the place itself. Typically I cannot have a cognition of something that is not spatially and temporally located for me. If I am at the right spatial location, e.g. a lecture room, but at the wrong temporal location, e.g. a day when the lecture is not occurring, I cannot have a cognition of the lecture.

The spontaneous character of human cognition is a more elusive element of Kant’s thought.\(^47\) It has one connotation of a response to received stimuli that is not determined and demanded by the character of those stimuli themselves. A further connotation concerns the notion of freedom, and those elements of our epistemic activity for which we as individuals can take responsibility. The higher cognitive faculties of understanding and reason are then the epistemic expression of our spontaneous nature. Since cognition

\(^{47}\) For discussion see (Pippin 1987; Longuenesse 1998; Callanan 2017).
is an achievement requiring both concepts and intuitions, it is an achievement exclusive to beings capable of both receptive and spontaneous capacities (A50/B74).

**Analytic/Synthetic**

Kant’s celebrated distinction between analytic and synthetic judgments no doubt cemented his image as a philosopher recasting metaphysical questions in a different form. While the distinction has clear historical antecedents, it is perhaps Kant’s framing of the relevance of the distinction that constituted its particular impact. Kant’s rhetoric involves the following claims: that the distinction itself has not been really appreciated in the history of philosophy; nor has it been realized that the distinction can be related to the *a priori/a posteriori* distinction in various ways; nor has it been noted (except perhaps obliquely by Hume) that core metaphysical questions are properly characterized as questions about the possibility of judgments that are both synthetic and *a priori* in character; nor that mathematical judgments are in fact synthetic *a priori* judgments (hence vindicating the possibility of the class of judgment in general). In this way, Kant presents the entire question of the possibility of metaphysics is in fact the question as to how synthetic *a priori* judgments could be possible (B19, A5-6/B9-10, A154/B193).

The distinction hinges on the notion of ‘containment’ and the idea that some concepts can be contained within others. For example, `<body>` contains the concept `<extension>` such that the judgment ‘all bodies are necessarily extended’ is true and its truth can be determined simply by consideration of what concepts are contained in the concept occupying the subject position of the judgment (A7-8/B11-12). Analytic judgments then are judgments that would appear tautologous if appropriately clarified. A judgment such as ‘all bodies possess weight’ could be true or false, but its truth or falsity cannot be determined by consideration of `<body>` alone, since that concept does not include the concept `<weight>` inside it. Such latter judgments Kant calls ‘synthetic’.

A natural but false assumption of the entire preceding philosophical tradition, Kant contends, is that the notions of apriority and aposteriority align with those of analyticity and syntheticity respectively. Kant’s guiding insight is that the limitations of that tradition can be expressed with the idea that they failed to see that there can be judgments that are synthetic yet *a priori*. Crucially – though controversially – Kant claims that mathematics consists of such judgments. The concept `<the sum of seven and five>` does not ‘contain’ (in whatever technical sense that connotes) `<twelve>` within it, Kant claims, and so the necessarily true judgment ‘the sum of seven and five is twelve’ cannot be an analytic one (B15-16). Kant claims (with a little more intuitive plausibility) that the necessarily true judgment ‘there is only one straight line between any two points’ is not contained within `<straight line>` or `<point>` (A24/B39).

In all these cases, Kant suggests, the truth of the judgment is secured because two concepts are connected via a third thing – a possible spatiotemporal intuition – that provides the referent of that judgment. Kant’s claim then is that the paradigmatically secure epistemic achievement of mathematical achievement is in fact coherent only because it is not simply a matter of conceptual analysis. Moreover, it provides firm ground for the idea that with regard to necessary truths in general the proper method might be to relate concepts to possible intuitions.

---

A more notorious part of Kant’s conceptual repertoire is the one between appearances and things in themselves. In some cases Kant appears to identify this distinction as the one between ‘phenomena’ and noumena’ respectively (A235/B394ff). It is fair to say that it is notorious in that, while the distinction is essential for the various positive philosophical tasks Kant envisages, the primary function of the distinction is a negative one. Kant claims that we can have no cognition of the domain of things as they are in themselves and that such cognition is restricted to things as they appear (A34-5/B51, A42/B59). One might wonder whether this negative claim is itself supported by cognition, and if it is not, what epistemic status it is supposed to have. Kant seems to draw attention to similar anxieties by virtue of the range of further claims he makes regarding things in themselves, such as that they exist, that they ground or underpin the appearances that we can cognize, and that they lack spatiotemporal features. It is still unclear over two centuries later on what grounds Kant thinks we can make such assertions about things in themselves.

