

Kant's First Paralogism

Ian Proops

University of Texas at Austin

Introduction

Kant pursues a variety of projects in the “Dialectic” of the *Critique of Pure Reason*; but chief among them is the program of “transcendental criticism”: pure reason’s systematic detection, correction, and explanation of its own excesses.¹ These excesses occur, according to Kant, because reason constantly strives to transcend the bounds of sense in its pursuit of speculative knowledge. In Kant’s view, all such attempts to “rise above the world of sense through the mere might of speculation” (A 591/B 619) are destined to fail.² In particular, pure speculative reason can afford us

My thanks to Omri Boehm, Jim Hankinson, Ivan Heyman, Martin Lin, Dustin Locke, Christopher Peacocke, Michael Della Rocca, Anat Schechtman, Nicole Smith, David Sosa, and Zoltán Szabó. Special thanks are due to Kenneth Winkler, who read several drafts of this essay and provided astute and generous comments on each draft. One ancestor of this essay was presented at the University of Texas in November 2007; another was discussed at a lunchtime meeting of the philosophy faculty at Yale University in the spring of 2008. I am grateful to those present on each occasion for their many helpful and thought-provoking questions. My thanks also to two anonymous referees for this journal who provided indispensable comments whose influence appears on almost every page.

1. Kant’s other goals include the promulgation of his doctrine of the regulative use of the Ideas of Reason (A 508/B 536ff.) and the rehearsal of his indirect argument (in the Antinomies chapter) for Transcendental Idealism (A 490/B 518ff.).

2. For passages from the first *Critique*, I have usually followed the translation by Paul Guyer and Allen Wood (1998). The points on which I have offered my own translation, when it concerns a substantive rather than merely stylistic alteration, are indicated by interpolating the relevant German word or phrase and by adding “translation my own” in parentheses. The basis for these translations is Kant 1902. Passages from this work are cited in the text by the relevant volume and page number. When I mark a translation as “my own,” it often departs only in minor respects from the Guyer-Wood translation,

no knowledge of the nature of the self as it is in itself, none of the world considered as the totality of appearances, and none of the existence (or nonexistence³) of God. Instead, the content of our speculative knowledge is confined to those principles that must be true if experience is to be possible, and such principles—Kant supposes—speak only of objects of possible experience.⁴

Since the views Kant criticizes as “dogmatic” in the *Dialectic* had been propounded by some of the brightest minds of the Early Modern period, he owes his reader a suitably charitable explanation of how thinkers of the caliber of, for example, Descartes and Leibniz could have gone as far astray as he contends.⁵ Recognizing this obligation, Kant elaborates a novel and ingenious account of the seductiveness of dogmatic error. The rationalist philosophers go wrong, he maintains, not because they fall prey to the “artful invention” of sophists—and certainly not because of any blundering intellectual clumsiness on their part (A 298/B 354)—but simply because they succumb to a powerful illusion grounded in the very nature of human reason itself (A 298/B 354).⁶ Because he takes even the most acute of human minds to be subject to this illusion,⁷ Kant can claim not to be implausibly charging his opponents

and I do not mean, in those instances, to imply a lack of indebtedness to them for the remainder of my translation.

3. Kant believes that we will not be able to prove the *nonexistence* of God, say, by discovering a contradiction in the concept of God. This feature of his view is an aspect of his so-called doctrine of practical faith. If we are to be moral, we must believe God to exist and, a fortiori, to be possible. This point is clear from a reflection from 1783–84: “In moral theology it is enough to presuppose that it is still possible that there is a God, *and that no one can ever prove the non-being thereof*.” R 6236; 18:519–20, emphasis added. The surrounding context makes clear that by “moral theology” Kant means the body of beliefs that we must hold if we are to be motivated by the moral law.

4. See B 294; A 471/B 499; 11:344, and compare B 289; A 702–3/B 730–31; 4:336n.

5. Kant’s immediate targets in the *Paralogisms* include both Leibniz himself and his followers—Wolff and Baumgarten in particular, and as Kant himself makes clear at B 413, on the issue of the immortality of the soul: Mendelssohn. Leibniz, as Margaret Wilson has argued, seems to be Kant’s target specifically in the second paralogism. Descartes, for his part, is explicitly identified as a target in the fourth paralogism. For details, see Longuenesse 2008, 20; and Wilson 1974.

6. In the case of Leibniz, this account of the source of metaphysical error is supplemented by a separate account, offered in the “Amphiboly to the Concepts of Reflection,” of how the empirical use of the understanding can be confused with one that seeks to go beyond experience. A 260/B 316–A 292/B 349.

7. The illusion, Kant says, “can fool even the most rational.” A 703/B 731.

with naïveté or sloppiness. Kant calls this supposedly pervasive illusion “transcendental illusion.”

In Kant's view, this distinctively intellectual species of illusion is not by itself sufficient to account for the existence of dogmatic metaphysics. It explains the pervasive and abiding *temptation* to dogmatism, but the actual commission of dogmatic error is explained only with the help of further resources. We form dogmatic views, Kant supposes, when, in the grip of transcendental illusion, we allow ourselves to be persuaded of the soundness of certain unsound arguments, which Kant terms “dialectical inferences of pure reason” (A 338/B 396). These arguments divide into three groups: “The paralogisms” are unsound arguments for various dogmatic theses about the nature of the self. “The antinomies”—or at least the “mathematical” subspecies of them⁸—are pairs of unsound arguments purporting to establish opposed doctrines about the world considered as the totality of appearances. Finally, the arguments of the “Ideal of Pure Reason” are unsound arguments that lead us dogmatically to assert the existence of God, conceived abstractly as the “*ens realissimum*”—a being that is “supremely perfect” in the sense of possessing the maximum possible degree of reality.

This account of Kant's diagnosis of dogmatic error acknowledges the importance of a distinction—originally drawn by Michelle Grier⁹—between the phenomenon of transcendental illusion, on the one hand, and the unsound arguments to which it disposes finite rational beings, on the other. The need to draw such a distinction is, to my knowledge, by now broadly accepted by Kant scholars. More controversial, and less well understood, are some of the details of how this diagnostic model applies to the specific doctrines criticized in the *Dialectic*. In particular, there is little agreement about how this model applies to Kant's criticisms of various rationalist doctrines about the self in the “Paralogisms” chapter.

The present essay seeks to solidify our understanding in this area by applying Kant's diagnostic model to the case of the self. Or, rather, it narrows its focus still further and seeks to apply this model to just one aspect of the Cartesian or Leibnizian dogmatist's theory of the self, namely, the view that the self is a “substance.” This task, I submit, is worth

8. In Kant's view, the third and fourth antinomies, which are known as the “dynamical” antinomies, do not comprise pairs of *unsound* arguments, for they are resolved by arguing that thesis and antithesis, taken the right way, can both be true. The chief point is that in these cases the thesis and antithesis only seem to be logically inconsistent because of the transcendentially realist assumptions encouraged by transcendental illusion.

9. I discuss Grier's work in notes 11 and 51 below. Citations are provided in note 19.

undertaking both for the light it sheds on Kant's conception of metaphysics as practiced by his immediate predecessors and for what it reveals about his conception of philosophical criticism. It should also be of interest for what it reveals about Kant's take on certain venerable philosophical notions, including the very notion of substance. We shall see that, in order to do justice to the richness of Kant's discussion, it will be necessary to distinguish no fewer than four rather different Kantian conceptions of substance.¹⁰

Our discussion recapitulates the bifurcation in Kant's diagnosis of dogmatism: its first part provides an account of how transcendental illusion creates a temptation to believe that the self is a substance, while its second offers an analysis of the fallacy that Kant supposes we commit when we succumb to this temptation. At higher resolution, the essay divides into seven sections and a conclusion. The first section discusses the phenomenon of transcendental illusion in general; the second asks how this illusion is supposed to generate in human beings an inclination to believe in the substantiality of the soul. The third asks what precisely Kant means by a "paralogism," while the fourth examines the B-edition's diagnosis of the formal fallacy involved in the first paralogism. The fifth section examines the rather different account of the first paralogism given in the A-edition and offers an account of why Kant came to see his earlier discussion as unsatisfactory. The penultimate section is devoted to defending my interpretation of the B-paralogism against possible objections, and the last to examining some further complications of the story told in the earlier sections. Along the way, I will identify and discuss a number of as yet unrecognized difficulties for Kant's account of the first paralogism and for his account of transcendental illusion in general. In spite of these problems, I hope to show that Kant's views on these matters are more promising—and, indeed, more philosophically illuminating—than many commentators have supposed. I will argue, in particular, that Kant's discussion of the B-edition's first paralogism yields an insight about the nature of conceivability that is of lasting value.