Nor is it even clear as to what the very distinction might mean. A natural reading would be to posit two entirely ontologically distinct domains of entity. This though is sometimes thought hard to square both with Kant’s account of our relation to the world of appearances (which is not supposed to function as a veil obscuring our insight into ‘true’ reality) and with his seeming reconceptualization of notions in epistemic terms. ‘Thing in itself’ might seem to connote a distinct but shadowy entity yet alternatively might connote simply reality insofar as it cannot be known by imperfect epistemic subjects. Kant seems to indicate the term functions mostly to draw attention to the epistemic achievements available to us that nevertheless fall short of such perfection, and thus to articulate the sense of our cognition having bounds, rather than connoting some special object perpetually floating beyond our epistemic horizon (A254-5/B310). Kant may have had in mind the ‘inner determinations’, ‘intrinsic properties’ or fundamental character of the very same things we encounter as appearances, though it is unclear as to which expressions Kant takes as primary for understanding the distinction. If though the thing in itself is the very object that we cognize as appearance, it is unclear how we are to understand (for example) the apparent entailment that the very same thing is both spatial and non-spatial. Fundamentally however, whatever one makes of the phenomena/noumena distinction it is clear that Kant thinks it ought not to be aligned with the distinction between appearance and reality, for the entire thrust of Kant’s strategy in the Critique is to show that a coherent notion of reality (that of a spatiotemporal world containing causally interacting substances) can be articulated while remaining entirely within the realm of appearances.

4. Transcendental Strategies

49 For discussion see (Allison 2004; Gardner 1999; Langton 1998; Allais 2006, 2015). Whether the phenomena/noumena distinction can be completely aligned with the distinction between appearances/things in themselves is a matter of debate.

50 A familiar range of positions on this issue can be found in (Adams 1997; Allais 2015; Allison 2004; Paul Guyer 1987; Langton 1998; Van Cleve 1999).
The Copernican turn was meant to draw explicit attention to the fact that cognitive capacities are essentially in play in both the conception and exercise of metaphysical speculation. These cognitive capacities are capacities to represent things as being thus and so. By shifting traditional characterizations of metaphysical concepts into the register of representational capacities, Kant thought he could also avail of a hitherto unseen justificatory and explanatory strategy. Specifically, he held show that those concepts gained an application by appeal to an additional constraint to which one could appeal in order to motivate a new form of argumentation. Those concepts, now construed as representational functions, are required at the very least not to contravene the possibility of coherent representational experience itself. However, Kant thought he could show that not only did they not contravene that condition, but that their application was necessitated by that very condition. Kant sought to show that these concepts were valid, and that this could be shown once one granted their construal as conditions on the possibility of ordinary experience. Such reasoning has come to be known as transcendental argumentation.  

A standard example of such argumentation appears to be at work in the Metaphysical Exposition of the Concept of Space in the Transcendental Aesthetic. There Kant argues that Space must originally be an a priori intuition. A portion of the argument for that conclusion argues that the notion of space per se could not have emerged a posteriori as a series of reflections upon experienced things manifesting spatial relations to each other. It is a necessary condition of representing anything as bearing spatial relations, Kant claims, that we represent something as already occupying some position within a spatial continuum (A23/B38). Thus the order of dependency is reversed and the representation of space per se provides the grounds for the possibility of spatial relations.

Kant also employed transcendental argumentation to the end of securing the more familiar and controversial claims of Transcendental Idealism, such as the non-spatiotemporality of things in themselves. Later in the Transcendental Aesthetic, in the Transcendental Exposition of Space, Kant had aimed to show that the judgments of geometry, having been showed to be synthetic a priori in character, could only be true if the space that was the proximate target of such judgments was in fact nothing more than a representation. It is a necessary condition of their being necessarily true, it is claimed, that they reflect nothing but determinations of the mind (B40-1). The argument is mostly opaque and insofar as it is transparent is not compelling. The seeming assumption that there are eternal truths of Euclidean geometry that are both accessed a priori and are known to reflect the essential character of physical reality has proven to be an unfortunate hostage to scientific fortune. It is also not clear exactly how space considered as an a priori intuition somehow suffices to explain the modal status of those geometrical truths, as Kant seems to think it does. Neither is it clear how, even if it is true that the representation of space is sufficient to explain the apriority of geometry, it thereby precludes spatiality also being a non-representational property of things in themselves.

52 For analysis of the ‘Argument from Geometry’ see (Allais 2010; Paul Guyer 1987; Shabel 2004).
53 A similar claim regarding the a priori status of the laws of the understanding appears at B164ff.
Kant’s strategy would be misconstrued if thought to be one of replacing abstract reasoning with a theory of spatiotemporal experience. For Kant the transcendental model of cognition and experience expressed the required synthesis of rationalist and empiricist elements. The rationalist focus upon propositional structure and logical form is retained though now recast as the logical structure of experience itself. For a single example, the hypothetical form of judgment \((if \, x, \, then \, y)\) is held to be the abstract logical expression of the category of causation, when is expressed in its ‘schematized’ manifestation, i.e. in a spatiotemporal context. In this way, Kant makes a remarkable claim whereby logical form is understood as implicitly woven into the structure of perceptual experience itself.

To take a further example, Kant’s claims that the difficulty of proving the nature of the relation between substance and accident and proving that it applies to a world of objects can be resolved if one translates these questions into ones about the role of \(\text{substance-representations}\) and \(\text{accident-representations}\) operative within a subject’s own fluctuating manifold of perceptual experiences. The argument of the so-called First Analogy attempts to show that the more mundane representational achievements of drawing a distinction between things happening \(\text{successively}\) and happening \(\text{simultaneously}\) in fact requires the representation of changes in time to be interpreted in terms of substance-representations (A182/B224ff). Thus so long as genuine temporal representation is possible then genuine substance-representation is possible, and doubt about the latter cannot be cotermiously maintained with confidence in the former. Even the hardened skeptic would find it difficult to deny the reality of temporal experience (or at least they would struggle to deny it without acknowledging that even the denial itself must be an event in time). In the Second Analogy Kant makes a similar argumentative strategy with regard to the concept of causation, responding to the Humean doubt about its objectivity that famously roused Kant from his dogmatic slumbers and that initially stimulated his inquiries in this transcendental direction.

The sheer scope of Kant’s vision here is startling. Kant’s approach is remarkable in that it attempts to provide a single unified narrative that accounts for (i) the phenomenology of ordinary human experience; (ii) traditional metaphysical principles and (iii) core propositions of mathematics and physics. For example, the First Analogy attempts to account for the possibility of our experience in duration in ordinary temporal experience, to justify the metaphysical thesis that all changes are change in an underlying unchanging substance and to provide support for Lavoisier’s law of the conservation of matter. Transcendental idealism attempts both to outline and vindicate a scientific and mathematically tractable metaphysics of the ordinary perceptual experience of human beings. Such aspirations now seem outrageously ambitious. There is little reason to think that there should be any purely philosophical reconstruction of a correlation between the structure of the phenomenology of ordinary human experience and the fundamental principles of physics, or that there should even be such a correlation. The former is plausibly formed upon the basis of contact with what are quite possibly merely local features of human beings’ environment. That conscious experience has the structure it has is perhaps due to perceptual capacities having evolved to navigate these physical features successfully. To think that the structure of such perception is isomorphic with either fundamental physics or fundamental ontology seems fanciful. Here the strategy inherent in Copernican Turn threatens to emerge as a forced apology for our metaphysical parochialism.
Kant’s most difficult and ambitious argumentation in the First Critique arises in the Transcendental Deduction.\textsuperscript{54} Here Kant attempts an argument to show that the categories in general apply to an objective world just because they constitute such a world (A158/B197). The argument itself is obscure and contentious; several striking features of it that might be remarked upon however, since here Kant – quite remarkably – both introduced and attempted a unified solution for several philosophical problems. Firstly, Kant begins from a quasi-Cartesian starting point, noting that for every possible judgment, the ‘I think’ is at least in principle capable of preceding it. This claim leads Kant to argue that that the possibility of judgment itself is dependent upon conditions that the self contributes. The judgment ‘the cat is black’ has both a logical form \((x \text{ is } y)\) and particular concepts that appear in the subject and predicate positions. It also has a propositional unity that, Kant notes, is something that requires explanation (thus introducing the question of the \textit{unity of the proposition}). Kant’s answer appears to be that such unity is achieved only on occasions of the propositional form and concepts being actively unified by a thinking self.

Moreover those conditions emerge as related to other crucial features of experience, ones previously not regarded as achievements at all. That our characterization of objects represents them as unified objects (the question of \textit{cognitive binding}); that the self’s own conscious states constitute a unity (the question of the \textit{unity of the self}); that the intentional content of a representation is distinguishable from the representation itself (the question of the \textit{nature of intentionality}); that the object is regarded as fully present in perception despite being only partially presented (the question of \textit{perceptual presence}) – all these achievements are secured by the transcendental activity of the self utilizing functional rules expressed by the Categories. The Deduction thus argues that they can all be solved and are in fact manifestations of the same problem of explaining how a thinking self represents anything at all. Given such ambition, it is perhaps unsurprising that both the details and the success of the Deduction remain contested.

The notion of the self that emerged in the course of the Transcendental Deduction is radically different from any that preceded the Critique. Kant’s self is the ‘I’ of any potential ‘I think’, which in turn only arises when there is some possible judgment being made. No further content to this ‘I’ (considered as a ‘I-in-itself’) is accessible to this self whether by introspection or any other mode of knowing (B152-3). Kant’s self is known, insofar as it is known at all, through what it does rather than what it is: its character is revealed indirectly through its products, namely in generating cognition of objects. Here again the ambiguity of Kant’s transcendental approach is evident, since it can be equally well interpreted as either an explanation of the fundamental elusiveness of a metaphysical self or as a sophisticated deflation of the very idea of such a self.\textsuperscript{55}

The truly radical and surprising character of Kant’s Copernican way of thinking fully emerges from reflection on the Deduction (rather than perhaps in any particular section

\textsuperscript{54} Kant performed the remarkable achievement of rewriting his supernaturally opaque A-edition Transcendental Deduction in the B-edition so as to offer profound new insights without any significant alteration in terms of clarity. I don’t examine the arguments of the versions or their differences, but merely present here some notable themes from both. Discussions of these topics can be found in (Ameriks 1978b, 1982; Förster 1989; Paul Guyer 1992; Henrich 1989; Howell 1992; Schulting 2012; Strawson 1966).

\textsuperscript{55} The latter reading is perhaps found in (Allison 2004; Melnick 2009). For a defence of Kant’s metaphysical intentions see (Ameriks 2000b; Marshall 2010).
of it). For it is here in the Deduction that Kant’s aim of reconfiguring the concepts with which philosophy is done is more on show than anywhere else. Whereas the Copernican turn sought to question the relation between the concepts of objectivity and reality, the Critique now shows itself to be geared towards reformulating the very nature of what it is to be a subject and what it is to be an object. More radically still, it emerges that these very concepts are inextricably interdependent: what it is to be a ‘subject’ is just to be a cognizer of ‘objects’, suitably defined; what it is to be an ‘object’ is just to be the target of a cognizing subject. Once these insights have been suitably internalized, one might become suspicious of just what it could be to have epistemic insight into a ‘self’ that was understood as something other than the ordinary functions of organizing perceptual experience, or indeed to have understanding of an ‘object’ that was supposed to involve something other than the familiar acquaintance we have of the quotidian spatiotemporal particulars of the manifest and scientifically tractable world.