1. Transcendental Illusion

As Kant explains it, "transcendental illusion" is the illusion that what are in fact merely so many maxims or rules guiding empirical inquiry amount

10. The first three conceptions are discussed in the main body of the essay. The fourth is discussed in note 54.

to objective truths (compare A 297/B 353). These rules, which human reason provides for its own direction, are instances of a more general demand of reason, which Kant formulates thus:

- P*: Find for the conditioned cognitions of the understanding the unconditioned whereby its [that is, the understanding's] unity is completed (compare A 307/B 364).

Such a prescription enjoins us to seek, among other things, ever more fundamental explanations and ever earlier causes for the phenomena we encounter in experience.

Kant maintains that, owing to transcendental illusion, rational human beings are inclined to take this "logical prescription" (A 309/B 365) for the descriptive claim that:

- D*: If the conditioned is given, the whole series of conditions subordinated to one another, a series which is therefore itself unconditioned, is likewise given (compare A 307–8/B 364 and B 436).¹¹

When through a "misunderstanding" we succumb to transcendental illusion, the "need of reason" expressed by *P* is "taken for a transcendental principle of reason" that "over hastily postulates . . . an unlimited completeness in the series of conditions" (compare A 309/B 366).¹² In other words, the prescription, *P*, is taken for the putative statement of fact,

11. In seeing transcendental illusion as consisting—as a first approximation—in the illusion that *D* is true, I am in general agreement with Grier's illuminating account of this phenomenon in Grier 1993 and 2001. (My labels "*P*" [for "prescription"] and "*D*" [for "description"] are merely intended as perhaps more suggestive replacements for Grier's "*P1*" and "*P2*.") Although I am in broad agreement with much of what Grier says—and in particular on the general need to distinguish transcendental illusion from error, I disagree with her on certain points of detail. First, I would not endorse her characterization of *D* as something we must assume as a condition of "using" *P*. Grier 2001, 126. For, being a prescription, *P* is not something that can be used but only something that can be complied with or violated. On my account, the correct thing to say about the relationship between *P* and *D* is rather that the illusion that *D* is true is one that is inevitably generated in finite rational minds in virtue of their being subject to the demand of reason expressed by the prescription *P*. Second, as we shall see in what follows, I believe that in the Paralogisms it is not quite *D* itself but a closely related claim that expresses the content of the illusion (see section 7). Third, I differ from Grier in my interpretation of the argument form of the first paralogism (see section 6 below, and especially note 51).

12. Kant gives this explanation in the form of a *question* in the course of announcing the topic to be investigated. The question is lengthy and hard to parse, but we might paraphrase it thus: "Has the need of reason expressed by *P* been taken for the transcendental principle expressed by *D*?" It is clear from his subsequent discussion that, as far as traditional rationalist metaphysics goes, Kant's answer is "yes."

D. In Kant's jargon, a "regulative" principle is taken for a "constitutive" principle.¹³

Although Kant characterizes the "misunderstanding" (A 309/B 365) that leads to the fallacious "inferences of reason" in terms of our having "taken" the need of reason expressed by the prescription *P* for the truth of the principle *D* (A 309/B 366),¹⁴ the path from the former to the latter can be fruitfully described in more subtle terms. It seems plausible that, being a command of reason, *P* must be viewed as an *authoritative* prescription. It is also plausibly experienced as a *reasonable* prescription since reason's authority strikes those subject to it as reasonably exercised. But since the prescription, *P*, strikes us as reasonable and authoritative, it is extremely natural to suppose that what it enjoins us to seek is in fact there to be found. Thus our consciousness of the standing authority of *P* generates in us the permanent inclination to regard *D* as true. We may, therefore, characterize transcendental illusion simply as the illusion that *D* is true—an illusion that is *explained* by our being subject to the prescription *P*.

Or so, at least, runs a plausible interpretation of Kant's first-pass view. However, I shall argue in section 7 that Kant's more considered view is that it is in fact not *D* itself that states the content of transcendental illusion but rather the related proposition *D'*: "If the conditioned is given, the unconditioned is likewise given." However, because the reasons for this qualification are rather complicated, it will be useful, for the time being, to continue our exposition with *D*.¹⁵

The phraseology of *P* and *D* immediately raises a number of questions: Are the members of the series of conditions alluded to by *P* and *D* to be understood as *cognitions* (as the phrasing of *P* might seem to suggest), or are they rather to be thought of as the various states, events,

13. See A 619–20/B 647–48; A 647/B 675; A 686/B 713; A 690/B 718; and *Prolegomena*, sec. 56, 4:350.

14. Kant also implies that *instances* of *P* are apt to be taken for *instances* of *D*, for he says that the various specific fundamental maxims and rules contributed by human reason and governing its use "look entirely like objective principles." A 297/B 313.

15. In characterizing the governing illusion behind all of rationalist metaphysics as the illusion that *D* is true, Kant may seem to be lumping his predecessors together in a rather high-handed way. Is it really true that *all* of rational metaphysics flows from a single principle? Such qualms may, however, be allayed somewhat by Omri Boehm's observation (Boehm 2009) that *D* can be regarded as one formulation of the Principle of Sufficient Reason, a principle that does indeed have a claim to be the single substantive principle generating all rationalist metaphysical claims. This idea might also explain why Kant refers to *D* as "this supreme principle of pure reason." A 308/B 365.

or properties that are the *objects* of cognitions? What is it for a condition to be “given”? And, perhaps most baffling of all: What would it mean for the unconditioned to “complete the unity” of the understanding?

My interpretation answers the first two of these questions as follows: First, the “conditions” in question are not cognitions but objective states of the world (or, in some cases, events or properties). Thus in enjoining us to seek the unconditioned for the “conditioned cognitions of the understanding,” *P* is enjoining us always to seek ever more fundamental conditions for the *objects* of our cognitions. Second, to speak of a condition as being “given” is just to say that it exists (or obtains). I shall not attempt to argue for these claims here as the justification for them is just that they make possible a coherent and plausible interpretation both of transcendental illusion itself and of the way this illusion plays out in the Antinomies and the first paralogism. The third question does, however, merit a more detailed answer.

In order to understand the notion of “completing the unity of the understanding,” it helps to note two relatively obscure but still important details of the Leibnizian background. First, in Baumgarten’s *Metaphysica*—the Leibnizian work on which Kant usually based his metaphysics lectures—the term “condition” is treated merely as a variant of “ground” (see Baumgarten 1902, sec. 14). Kant seems to follow Baumgarten on this point, for he occasionally speaks indifferently of “grounds or conditions” (see, for example, *Real Progress*, 20:328). The second point to note is that, in Baumgarten’s view, grounds and conditions are precisely what one cites in answering “why”-questions (Baumgarten 1902, sec. 14). Kant shares this opinion (*Vienna Logic*, 24:921), but he takes issue with Baumgarten’s claim that one can *define* a ground as “that from which it can be cognized why something is.” Such a definition, Kant contends, would be circular because “why” just means “through which ground” (*Vienna Logic*, 24:921). Importantly, however, while he rejects Baumgarten’s attempted definition as circular, he still regards it as *true* (*Vienna Logic*, 24:921). So, for Kant, whenever one states a ground or condition, one thereby answers a “why”-question. Taken together, these observations suggest that Kant is thinking of the demand that we seek the unconditioned condition for something as the demand that we seek something which, if found, would answer the final answerable “why”-question in a series of such questions. If that is right, then to have found an unconditioned condition would be to possess the most *complete understanding* possible of the original phenomenon to be explained. In Kant’s language, one would have “completed the unity” of the understanding.

In Kant's view, we are inclined to regard *D* not just as *true* but as *necessarily true*. This circumstance, he maintains, is owed to our readiness to mistake the "subjective necessity" of "a certain connection of our concepts" for an "objective necessity, the determination of things in themselves" (A 297/B 353). Kant does not explain what this "subjective necessity" consists in; but we might plausibly think of it as the necessity involved in the following conditional injunction: "If you wish to proceed rationally in inquiry, you *must* seek, for the object of every cognition, the series constituting its unconditioned condition." The necessity involved here counts as "subjective" because the imperative is conditional upon our having a certain *desire*. Should one not wish to proceed rationally in inquiry, one will stand under no obligation of the kind specified in the conditional's consequent.

Kant supposes that, owing to transcendental illusion, this subjective necessity is apt to be misconstrued as the objective necessity invoked by the claim: "Necessarily, for the object of any cognition, the series constituting its unconditioned condition exists." The illusion that *D* holds of *necessity* is thus itself a component of transcendental illusion. The slide Kant envisions here may be made clearer by breaking it down into two transitions. Consider the following three statements:

- [A] If you desire to proceed rationally in inquiry, you *must* seek, for the object of any cognition, the series constituting its unconditioned condition.
- [B] You *must* seek, for the object of any cognition, the series constituting its unconditioned condition.
- [C] Necessarily, for the object of any cognition, the series constituting its unconditioned condition exists.

Claim [C] is just the claim that principle *D* (slightly reformulated) holds of necessity. As I read him, Kant supposes we are apt to confuse [A] with [B] simply because, as rational beings, we naturally desire to proceed rationally in inquiry. Because this desire is so natural and pervasive, its presence is easily overlooked, and we thus fail to appreciate that the necessity in [B] is merely the conditional necessity expressed in the consequent of [A]. Once we have mistaken [A] for [B], transcendental illusion inclines us to slide further to [C]: for we hear the injunction of [B] as an injunction of reason, and so regard it as rational. This in turn leads us to believe that what we are enjoined to seek actually exists. We thus misconstrue the "must" of a binding intellectual categorical imperative—

that is, [B]—as the “must” of an indicative claim that purports to express a necessary fact—that is, [C].

In support of this explanation, we may observe that Kant himself on one occasion draws the distinction between objective and subjective necessity in precisely these terms—albeit in a somewhat different connection. In his *Lectures on the Philosophical Doctrine of Religion*, which stem from the mid-1780s, Kant remarks regarding his attempted proof of the *ens realissimum* (or “most real being”) in his 1763 work *The Only Possible Basis for a Demonstration of the Existence of God* that this proof

is unable to establish the objective necessity of an original being; rather it establishes only the subjective necessity of accepting such a being. But this proof can in no way be refuted, because it has its ground in the nature of human reason. For my reason makes it absolutely necessary for me to accept a being which is the ground of everything possible, because otherwise I would be unable to cognize what in general the possibility of something consists in. (28:1034)¹⁶

Kant is warning us here not to take his 1763 proof to establish the *objective* necessity of the claim that an *ens realissimum* exists. Rather, it should be taken to establish the merely *subjective* necessity involved in the conditional claim: “If we want to cognize in what in general the possibility of something consists, then we must accept that the *ens realissimum* exists.” Having noted this parallel, we should concede that in the case of the hypothetical imperative governing inquiry, [A], what figures in the consequent of the conditional is not the notion of *accepting* a proposition but rather that of *seeking* the unconditioned. Nonetheless, this difference does not undermine the parallel between these passages insofar as it relates to Kant’s use of the terms “subjective” and “objective.”

Kant conceives of transcendental illusion as sharing three features with cases of ordinary perceptual—or “empirical” (A 295/B 351)—illusion. First, it is *natural* and *unavoidable* (*unvermeidlich*) (A 298/B 354): just as anyone who possesses a properly functioning visual faculty will, Kant supposes, be subject to the moon illusion;¹⁷ so anyone who possesses a properly functioning faculty of reason will experience transcendental illusion. Presumably because he wishes to exempt a possible divine reason

16. This passage is cited in a fascinating essay by Andrew Chignell, 2007, 349. I am indebted to him for raising it to prominence, and I follow his translation.

17. The “moon illusion” refers to the illusion that the moon is larger on the horizon than it is at its apex. Kant assumes that such optical illusions are universal (A 297/B 354); whether that is in fact so is, of course, a question for empirical psychology.

from transcendental illusion, Kant tends to characterize this kind of illusion as inevitable for *human* reason. Such an exemption would be well judged since a divine rational faculty, should it exist, would need to be conceived of as immune from transcendental illusion on pain of compromising the Divinity's perfection. (Sometimes in what follows the qualification "human" will be omitted, but only for stylistic reasons.)

The second feature transcendental illusion shares with perceptual illusion is that its usual consequences are *correctible*:¹⁸ just as one need not be "taken in" by an optical illusion in the sense of actually forming an erroneous judgment or belief on its basis, so one need not be led into actual error by transcendental illusion. As Grier has emphasized,¹⁹ this distinction between unavoidable illusion, on the one hand, and avoidable error, on the other, is essential for understanding Kant's critical enterprise: without it Kant would seem to be making the manifestly implausible claim that the unsound arguments themselves are inevitable. Thus, rather than urging us to the impossible feat of avoiding inevitable sophistry, Kant is aiming to help us to avoid unsound arguments, even while insisting that the illusion that inclines us toward finding them compelling cannot be removed.

This brings us to a third feature that transcendental illusion shares with empirical illusion, namely: its *persistence*. Transcendental illusion "does not cease even though it is uncovered and its nullity clearly seen into by transcendental criticism" (A 297/B 353). In other words, transcendental illusion survives detection and even explanation. Owing to transcendental illusion's persistence, human reason is continually propelled into "momentary aberrations that always need to be removed" (A 298/B 355). Accordingly, the work of transcendental criticism is never complete.

Kant's account of transcendental illusion is clear; but how does it play out in the case of a particular "transcendental idea"? This is easy to see in the case of the transcendental idea of the world, which is associated with certain arguments—the antinomies—that purport to establish conclusions about the world as a whole. In the first antinomy, for example,

18. I avoid saying that the illusion itself is correctible since this might create the misleading impression that the illusion is capable of being neutralized by a counterillusion, in the way that, for instance, some forms of "corrective" lens provide a compensating counterillusion. I also avoid saying that the illusion is "resistible" since we cannot get rid of it—a point, indeed, that Kant puts by saying that it is "irresistible" [*unwiderstehlichen*]. A 642/B 670.

19. See Grier 1993, 263–64 and Grier 2001, 116 and 128ff.

transcendental illusion tempts us to embrace both the belief that the world must have had a beginning (the thesis²⁰) and also the apparently contradictory belief that the world cannot have had a beginning (the antithesis).²¹ And, indeed, it tempts us to view these beliefs as amounting to rationally grounded knowledge. In the case of the thesis, the whole regressive series of “conditions”—in this context, ever earlier states of the world—is taken to terminate in an earliest “condition” of the series, which is itself unconditioned (compare A 418/B 446). In the case of the antithesis, on the other hand, the regress is taken not to terminate, but the whole infinite series without its first member—namely, the present state of the world—is taken to be the unconditioned condition of that first member (compare A 418/B 446, including footnote “*”). The first antinomy arises because we thus have—or, rather, imagine we have—two incompatible ways of finding the completeness that reason enjoins us to seek in the series of conditions. Dogmatic theorizing about the age (and size) of the world will thus lead to unending controversies since its practitioners will divide between these apparently incompatible alternatives according to their individual proclivities—including their taste, or lack of taste, for theism.²²

20. Or, strictly speaking, *part* of it, since the antinomy also concerns the question whether the world is bounded in space.