5. Transcendental Realism and Illusion

Kant begins the First Critique not with a statement regarding reason’s remarkable capacity to vindicate Newtonian physics, mathematics, and scientific rationality in general but with a statement regarding reason’s essentially vexed predicament:

Human reason has the peculiar fate in one species of its cognitions that it is burdened with questions which it cannot dismiss, since they are given to it as problems by the nature of reason itself, but which it also cannot answer, since they transcend every capacity of human reason.  

(Avii)

This is a remarkable enough claim, since it in effect maintains that human rationality is by its very nature incapable of satisfying itself. More striking still is the later claim that the questions that reason is determined to produce but cannot answer relate to the notions of God, freedom and immortality (A3/B7, B395–note, A798/B826). One possible reading of Kant’s motives here attributes a privately maintained scepticism regarding the validity of these notions and that the arguments of the First Critique play the diplomatic role of securing the impossibility of deciding either way as to their validity or invalidity. The reading would maintain that Kant first exerted his energies in rescuing scientific cognition and then threw whatever resources remained so as to grant these notions – so difficult to excise from human thinking – some lower-grade but nevertheless positive epistemic role within human beings’ lives. It is clear that Kant does not think that once these ideas have been exposed for what they are that one will then cease to be concerned with the questions that they generate. Even though critique is therefore different from therapy (the latter which hopes to unburden oneself of troubling questions altogether), it can appear that Kant is aiming at a kind of quarantining of their troublesomeness.  

Such a reading is implausible however. Leszek Kolakowski once remarked that no thinker before or after Kant has exercised so much energy or ingenuity in showing that their own most deeply-held convictions are entirely incapable of theoretical substantiation. It is part of Kant’s peculiar frame of mind that it was his steadfast personal commitment to these notions that drove him to remove them from the domain

56 For overviews of Kant’s strategy in the Transcendental Dialectic, see (Allison 2004; Grier 2001, 2010; Bennett 1974).

57 (Kolakowski 1978, 47–8).
of theoretical proof or disproof. Part of Kant’s motivation for distinguishing between a concept and its proper domain of application was just so that he could identify these concepts as ones whose proper referents must lie beyond the domain of possible human experience. Kant’s commitment was rather to the thoughts (a) that these concepts are ineliminable from the human conceptual repertoire, (b) that they were incapable of theoretical proof or disproof, and (c) that adopting a non-theoretical yet rational stance towards them is the appropriate attitude for a human being (Bxxix-xcx). Kant’s famous claim that he wished to deny knowledge in order to make room for faith [Glaube] (Bxxvii) explicitly concerns the thought that there is – indeed must be – a positive non-theoretical function for these concepts that manifests just because we are committed to them yet cannot explain them with theoretical resources.

Kant’s approach in the Transcendental Dialectic then is to attack reason so as to save it. The suspicion had long been sowed by Montaigne and Bayle, and then tended by Hume and cultivated by Rousseau, that human rationality was inherently malformed, and constitutively directed towards illusion, misdirection, self-deception and self-contradiction. While he came to acknowledge the force of the suspicion, it was not tolerable to Kant’s providential perspective on human existence that there could be rational outputs of the human mind with such a fundamentally negative effect. An essential aspect of Kant’s frame of mind – and perhaps its greatest weakness – regards his default confidence in the existence of some proper truth-conducive function for every constitutive aspect of the human mind. (Kant 1996, xxix) Nevertheless, without a resolution of the contradictions and an explanation of the origin of the very appearance of these deceptive elements in the first place, we would be left with a ‘just suspicion’ of reason itself (Axi- note).

Kant’s ingenious response was not to deny that these phenomena are products of rationality itself but rather to characterize them as misjudged over-applications of reason. The errors of reason are not entirely due to a constitutional tendency of reason (such that one might begin to suspect the faculty in itself) but are properly characterised as misuses of a reliable faculty beyond its proper domain of application. In generating illusions and contradictions, human beings reason unwisely because they reason too much. The proposal can be made because a criterion for what ‘too much’ reasoning consists in has been made possible by the distinction drawn between a capacity and its proper domain of application in the preceding portion of the First Critique. It was shown there that rational capacities acting in isolation cannot secure cognition, but instead require the cooperation of some distinct non-rational capacity, i.e. sensible intuition, to make a separate contribution. This non-rational contribution now performs an equally valuable negative contribution however, since it offers a general constraint on concept-use. Those concepts to which no possible spatiotemporal intuition can be provided are concepts with which we can never secure reference and with which we can never generate cognition. The spatiotemporal domain therefore sets the limits of the preeminent operation of rationality within which our representations can have ‘sense and significance’ (B148-9, A156/B195, A240/B299, A252/B308, A287/B343). Metaphysical speculation that fails to respect this constraint can only ever produce what

59 To say that a concept lacks sense and significance is not to say that it lacks all meaning whatsoever however (as we shall see); it merely connotes that the concept lacks a proper place within discourse regarding what can be objectively represented.
he calls ‘transcendental illusion’ (A295/B352) and a body of *faux* metaphysical knowledge he (somewhat confusingly) entitles ‘transcendent’ metaphysics (A295-6/B352-3, A703/B731, A845/B873).