21. Kant's solution to the first two antinomies consists in arguing that thesis and antithesis are contraries rather than contradictories and that both are false.

22. Kant's account of what I have called these “proclivities” is in fact rather involved, and it has an interesting philosophical motivation. He supposes that if we could consider the thesis and antithesis positions purely on their intrinsic merits—that is, according to their “grounds” (A 475/B 503), we would be led to an “unceasing vacillation.” Ibid. In fact, Kant supposes that we would flip-flop between thesis and antithesis on a daily basis. Ibid. In practice, however, individuals who consider the antinomial questions tend to form themselves into relatively stable camps or factions. In one corner, favoring the thesis position, are the “dogmatists” (compare A 466/B 494) or advocates of “Platonism” (A 471/B 499). In the other, favoring the antithesis, are the “empiricists” (A 468/B 496) or champions of “Epicureanism” (A 471/B 499). In a broader sense, of course, both camps are “dogmatic,” and it is clear that Kant views his labels for these positions as loose and possibly historically inaccurate. A 471/B 499.

Although Kant doesn't himself explicitly mention the point, this apparent mismatch between Kant's theory of transcendental illusion, which predicts extreme doctrinal instability, and observation, which reveals a measure of stability, constitutes a *prima facie* difficulty for his theory. To address it, Kant offers a supplementary explanation: he suggests that in practice certain factors unrelated to the intrinsic merits of the claims influence our choice between them. These factors, which Kant terms “interests of reason,” are as follows: First, there is a certain “practical” interest of reason, namely, the need to employ the four thesis positions as “cornerstones” of religion and morality. A 466/B 494. Second,

This is only the briefest sketch of a good deal of complex material, but it does, I would claim, capture the general contours of Kant's application of the model of transcendental illusion to the case of the Antinomies. Unfortunately, it takes considerably more effort to see how this model might apply to the case of the self as it is discussed in the Paralogisms. It is to this question that we now turn.

2. Transcendental Illusion and the Self

The proponent of "rational psychology"—the rationalist's putative science of the self—presumes that we can know the character of the self as a simple, incorruptible, immaterial, and naturally immortal substance (A 345/B 403, A 349)—a "soul" as conceived by the Christian tradition. This traditional conception of the soul should not be confused with Kant's own more abstract or formal conception, according to which a soul (*Seele*) is just "a thinking being" (*ein denkend Wesen*) (A 348)—where the notion of a "being" is maximally thin and does not, for example, entail substantiality.²³ Kant maintains that transcendental illusion explains the prevalence of the belief in the richer, traditional conception of the soul because it unceasingly tempts us to draw certain fallacious inferences about the character of this "thinking being." In Kant's view, however, the

there is the interest of "popularity," which favors the thesis because—Kant supposes—the common understanding has no scruples about the idea of an unconditioned. A 467/B 495. One supposes that Kant would view this lack of scruples as owed, at least in part, to the ordinary person's ignorance of the Principle of Sufficient Reason. Third, there is a certain "architectonic interest,"—a kind of penchant for system building and foundationalism—which favors the thesis of each antinomy. A 474–75/B 503. A fourth "interest of reason" does not so much stabilize the oscillation as fuel it. This is the "speculative" interest of reason, which can work in favor of either the thesis or the antithesis. A 466–67/B 494–95. Kant is not explicit about the nature of this interest, but he may well have in mind the demand of reason expressed by *P*, which, as we have seen, can be satisfied in either of two incompatible ways. (I am grateful to an anonymous referee for this suggestion.) Notice that since each of the stabilizing interests is supposed to favor the thesis position, Kant is still left with a residual problem of explaining the stability of the antithesis position.

23. Kant uses a range of expressions for his own, much thinner, notion of the soul in the first *Critique*. These include: "the soul" [*die Seele*] (A 351; A 361); "the thinking I" [*das denkende Ich*] (A 351; A 361; A 401); "the thinking self" [*das denkende Selbst*] (B 430), and "[the] transcendental subject of thoughts." A 346/B 404. At A 361 he suggests that the first of these expressions refers to "the transcendental object of inner sense," which presumably means the self that appears to itself in inner sense as "the empirical self." Kant's remarks at B 429 are consistent with taking "the thinking self" at B 430 to have the same reference. Thus the thinking I or thinking self is both the transcendental subject of *thoughts* and the transcendental object of inner *sense*.

thinking being that we refer to by means of the word "I" is something about whose inner nature we must remain forever ignorant. (In what follows I will use the terms "the self" and "the soul" indifferently when reporting Kant's views about this "thinking being.")

In the case of the B-edition's first paralogism, transcendental illusion is supposed to encourage the belief that the self is a substance in the sense of "something that can exist only as **subject**, not as a mere determination of other things." (B 288; compare B 149, B 289). A substance in this sense is something that cannot *inhere* in another thing as a determination of it. And by a "determination" in this context Kant means something that is either a property or a "mode." (The latter is an entity whose existence consists in the exemplification of a property or relation by another thing [or other things]. Common examples include dents and fists.²⁴) Kant also emphasizes that it is something that *need not* so inhere because it is a "self-**subsisting being**" (B 407; compare B 412, B 149). In the B-edition, Kant intends to be speaking of substance in this sense whenever he invokes the "unschematized" category of substance; but in both editions he also appeals to a distinct concept of substance corresponding to the "schematized" category. This is the concept of something that "persists" (A 144/B 183, A 182) in the sense of existing at every moment (A 182). To keep ideas straight, I will follow Jonathan Bennett's convention of referring to the first of these notions as "substance₁" and to the second as "substance₂" (Bennett 1966, 182ff.). Even though Kant uses the same term, "*Substanz*," for both notions, he is good at not confusing them.

Kant is clear that we cannot know whether or not there are any substances₁—or even whether or not their existence is a real possibility:

[In the "General Note on the System of Principles" (B 288–B 294) and the chapter on "*Phenomena* and *Noumena*" (A 235–60/B 294–B 315)] it was proved that the concept of a thing that can exist for itself as subject but not as mere predicate [*als bloßes Prädikat*] carries with it no objective reality at all, that is, that one cannot know whether it can apply [*zukommen könne*] to any object, since one has no insight into the possibility of such a way of existing, and consequently that it yields absolutely no cognition. (B 412, translation my own).

Because the concept of a substance₁ contains a mark of necessity (namely, "cannot occur as a predicate"), which no experience can yield, we are

24. For this understanding of the notions of "determination" and "mode" I am indebted to James van Cleve's discussion in van Cleve 1999, 105.

incapable of having an intuition of a substance₁, and so cannot thereby know the concept of substance₁ to be capable of having instances—that is, to be “objectively real.” Nor can we know it to have instances in the way that, according to Kant, we can (and do) know that (say) the concept of cause has instances; for being an *unschematized* category, the concept of substance₁ will figure in no a priori synthetic “principle of experience” that could guarantee its actuality. For these reasons, we cannot know there is not some unforeseen “real repugnance” between the determinations represented by the various component marks of the concept of substance₁, and thus we cannot know whether or not a substance₁ is a real possibility. Nor, as matters stand, do we have any reason to suppose that substance₁ does *not* exist, for its concept seems (as far as we can tell) to be *logically* consistent. To avoid dogmatism, then, we must be agnostic about the existence or nonexistence of any substance₁—and yet transcendental illusion tempts us to assert, dogmatically, that there is at least one.²⁵

In order to apply Kant’s model of transcendental illusion to the case of substance₁, we shall need to arrive at a plausible specification of the series of conditions that reason is, in this case, supposed to enjoin us to complete. We saw that, in the case of the world’s existence in time, the relevant series is composed of temporally ordered states of the world. In the present case, a very different series must be under consideration. But of what kind of things would it be composed?

In the first *Critique* Kant’s view on this matter is hard to discern, but in the *Prolegomena* it is spelled out clearly:

Pure reason demands that for each predicate of a thing we should seek its own [*zugehöriges*] subject, but that for this subject, which is in turn necessarily only a predicate, we should seek its subject again, and so forth to infinity (or as far as we get). But from this it follows that we should take nothing that we can attain for a final [*letztes*] subject, and that the substantial itself could never be thought²⁶ by our ever-so-deeply

25. I owe a debt of gratitude to an anonymous referee for much of the substance of this paragraph. James van Cleve (1999, 173) attributes to Kant the view that the A-edition first paralogism succeeds in establishing that the self is a substance₁. However, as will shortly become apparent, this is to attribute to Kant a position he would have deemed unacceptably dogmatic in *both* editions of the first *Critique*.

26. Here one expects “cognized” or “known” rather than “thought.” Fortunately, however, a case can be made that Kant sometimes uses “thought” where he means “cognized” or “known.” Consider, for example, a remark from the “amphiboly of the concepts of reflection”: “If by merely intelligible objects we understand those things that are thought

penetrating understanding, even if the whole of nature were laid bare before it. (*Prolegomena*, sec. 46; 4:333, translation my own)

The relation of predicate to subject generates a series in which each subject can be conceived of as (really) a predicate of a further subject that conditions or grounds it. Kant, moreover, is clear that the regressive series of conditions in question here is supposed to stand on all fours with the various regressive series comprehended under the “cosmological idea” (that is, the idea of the world). He makes this point in passing when, as it happens, denying that the “theological idea” generates a similar series:

[In the case of the theological idea] reason does not, *as with the psychological and the cosmological idea, start from experience and become seduced by the ascending sequence of grounds into aspiring, if possible, to absolute completeness in their series*, but instead breaks off entirely from experience and descends from bare concepts of what would constitute the absolute completeness of a thing in general . . . to the determination of the possibility, hence the reality, of all other things. (*Prolegomena*, sec. 55, 4:348, emphasis added)

The passage quoted from *Prolegomena*, sec. 46 makes clear what kind of regressive or “ascending” series Kant has in mind in connection with the self: it is a series of ever more fundamental *subjects* (that is, bearers of properties).²⁷ However, while helpful, the passage is also potentially misleading. It can create the impression that Kant means to be discussing a demand of reason that might be expressed in the injunction: “For each subject, seek another in which the first inheres as a predicate.” But although such a picture is certainly suggested by Kant’s words, it cannot be quite what he intends. For this injunction implies that we should never rest with the idea that we’ve found an ultimate subject. In consequence, such an injunction would tend to generate in minds subject to transcendental illusion the belief that there is *no* ultimate subject, and

through pure categories, without any schema of sensibility, then things of this sort are impossible.” A 286/B 342. If we took this remark at face value, it would sit awkwardly with Kant’s view that the concept of a noumenon, taken merely problematically, is an unavoidable concept that limits sensibility. A 256/B 311. For that seems to imply that noumena are at least conceivable. Interestingly, however, Kant altered his own copy of the first edition to read not “are thought” but “are cognized by us” (see A 286/B 342, Erdmann 1881, remark 150, 46; 23:49); but, having done so, he then neglected to incorporate this correction into the second edition. Since that is so, we may suppose that *occasionally* when he uses “thought” in a published work, he means “cognized.”

27. Here we need to keep in mind that accidents—which Kant usually refers to simply as “determinations”—can be considered “subjects” relative to further, less fundamental, accidents that inhere in them.

hence that nothing—not even the self—is a substance₁. But such a belief is obviously not Kant’s target in the first paralogism; for he means to be criticizing and diagnosing our belief that there is at least one substance₁, namely, the self.

A better story, which still preserves the intended parallel with the cosmological idea, might run as follows. For each subject we encounter, we are enjoined to seek some ultimate subject in which the associated regressive series of subjects might be thought to terminate. Accordingly, we must try to construe each subject we encounter as in turn a “predicate” (that is, property or mode) of something else, hoping as we do so that at some stage our attempt will be unsuccessful precisely because we will have arrived at an *ultimate subject*. In obeying the injunction to proceed in this manner we will of course be acting as if we are trying to show that there are no ultimate subjects; but we will be doing so only as a means to discovering a subject that resists our efforts to construe it as a predicate of something else. Our *ultimate* goal, then, will be to discover an ultimate or “absolute” subject. Understanding the injunction in this way yields the satisfactory result that the dogmatic belief generated when we are taken in by the relevant instance of transcendental illusion will be the belief in the existence of some ultimate subject—a “substance₁” in Bennett’s jargon.

Now, although Kant is certainly guilty of some carelessness here, I think he does, nonetheless, appreciate that this is how his account ought to run. For the discussion in the *Prolegomena* continues in a way that agrees with this account:

Now it does appear as if we have something substantial in the consciousness of ourselves (i.e., in the thinking subject), and indeed have it in an immediate intuition; for all the predicates of inner sense are referred to the **I** as subject, and this cannot again be thought as the predicate of some other subject. It therefore appears that in this case completeness in the referring of the given concepts to a subject as predicates is not a mere idea, but that the object, namely, the **absolute subject** itself, is given in experience. But this expectation is disappointed. (*Prolegomena*, sec. 46, 4:334, translation my own)

Here Kant is suggesting that the injunction to seek an ultimate subject generates the belief that there is such a thing in the case of at least one series of subjects. In the case of the series apparently terminating in *the self*, we seem to have a subject that cannot be conceived of as a predicate of another thing. But this appearance, he adds, gives rise to an

expectation that is “disappointed”: we expect to encounter a substance₁ in introspection, but no intuition of such a thing is forthcoming. Kant is thus concluding this remark with a Humean point about the failure of introspection to reveal a self or soul with the properties—of simplicity, endurance, substantiality, and so forth—that are commonly attributed to it by rationalist metaphysicians.

The theory of transcendental illusion predicts that, other things being equal, each of us (insofar as we are rational) will be inclined to believe that there are substances₁; but Kant supposes that we shall actually succumb to this inclination only when a plausible yet fallacious argument accompanies this illusion. In the case of our belief that the self is a substance₁, this argument is the first paralogism. But what exactly is a “paralogism”? And why does Kant treat the paralogisms as a distinctive class of “dialectical inferences”?

3. The Notion of a Paralogism

It is a simple matter to describe what the arguments grouped under the heading of “The Paralogisms of Pure Reason” are supposed to have in common: they are, one and all, supposed to be sophistical categorical syllogisms (A 405–6/B 432–33), whose status as sophistical stems from their being “false” due to their “form” (A 341/B 399). Each, moreover, is supposed to be a “transcendental” rather than a “logical” paralogism, where this means that there is “a transcendental ground for inferring falsely due to [the argument’s] form” (A 341/B 399). This “transcendental ground” is, presumably, transcendental illusion.²⁸

It is, however, harder to say what exactly Kant *means* by the term “paralogism”—and harder still to answer this question in a way that respects his idea that the Paralogisms constitute a distinctive class of

28. Although Kant unquestionably maintains that the dialectical inferences are grounded in transcendental illusion, he also sometimes says that they *ground* the illusion. For example, in the very first sentence of the Antinomies chapter, Kant announces that his goal is to show that every “transcendental illusion of pure reason *rests on* dialectical inferences” [*auf dialektischen Schlüssen beruhe*]. A 405/B 432, emphasis added; compare A 606/B 634. To take account of this remark, one needs to distinguish between the general phenomenon of transcendental illusion (that is, the appearance that *D* is true) and the particular illusions to which it gives rise. Kant’s position appears to be that we are to think of the particular illusions as “resting on” certain particular dialectical inferences, which are their immediate grounds, but also that we are to think of these inferences themselves as resting, in turn, on the general illusion that *D* is true.