Kant’s most famous attack on the pretensions of human reason to secure metaphysical results independently concerns the proof of the existence of God in the section on the Transcendental Ideal (A583/B61/ff). Kant had long held to the view that a fundamental problem of rationalism in general was the assumption that a conclusion that something exists might be established by arguing from premises regarding what reason *must* think (A340/B398). Unless reason is something that can actually bring things into existence by virtue of its demands – which it is not – then no such strategy could ever succeed, Kant holds. ‘Existence’ is simply not the *kind* of thing amenable to rational establishment. Since what reason *can* establish is that various concepts are contained in others or not, such that they can be stated as the predicates in analytic judgments, Kant’s view is pithily caught in his claim that existence is fundamentally not a real predicate in the required sense (A598/B626). For claims that something exists we require the testimony of some additional non-rational representational resource, such as intuition. God is however not plausibly the kind of thing that manifests itself in terms of spatiotemporal intuition. Since all attempted proofs God’s existence must proceed with some appeal to what can be rationally determined to exist, this analysis has had a lasting undercutting effect on the very idea of such a *verum*. The negative outcome is more famous than the positive one that Kant surely intended the ordinary reader to appreciate, which is that he had made philosophically respectable how faith in the face of incomprehension might be the appropriate mental attitude to bear to a divine being, and to explain how those who do not see yet believe might be more blessed on account of their attitude in the face of a lack of evidence.

When Kant states in the Canon of Pure Reason of securing two ‘articles of faith’, namely those of the immortality of the soul and the existence of God (A830/B859), he is making an appeal to Rousseau’s terminology employed in *Emile*, one many of his readers would have easily recognized. There the Savoyard Vicar professed that philosophy was entirely incapable of proving the existence of God, freedom or the soul, but that these concepts gained in their dignity and importance for the human being by virtue of being removed from the domain of theoretical speculation. The impact of Kant’s encounter with Rousseau cannot be overstated then, since the *Critique of Pure Reason* is crucially directed towards providing a philosophical grounding for Rousseau’s avowedly anti-philosophical results. Rousseau had reached these results by way of rhetoric that undermined the importance of rationality in guiding human affairs; Kant’s aim was to achieve the same ends without sacrificing the authority of rationality.

For all this, Kant perhaps still leaves us with a sense in which reason is inherently misleading. Although the misuse of the ideas of reason is due to our failing to heed the distinction between a concept and its domain of application, it is nevertheless the case that reason *naturally* produces conceptual representations that cannot receive a spatiotemporal instantiation. Thus we are naturally constituted to think in accordance

---

60 See also the *Only Possible Proof* in (Kant 1992).
61 Kant raises the challenge with regard to the Ontological Proof, which he believes to underpin all other possible proofs (A607/B635, A629/B657).
62 This view would have been of course theologically orthodox, cf. John 20: 29.
63 (Rousseau 1762/1979, 273 ff.)
with ideas that we cannot validate. Even if one came to accept that this is the case, and came to adopt defensive strategies for avoiding being misled by one’s own reason, the diagnosis threatens to frame the human being as one whose rational capacities are in a certain sense profoundly inapt for ordering human lives.

6. Freedom and Morality

Kant attempts to vindicate the ideas of reason in two ways. Firstly he claims that, despite lacking that paradigmatic application relating to cognition, the ideas of reason nevertheless have an important application within our theoretical activities. He distinguishes between ‘constitutive’ and ‘regulative’ notions. The former are the Categories, the concepts that constitute the very possibility of our experience of the empirical world, and which are deeply connected to the possibility of cognition. The latter concepts, such as the ideas of reason, are those that do not perform the function of making spatiotemporal representation itself possible, but instead play an indispensable role in ordering and regulating the systematic co-ordination of our knowledge claims (A616/B644, A619/B647, A642/B670ff). For example the idea of the soul, though lacking a particular instance in which it might be instantiated so as to yield cognition, nevertheless plays a necessary place-holding role in setting the topic of the diverse scientific inquiries known under the branch of psychology (A672/B700).

Secondly, and more importantly, Kant argues that the theoretical elusiveness of these notions is a necessary condition of our recognizing their practical relevance and urgency. The second response is most clearly displayed with regard to his account of freedom. Kant’s negative attack on the possibility of proving our freedom theoretically occurs in the Third Antinomy (A445/B473ff). If, as F. Scott Fitzgerald maintained, the test of a first-rate intelligence is the ability to hold two opposed ideas in the mind at the same time then the Antinomies might seem deliberately structured to test one’s mental capacities. Remarkably, Kant here shows how proofs of the existence of at least one uncaused physical cause are available as well as proofs that all events are thoroughly determined by empirical causes. Not only this, but each proof begins with the conclusion from the other proof. Thus Kant tries to show that the possibility of freedom can be deduced from the premise that everything is determined, and vice versa. The resolution of what would be a proof of reason’s inherent contradictoriness is provided of course by the metaphysics of transcendental idealism, which allow for the thought that determinism might be true of the phenomenal domain, while a power to bring about events by the will alone is made possible by locating that will (qua will-in-itself) in the noumenal domain (A532/B560ff).