“dialectical inferences,” which are set over against the Antinomies, on the one hand, and the arguments of the Ideal of Pure Reason, on the other.

In the *Jäsche Logic* Kant divides formal fallacies into “paralogisms” and “sophisms” as follows: a fallacy counts as a *paralogism* insofar as one deceives oneself through it, but as a *sophism* insofar as one intentionally seeks to deceive another (9:134–35). This contrast, however, is unlikely to be the one guiding Kant’s use of the term “paralogism” in the first *Critique*; for if it were, he would be implying that those “dialectical inferences of reason” (A 309/B 366) that are discussed outside the Paralogisms chapter—those found in the Antinomies and in the Ideal of Pure Reason²⁹—are in fact merely sophisms. And yet it is scarcely credible that Kant would have taken *all* of the unsound arguments criticized in these other chapters to have been propounded deceitfully, even if he supposes that some of them might have been.³⁰ Indeed, Kant seems to imply that these arguments are often given in good faith. For example, speaking quite generally of the “dialectical witnesses” that “transcendent reason brings forth on behalf of its pretensions”—and so, presumably, of the arguments of the Antinomies and Ideal among others, he says: “we knew beforehand that all their allegations, *while perhaps honestly meant*, had to be absolutely null and void.” (A 703/B 731, emphasis added). For these reasons, the paralogism-sophism contrast does not seem to be a good guide to the thinking behind Kant’s decision in the first *Critique* to style a particular class of dialectical inferences “paralogisms.”

A better text for these purposes occurs at the beginning of the Paralogisms chapter itself:

A logical paralogism consists in the falsity of a syllogism due to its form, whatever its content may otherwise be. A transcendental paralogism, however, has a transcendental ground for inferring falsely due to its form. (A 341/B 399)

29. In the case of the Ideas of the world and God, Kant also takes transcendental illusion to lead to certain “dialectical inferences.” He maintains that in the former case these inferences take the form of hypothetical syllogisms, in the later, disjunctive syllogisms. A 340/B 398; A 406/B 432–33.

30. Kant does suggest that some of these arguments are deployed deceitfully. See, for example, his discussion in the *Hechsel Logic* of the fallacy “*sophisma heterozeteseos*” (the fallacy of searching for something else) as it occurs in the proof of the existence of God. Kant suggests that in this case one pretends to give a proof of an interesting claim while actually giving a proof of a claim that is much weaker and consequently less interesting. Young 1992, 410–11. These notes, which are based on lectures Kant gave in the early 1780s, were not discovered until the 1980s and so are not included in Kant 1902. For a discussion of their dating, see Pinder 1987.

The part of this passage that deals with *logical* paralogisms is echoed both by the *Blomberg Logic*, where Kant says that “Every inference false in form is a *paralogismus*” (24:287), and by a marginal note to the first edition of the first *Critique*, where Kant says that “A paralogism is a syllogism that is false *in forma*” (Erdmann 1881, 154; 23:38, quoted in Guyer and Wood 1998, 411). Clearly, then, a paralogism is a formally invalid syllogism. But there is more to it than that, as a reflection from the *Duisburg Nachlaß* makes clear:

A paralogism is a syllogistic inference that is false as far as its form is concerned, although as far as its matter (the antecedent propositions) are concerned [*bei Vordersätzen*], it is correct. (R 5552; 18:218, translation my own³¹)

So a paralogism is false in form,³² but correct in matter.³³ This remark suggests that what Kant means by a “paralogism” in the first *Critique* is

31. In rendering the dative plural phrase “*bei Vordersätzen*” as “the antecedent propositions,” I follow the practice of J. Michael Young. See his translation of the *Jäsche Logic*, especially sec. 59 (9:121) in Young 1992. It should be noted that the translators of the Cambridge edition of Kant’s reflections are misleading here. They render this plural noun as “the major premise” (Guyer 2005, 236), thereby giving the false impression that Kant once supposed that the major premise alone constituted the “matter” of a syllogistic inference. Kant, however, always supposed that both premises constituted the matter of the argument.

32. In the *Hechsel Logic* Kant glosses an argument’s being “false as to form” as its being a “fallacy” (16:110); so there is reason to think that, in characterizing a paralogism as false “due to its form” in the first *Critique* (A 441/B 399), Kant is simply characterizing it as formally fallacious. This apparently obvious point is worth emphasizing because Kant elsewhere equates an argument’s “form” with its *conclusion* (for example, *Jäsche Logic*, sec. 59). However, if Kant really thought we knew a paralogism’s conclusion to be false, he would be dogmatically committed to its negation. His claim, for example, that the first paralogism is false “due to its form” would be tantamount to the claim that the self is not a substance₁. To avoid saddling Kant with such an implausibly dogmatic stance, we need to allow ourselves to be guided on this point by Hechsel’s notes in preference to Jäsche’s and to understand a paralogism’s “form” as what we would today call its logical form rather than its conclusion.

33. A historical precedent for treating the premises of an argument as its “matter” and its logical form as its “form” occurs in the writings of Christian Wolff: “with inferences one is to look in part to the correctness of the propositions, in part to their mode of connection with one another, that is, as one usually says for short, in part to the matter, in part to the form.” Wolff 1747, sec. 390, translation my own. It is also present in St. Thomas Aquinas’s commentary on Aristotle’s *Physics*: “Each argument of both Melissus and Parmenides is sophistic, for they err in matter, whence [Aristotle] says that they have accepted what is false, that is, they assume false propositions, and they err in form, whence he says they are not syllogizing.” Aquinas 1999, 10. This kind of application of the matter/form distinction to *arguments* is absent from the passages of Aristotle on which St. Thomas is commenting here.

an inference that is invalid, even though its premises are true. (We shall soon see that there is confirming evidence for this interpretation in the Paralogisms chapter itself; but it will be convenient to defer its discussion until the next section.) If this idea is correct, it would serve to distinguish the Paralogisms from the arguments of the Antinomies and the Ideal since, among these three groups of dialectical inferences, only the Paralogisms are such that every member of the group is taken to be an invalid argument with true premises.

It is worth noting, further, that each paralogism is supposed to involve a fallacy of equivocation. That much is clear from the continuation of the reflection from the *Duisburg Nachlaß*:

[A paralogism] arises when the middle concept is taken in different senses in the two premises—when, namely, the *logical* relation (in thinking) in one of the premises is taken as a real one (of the objects of intuition) in the other. (R 5552; 18:218)

What this reflection proceeds to discuss are not fallacious inferences about the self but rather the body of fallacious inferences “from the mere actions of reflection” (A 278/B 334) that Kant terms “the amphiboly of the concepts of reflection.” And although his deletions indicate that he later revised his view that the concepts of reflection “lead to paralogisms” (R 5552; 18:218), the fact that he once considered applying the concept of a paralogism to subject-matter unconnected with the self suggests that “paralogisms” are arguments of the kind just described, whether or not they concern the self.

As we might have expected given that Kant’s focus in the Paralogisms chapter lies specifically on *transcendental* paralogisms, he is at pains to emphasize the naturalness and inevitability of our tendency to want to engage in paralogistic reasoning. Speaking of dialectical inferences in general (and so of the transcendental paralogisms among other things), he says:

They are sophistries not of human beings but of pure reason itself, and even the wisest of all human beings cannot get free of them; perhaps after much effort he may guard himself from error, but he can never be wholly rid of the illusion, which ceaselessly teases and mocks him. (A 339/B 397)

The paralogisms lead us to believe that we can have some substantive (that is, synthetic) knowledge of the self as it is in itself. In truth, however:

The subject of inherence [that in which thoughts inhere] is designated only transcendently through the I that is attached [*angehängte*]³⁴ to thoughts, without [our] noting the least property of it, or being acquainted with it or knowing it [through propositions] at all. (A 355, translation my own³⁵)

Kant is saying both that we cannot intuit the self as it is in itself and that we cannot attain to propositional knowledge of it—and presumably he means propositional knowledge *of its inner character or nature*; for he had earlier claimed that the self is “cognized [*erkannt*] through the thoughts that are its predicates” (A 346/B 404, translation my own). Thus, for example, while I can know that I am thinking that it’s raining, I cannot know whether or not the subject of this thought is a substance₁.

4. The Paralogistic Fallacy in the B-edition

Kant gives different accounts of the dialectical inference constituting the first paralogism in the two editions of the first *Critique*. I will focus on the second edition’s version because it belongs to a chapter that Kant rewrote and which he expressly recommends as an improvement on the text it replaces (see B xxxviii; compare *Prolegomena*, 4:381). I will, however, briefly discuss the A-edition version in the next section.

Precisely how is the formal fallacy of the B-edition paralogism to be explained? Kant’s interpreters are far from agreement on this question, and, as it seems to me, none has yet arrived at a wholly satisfactory

34. Here I follow Kemp Smith (1933) in favoring the looser “attached” to the more literal “appended” since the “I think” is typically prefixed to thoughts rather than appended to them.

35. My rendering of this passage diverges from several established translations in its treatment of the troublesome phrase “*ohne . . . überhaupt etwas von ihm zu kennen, oder zu wissen*.” Guyer and Wood offer “without . . . cognizing or knowing anything at all about it.” But according to their own principles of uniform translation (Guyer and Wood 1998, 73) “cognizing” would be appropriate only if the corresponding verb was “*erkennen*.” Kemp Smith (1933) offers “without knowing anything of it either by direct acquaintance or otherwise.” However, as Kitcher notes, this both misleadingly deploys the Russellian term of art, “direct acquaintance” and fails to bring out the nature of whatever notion it is with which “*kennen*” is being contrasted. Kitcher 1982, 528. Kitcher herself renders the sentence in which the phrase occurs “[we have no knowledge of the thing that thinks] through awareness or through reasoning.” Ibid. However “*wissen*” carries no implication of reasoning but rather, through the contrast with “*kennen*,” one of having *propositional* knowledge. My own translation resorts to an interpolation to bring out the (familiar enough) contrast between “*kennen*” and “*wissen*” as relating to “knowledge of/acquaintance with,” on the one hand, and “knowledge about,” on the other.

answer. Before offering my own account, it will be useful to state some constraints on any adequate interpretation.

First, as our consideration of Kant's use of the word "paralogism" has suggested, any textually faithful reading will need to treat the first paralogism as a fallacious categorical syllogism with an ambiguous middle term. That this constraint is operative in the first paralogism is confirmed by Kant's remark that the conclusion of the first paralogism is drawn "*per sophisma figurae dictionis*" (A 402 and B 411), which, as he explains in his logic lectures, means that the paralogism has an ambiguous middle term (*Jäsche Logic*, sec. 90, note; 9:135).

Second, as our reflections in the preceding section also suggested, any adequate reading must portray the paralogism's premises as true. This conclusion is confirmed by Kant's remark in the A-edition: "If one wants to give a logical title to the paralogism in the dialectical syllogisms of the rational doctrine of the soul, *insofar as they have correct premises*, then it can count as *sophisma figurae dictionis*" (A 402, first emphasis added). Since in the B-edition Kant again says that the first paralogism's conclusion is drawn "*per Sophisma figurae dictionis*" (B 411), it seems likely that by his lights the B-edition version of the argument also contains "correct" (that is, true) premises.

A third constraint on any adequate reading is that it should portray the fallacious inference as highly seductive. What transcendental illusion produces in us, after all, is not a *compulsion* to err, but merely a *temptation*. And this temptation cannot be adjudged so forceful that it could induce someone to accept a patent non sequitur as a valid argument.

With these constraints in hand, we are finally in a position to analyze the first paralogism. In the B-edition it is presented as consisting in the following categorical syllogism:

What cannot be thought otherwise than as subject does not exist otherwise than as subject, and therefore is substance.

Now a thinking being, considered merely as such, cannot be thought otherwise than as subject.

Therefore it also exists only as such a thing, that is, as substance. (B 410–11)

This argument gives the misleading impression of being an instance of a valid argument-schema, namely, the first-figure categorical syllogism in Barbara:

Kant's First Paralogism

All *M* are *P*—major premise
All *S* are *M*—minor premise
Therefore: All *S* are *P*—conclusion

The B-paralogism—or, rather, a slightly more regimented (and clarified) version of it—is obtained by replacing the schematic letters as follows:

M: entities that cannot be thought otherwise than as subjects
P: entities that cannot exist otherwise than as subjects, and therefore (by definition) are substances
S: entities that are thinking beings (considered merely as such)

Two remarks about this regimented version of the paralogism are in order. First, I have supplied the term “entities,” which is intended to be categorially neutral, merely in order to massage the argument Kant actually presents into the (apparent) form of a first-figure categorical syllogism in Barbara. This requires construing each of the predicates as a plural noun-phrase. Second, in order to be consistent with Kant’s definitions of substance at B 149, B 288, and B 289, the predicate-term, *P*, needs to read “entities that *cannot* exist otherwise than as subjects . . .” rather than “entities that *do not* exist otherwise than as subjects . . .” Thus our formulation of *P* involves a modal strengthening of what Kant actually writes. The need for this stronger formulation is further indicated by the footnote of B 411–12, in which Kant presents a consequence of the argument’s intended (but not attained) conclusion as being that “I *cannot* exist otherwise than as a subject” (emphasis added). In what follows I will treat Kant’s omission of the needed modal force in his main statement of the argument as merely a slip.

As we have noted, the argument does not actually have the form of a first-figure categorical syllogism in Barbara on account of the ambiguity of its middle term (A 402). Kant locates the source of this ambiguity in the word “thought”:

‘Thought’ [*Das Denken*] is taken in an entirely different meaning [*Bedeutung*] in the two premises: in the major premise, as it applies to an object in general (hence as it may be given in intuition); but in the minor premise only as it subsists in relation to self-consciousness, where, therefore, no object is thought, but only the relation to oneself as subject (as the form of thinking) is represented. (footnote, B 411–12, translation my own)

The interpreter’s task is to identify these two senses of “thought.” We will aim to do so by considering the intent of each of the premises.

The major premise says, in effect, that if it is impossible to *conceive* of some entity as a property or mode, then that entity cannot *exist* as a property or mode. Since the premises of the paralogism are supposed to be true, Kant must agree with this generalized conditional. He nonetheless avoids dogmatism because he also insists that we can never know any instance of the antecedent to be true.³⁶ As he sees it, the conclusion that something cannot be conceived of as merely a property or mode of something else must always be premature since it amounts to the dogmatic assertion that that thing is an ultimate subject.³⁷

Kant, then, has a general reason for denying that the predicate figuring in the antecedent of the major premise—"cannot be thought otherwise than as subject"—can be known to apply to anything. But, in addition, he has a specific reason for denying that this predicate can be known to hold of a thinking being considered merely as such. For it is at least arguable that the self—to take one such being—*can* be conceived "otherwise than as a subject" (that is, as a property or mode of something else). Indeed, one version of such a view is standardly attributed to Spinoza, who holds the human mind to be a mode inhering in the one true substance: God.³⁸ Spinoza may or may not be right about this, but his position is at the very least not *known* to be unintelligible—and the same can be said for the alternative ideas that the self is merely a power of a thing,³⁹ and that it is a configuration of other things. But such facts are

36. One might reasonably doubt whether by the lights of his own epistemology Kant could really be entitled to grant the truth of the major premise. For it is not clear upon what grounds he could take himself to know that such a claim is true. Since the issue is rather involved, I reserve its discussion for section 6 below, "Objections and Replies."

37. Compare a reflection conjecturally dated to 1773–75: "Our reason is not a faculty the progress of whose use we can arbitrarily cut off. . . . It is contrary to reason to cut off further research and wantonly suspend all further effort." R 4741; 17:694.