Even if the arguments are deemed successful, they are by their very nature limited in their conclusions. The possibility of freedom requires not just the negotiation of the apparent truth of determinism but also an account of how something belonging to an entirely different ontological domain might intervene in the happenings of the physical domain such that an action might be plausibly explained both in terms of its phenomenal cause or its noumenal ground. A host of other problems arise, not least the question of how such beginnings of action can come about at all, since the domain in which they emerge is, Kant has insisted, an inherently timeless one. Kant’s ambition is to make

64 For analysis see (Allison 1990; Di Giovanni 2005; Ameriks 2000a; Paul Guyer 2005; R. L. Velkley 1989; Wood 2010, 1984)

65 For some discussion see (Freyenhagen 2008).
some minimal conceptual space for freedom while conceding that its actuality and the structure of its operations are ineliminable mysteries.

The requirement of such a space for freedom is crucial. All the ideas of reason, he claims, are necessary presuppositions for the possibility of moral life. If the soul were merely a material and unfree thing, if there were no supreme being, Kant claims, then moral ideas would ‘lose all their validity’ (A468/B496). In the Critique Kant alludes to the ubiquitous commitment we have to the idea that the moral ‘ought’ finds some real place in our cognitive economy (A547/B575, A550/B578, A633/B661, A806/B834). Yet since that representation finds no grounding in empirical nature, we are driven not to reject it, but instead to insist on the mere non-contradictoriness of the idea that such a representation finds a ground in another domain.

Kant had previously recognized the threat the failure of proofs of the existence of God, the soul and freedom might pose to the supposedly truth-conducive character of reason; he discovered in transcendental idealism a means not of securing those proofs but of rationally explaining their necessary failure. In so doing he recasts that failure as offering a valuable aspect of human beings’ existential predicament when those concepts are positioned within an order of explanation other than of theoretical achievement (A800/B828). In the case of those theoretical propositions that underpin scientific explanation, the goal of the First Critique is to explain how it is that we can grasp their possibility and comprehensibility; in the case of those metaphysical propositions that underpin the practical domain, the goal is to explain how it is that we can appreciate how a concession of their fundamental incomprehensibility does not thereby undermine their possibility.

Kant explores these themes more fully in the Groundwork of the Metaphysics of Morals, where he begins with some strikingly similar claims from the theoretical philosophy.66 It is clear for instance that Kant thinks that morality consists in a priori cognition and that it is a common phenomenon, registered by agents without special education or sophisticated reflective awareness.67 It concerns truths that are universal in their scope in two senses: they are truths that prescribe rules regarding what must always be done without exception, and they prescribe rules that hold for every possible finite rational agent.68 In this way the practical project can seem like a Newtonian extension of a science of necessary rules to the domain of the ethical: whereas the former set out the laws of what must happen in the physical universe, the latter set ought the laws of what ought to happen.69 The latter laws are crucially different of course in that they allow of a failure condition, whereby it is possible that laws that are in one sense necessary while in another sense might not be obeyed.70

The focus upon laws explains Kant’s focus upon a single kind of response – the registering of obligations as one’s duty – as the paradigmatic moral phenomenon.71 In

---

66 For some overviews of Kant’s moral theory see (Sedgwick 2008; Wood 2008, 1999). For detailed analysis of the Groundwork see (Allison 2011; Timmermann 2007; Schonecker and Wood 2015).
67 Groundwork 4: 389 in (Kant 1996, 44).
68 Groundwork 4: 389 in (Kant 1996, 44-5)
69 Groundwork 4: 408 in (Kant 1996, 62).
70 Groundwork 4: 388 in (Kant 1996, 43).
71 Groundwork 4: 389 in (Kant 1996, 44).
registering duty, Kant claims, we do not recognize a distinct source of moral authority that sanctions the behaviour in question but rather recognize a certain normative force generated by the mere rational recognition of the law-like character [Gesetzmassigkeit] of what is under consideration.\(^7\) It is this recognition that motivates the so-called ‘categorical imperative test’, whereby the moral character of a proposed plan of action can be determined solely by consideration of this formal character and by whether it can be consistently represented as universally endorsed by all possible agents.\(^5\)

Kant strikingly argues then that morality itself can be identified with the demands of rationality without recourse to any subject-independent source of moral value. Rather he claims that valid moral laws must be capable as being viewed as issued solely from the agent’s own rational capacities.\(^7\) This autonomy constraint – that laws must be capable of being viewed as self-legislated in order to be viewed as valid and binding – is yet another Rousseauian inheritance, this time from the latter’s political philosophy.\(^7\) Here it is put to work to generate a further instance of the Copernican strategy, since once again it seems that for Kant all valid cognition of necessary truths must be viewed as issuing from the rational resources of the subject. It is this subject-centred axiology that motivates some of Kant’s most famous reflections upon the value of human beings as valuable ends in themselves. A fundamental tenet of modern liberal society finds its root in Kant’s thought that human beings cannot be treated as mere means to an end, and instead are inherently deserving of respect irrespective of their station or circumstances.\(^7\)

Having reached a definition of the supreme principle of morality as rational autonomy, Kant turns in the third and final section of the *Groundwork* to the task of showing its possibility. The reader is here treated to a cascade of disappointments. Firstly, it emerges that what will be explained is not the possibility of autonomy but that of freedom. The concept of autonomy is claimed to be ‘reciprocal’ with that of freedom, thus justifying the strategic substitution whereby the explanation of the latter is taken as sufficient for the possibility of the former.\(^7\) Secondly, it emerges that freedom’s reality cannot be proven or even comprehended.\(^7\) The negative result is supposed to be importantly offset by the recognition that human beings must nevertheless act ‘under the idea’ of freedom, under a condition that they must take their actions to be in principle capable of being determined by a free power of choice, irrespective of the accompanying negative claims.\(^7\) In this way, Kant claims, we ‘we can thus liberate ourselves from the burden that weighs upon theory’.\(^8\) The incomprehensibility of freedom would have been less surprising to the reader of the First *Critique* of course. That reader would see that the motivation for substituting autonomy for freedom as the explanatory basis of morality was made *just for the purpose of putting it in a domain that has the room for its metaphysical possibility.*

---

\(^7\) *Groundwork* 4: 402 in (Kant 1996, 56-7).

\(^3\) *Groundwork* 4: 402 in (Kant 1996, 56-7). For discussion see (O’Neill 1975; Wood 2008; Geiger 2010; Galvin 2009).

\(^4\) *Groundwork* 4: 432-3 in (Kant 1996, 82-3).

\(^5\) See (Rousseau 1762/2009).

\(^6\) *Groundwork* 4: 430-1 in (Kant 1996, 81).

\(^7\) See *Groundwork* 4: 446-7 in (Kant 1996, 94–5). For discussion see (Allison 1986).

\(^8\) *Groundwork* 4: 463 in (Kant 1996, 108).

\(^9\) *Groundwork* 4: 447-8 in (Kant 1996, 95–6).

\(^10\) *Groundwork*, 4: 448 in (Kant 1996, 96, note).
If Kant is correct that genuine moral norms require a genuine libertarian model of freedom then the possibility of morality might seem to demand a two-order metaphysics. It is to this end that transcendental idealism then reappears in dramatic fashion at the end of the *Groundwork*. It allows for the possibility, if not the comprehensibility, that morality is real and manifested solely by beings whose constitution is such that they straddle the phenomenal and noumenal domains. According to transcendental idealism, human beings are uniquely both rational intelligences and ‘members of the world of sense’.81 When they succeed in responding to the moral law, Kant claims, it is because they have allowed the laws of freedom to influence their wills through a rational subordination of their desires, the latter which reflect the other side of their existence. The failure condition for moral normativity is explained as occasions when the laws of desire are allowed to hold sway over what ought to be our ‘proper self’, i.e. the self-in-itself whose cognitive powers are responsible for the very possibility of experience of the empirical world.82

In keeping with the broad trend of the Critical project, Kant’s aim in the *Groundwork* is not to deny the profound limitation of human rationality but to come to a rational understanding of that limitation. This allows reason to understand itself and thereby to fulfill the traditional philosophical ambition of self-knowledge by making sense of those limitations as existentially important aspects of the human condition. The inquiry, Kant claims, begins and ends with common moral cognition.83 The *Groundwork* concludes that our ordinary first-order moral responses have been shown to be at least really possible. Furthermore, his Rousseauian commitment concerning the irrelevance of sophisticated metaphysical understanding to simple moral self-orientation has also been vindicated by the revelation that the ultimate grounds of moral explanation are an uncognizable noumenal matter. Kant presents a model that fails to offer comprehension of freedom but instead insists that we can ‘comprehend its incomprehensibility’ while simultaneously viewing ourselves as nonetheless practically committed to its reality.84 This, Kant insists, secures a result that in fact tallies appropriately with our everyday conceptions and commitments.

7. The Limits of Transcendental Idealism

Kant’s project of transcendental idealism is the zenith (or nadir) of the Enlightenment project of devising a rational metaphysics for the physical sciences. Part of the advantage of that physics was its promise in regimenting explanation itself and banishing appeal to occult concepts. A looming worry for any such approach was that it would emerge as proving too much. It might establish that spirits, supernatural forces, final causes, etc. would be removed from nature but so too would the self, freedom, moral value, and any hint of divine presence. Kant’s project is striking in that it affirms just this consequence—none of these latter notions, including those concepts that seem so intimate and important to human beings’ world-view do ultimately find a proper place in the natural physical order.

Kant’s strategy was to insist upon the existence of a second order wherein the referents

81 *Groundwork* 4: 462 in (Kant 1996, 107–8).
82 *Groundwork* 4: 457 in (Kant 1996, 104).
83 *Groundwork* 4: 392 in (Kant 1996, 47).
84 *Groundwork* 4: 463 in (Kant 1996, 108).
of those crucial concepts might be reasonably hoped, though never known, to reside (A805/B833). Kant’s metaontological perspective is highly idiosyncratic therefore: we have stronger epistemological relations to the less fundamental ontological order (the world of appearances) and weaker epistemological relations to the more fundamental ontological order (the world of things in themselves). The postulation of a timeless, unknowable and incomprehensible metaphysical order existing alongside the natural order and manifesting complex ontological relations to it might have seemed to many as an heroic salvaging of what needed to be salvaged; to scientific critics it might have seemed as a reductio of any attempt to salvage those latter concepts at all. To anti-scientistic critics it might have seemed as a reductio of any philosophical ambition that started where Kant’s had started, with the project of regimenting rationality by first appealing to its scientific form of expression and using that form to determine the scope, possibility, and value of any other expression of human mentality.

Even if one were to defend Kant’s ambitions, it must be acknowledged that the breathtaking inventiveness of transcendental idealism’s proposed reconciliation of these orders nevertheless comes at serious costs. A fundamental worry is of course how Kant can maintain such an important role for things in themselves while denying us the preferred form of epistemic access to them, and then what kind of epistemic warrant is available for Kant’s positive and negative claims about them. A related worry concerns how we are to think of the epistemic perspective of the transcendental inquirer: how the elucidation of timeless necessary truths about the structure of the mind is even possible, and of how the limits of human cognition might be perspicuously drawn without claiming epistemic powers greater than those Kant is supposedly attributing to human beings. In Strawson’s memorable description, Kant seems on occasion to draw the bounds of sense and also to traverse them. Similarly, one might query how the propositions that constitute the content of transcendental idealism can themselves even be expressed. The sentences in the Critique of Pure Reason do not appear to be analytic judgments, and if synthetic a priori then they seem to lack the required corresponding intuitions that Kant insists is required so as to grant them ‘sense and significance’, let alone necessary truth. If they lack sense and significance but possess some lesser kind of meaning, one might ask what compels us to accept a philosophy that cannot express itself in its own vocabulary.

Another anxiety is one that relates to all dualisms, namely the problem of explaining the connection and ultimate unity of that which has been bifurcated. Kant’s strategy for bringing the two orders into reconciliation is characteristically unorthodox. The first stage in the Kantian reconciliation involves recognizing the fundamental theoretical incompatibility of the phenomenal and noumenal orders. The second stage posits that first moment of acknowledging the theoretical incompatibility of the two orders as required for a second moment whereby one recognizes their practical compatibility. The driving of the reader to a recognition of the reality of an entirely non-theoretical yet nonetheless rational practical context of philosophical explanation is arguably the key agenda of the First Critique, nevertheless it can appear that the reader might feel compelled only so far as they already share the requisite practical commitments.

85 For discussion of the relevance of Kant’s question ‘What may I hope?’ to metaphysics see (Chignell 2017).
86 A clear evocation of associated worries can be found in the introduction in (Langton 1998).
87 (Strawson 1966, Preface).
A final worry is that Kant's rescue of both traditional metaphysics and scientific rationality threatens to undermine itself. One might have thought that metaphysics is the fundamental domain of discourse. Kant's strategy explicitly requires that transcendental philosophy thus now functions as the fundamental domain of discourse, since it is concerned with the conditions of the very possibility of domains of discourse in general – and as such is prior to and independent of all other domains, even those of science and metaphysics. Thus while it might serves the essential goal of vindicating traditionally conceived metaphysics, it does so only by reconceiving just what those metaphysical concepts are. In a crucial sense then it supplants metaphysics, since it now insists upon a distinct domain of philosophical reflection as the *prima philosophia*. If one accepts this move however, it is open to question whether traditional metaphysics and the scientific expression of rationality really are the proper primary objects of theoretical concern. Perhaps transcendental philosophy, since it is independent of any first-order commitments is beholden to none, and has limitless autonomy to decide what its proper self-conception can be. Kant's method for preserving the priority of metaphysics and the sciences within the hierarchy of human knowledge arguably ushered in the means to their relegation.

The result of concerns like these is a perennial worry that Kant's relentless mediating between opposed traditions results in a philosophy that is a cable of strands artificially bound together and constantly threatening to unravel in various different directions. While there are many who have found deep inspiration in his writings, Kant has the peculiar fate of being the philosopher who inspired successors (often with the refrain 'back to Kant!') to each identify different strands of his thought as representing *the* key insight. Each recuperation declares the other strands to be superfluous and misguided, and to conclude that only in developing *their* single chosen strand can the true Kantian insight be brought to its natural completion. Whatever the merits of such strategies, it must be borne in mind that they are each premised on the thought that whatever property of Kant's project mark it out as valuable, it is not the synthesizing vision that he himself took to be his crucial contribution to philosophy.

At the risk of utilizing a hackneyed metaphor, it would seem that the advocates of transcendental idealism – admittedly few in number at this time of writing – would secure for themselves all the questionable delights of tightrope walking. There is (despite what generations of students might nearly unanimously claim) a sense of thrill from grasping both the details and the scope of the project. It is hard to resist the niggling notion that the main reason for engaging upon the project is the constant awareness that really one ought to fall on one side or another of a nearly inscrutable line that is claimed to support the entire endeavour. There is inevitably the dismissive claim that the inherent instability of the project is somehow a key part of its appeal. For all this, those attracted to Kant's thought might claim that such responses are unavoidable with regard to a project where a synthesis of opposing viewpoints is attempted with a vision and virtuosity unmatched in the history of philosophy.

**Bibliography**