38. In the *Ethics* Spinoza says that "the human mind is part of the infinite intellect of God." *Ethics* 2, P 11, Corollary; Spinoza 1992, 70. Everything for Spinoza is a mode or a substance. *Ethics* 1, P 15, Proof; Spinoza 1992, 40. Since God is the sole substance, the self is not a substance and so must be a mode.

39. Kant was unquestionably aware of one view according to which the soul is not a substance.¹ This is Baumgarten's view according to which the soul is to be identified with a representational power, namely, the power of representing the universe from a particular point of view. See Baumgarten 1902, sec. 741; and see 29:906 for Kant's awareness that Baumgarten held such a view. As Karl Ameriks notes (2000, 82, n. 94), Kant attributes a similar view to Wolff (28:261), and he takes it to imply that the soul is an accident. 28:261. However, Kant gives no citation when he makes this attribution, and one has to wonder whether Wolff ever held such a view. Certainly, if he did, he was not consistent in doing so; for in Wolff 1747 he clearly maintains that the soul is a self-subsisting thing that *has* powers (see secs. 743–44). What matters for our purposes, however, is that although

all Kant requires in order to argue that the minor premise, interpreted as it would need to be in order to mesh with the major, is not known to be true. So, even if the argument were interpreted so as to be valid, it would still not be known to be sound, and so would not constitute a demonstration.

The major premise, I contend, involves treating “thought” as meaning “conception.” I take Kant’s position to be that in the minor premise “thought” does not have this sense, even though, when in the grip of transcendental illusion, we are tempted to think that it does. But then what sense *does* it have?

The best place to begin the examination of this question is with the continuation of the footnote from B 411–12. Kant says:

In the first premise, things are talked about that cannot be thought of other than as subjects; the second premise, however, talks not about **things** but about **thought** (in that one abstracts from every object), in which the I always serves as subject of consciousness; hence in the conclusion it cannot follow that I cannot exist otherwise than as subject, but rather only that, in thinking my existence, I can use myself only as the subject of the judgment, which is an identical proposition that discloses absolutely nothing about the manner of my existence. (footnote B 411–B 412, translation my own)

Let us focus on the conclusion about the self that Kant supposes we *are* entitled to draw from the premises of the first paralogism, namely:

- [1] In thinking my existence I can use myself only as the subject of the judgment.

Because Kant supposes we are entitled to infer [1] from the premises of the argument, and because he takes those premises to be true, he must also take [1] to be true. So a charitable interpretation of [1] is in order. But any charitable reading of this claim must take Kant to be committing an error that is the analogue at the level of judgment of a confusion of use and mention. What I “use” as the subject of a judgment expressing my existence, after all, is not *my self* but only a judgmental component representing my self.⁴⁰ So Kant must mean:

Kant takes issue with the view that the soul is a power, he does not seem to regard it as unintelligible. See 28:261–62 and 29:906.

40. Kant’s talk of “using” the self or the “I” can be clarified by comparing it to his discussion of the use of more ordinary concepts in judgments. In the *Metaphysical Foundations of Natural Science*, in the course of explaining the notions of substance and accident (understood as pure concepts of the understanding), he says:

- [2] In thinking my existence I can use the representation of my self only as the subject of the judgment.

This is better, but we are not yet finished with our charitable reformulations; for [2], as it stands, is, strictly speaking, false. To see why consider the judgment: “Something instantiates the property of being identical with me.” In this judgment I think my existence without using the representation of my self as a subject, for it is used only as a component of the predicate. Nor do I use the representation of my self as a subject in the judgment: “The property of being me is exemplified,” for here my representation, “me,” is used only as a component of the subject concept. To do justice to Kant’s thought, therefore, we will have to suppose that when he says such things as [1] he does so against the background of the assumption that any subsentential component of a categorical judgmental occurs either as a subject or as a predicate.⁴¹ Although this assumption is, strictly speaking, false, it does little harm to Kant’s reasoning. To correct for it, we need only suppose that, when he says that a

In the categorical judgment **The stone is hard**, the **stone** is used as subject, and **hard** as predicate, in such a way that the understanding is still free to exchange the logical function of these concepts, and to say that something hard is a stone. (4:475, continuation of note “*”)

Here when Kant says “the **stone** is used as subject” it is clear that he means that the *concept stone* is used as subject. The emphasis on the word “stone” thus seems intended to indicate that in this context this word does not refer to an object (a stone) or even to stones in general (the plurality of stones) but only to a concept. The same typographical convention seems to be in play in Kant’s use of the phrase “the **I**” at A 349 when he says: “in all our thinking the **I** is the subject, in which thoughts inhere only as determinations, and this **I** cannot be used as the determination of another thing.” The parallel with the phrase “the **stone**” suggests that the emphasis on the word “I” in the phrases “the **I**” and “this **I**”—indicated in the Akademie edition by spaced typesetting—is intended to indicate that Kant means to be referring to the judgmental component expressed by the word “I”—which component is presumably the concept of the soul spoken of at A 351. His point, therefore, is that the *concept* expressed by the word “I” cannot be employed in a judgment in a predicative role. Note: some editions of the first *Critique*, including the Guyer-Wood translation, depart from the Akademie edition by omitting the emphasis on “I” in the phrase “this **I**” while including the emphasis in the phrase “the **I**.” For all I know, this may be because they are relying on a text that is (at this point) more authoritative than the Akademie edition. But even if that should be so, the fact that the phrase “this **I**” is anaphoric on the phrase “the **I**” strongly suggests that Kant would have intended the emphasis to be present in both phrases.

41. Kant does not treat the copula as expressing a judgmental component. Instead, he treats it as expressing the judgment’s form. See, for example, *Jäsche Logic*, sec. 24, (9:105).

judgmental component “can be used only as [the] subject of the judgment,” what he really means is that it “cannot be used as [the] predicate of the judgment”—whereby a “predicate” *here* he means a judgmental component in predicative role, rather than a property or mode. At the end of the day, then, what Kant must mean by [1] is just:

- [3] In thinking my existence I cannot use the representation of my self as [the] predicate of the judgment.⁴²

At this point it would be well to recall what Kant says about [3] in the footnote to B 411–12. He says that all that follows from the argument is [3] in contradistinction to the more metaphysically loaded claim that “I cannot exist otherwise than as subject.” The focus on existence in each of these claims suggests that Kant means to be saying that:

- [4] All that follows from the premises of the paralogism *as far as my existence is concerned* is *not* that I cannot exist otherwise than as subject but merely that [3].

It is easy to see why Kant should take [3] to follow from the premises of the paralogism; for it follows from a combination of the minor premise, construed as saying

- [5] The representation of a thinking being, considered merely as such, cannot be used as the predicate of a judgment,

with the independently plausible claim that

- [6] The representation “I” (or “me”)—or, strictly, the judgmental component expressed by these words—is the representation of a thinking being considered merely as such.

There is, then, good reason—provided by footnote B 411–12—to take the minor premise as meaning [5]. And, as we required, the minor premise, so construed, seems to be *true*. For the representation of any thinking being considered merely as such will be a singular represen-

42. Kant describes [1] as an “identical proposition,” which means that it is an *analytic* judgment. A 7/B 10; *Jäsche Logic*, 9:111. One supposes that he must view it this way because it (a) contains necessity but (b) is not plausibly construed as a claim whose truth is a condition of the possibility of experience. One supposes that Kant would have regarded [1] as analytic because it amounts to [3] and because it is analytic of our concept of the word “*Ich*” that it has the grammatical distribution it does. Presumably, it would be thought of as “That German word whose phonetic signal (in such-and-such a dialect) is so-and-so and whose meaning is such-and-such and whose grammatical distribution is so-and-so, where this last “so-and-so” would contain the component “does not occur as a predicate.” Such an account might be satisfactory for the word “*Ich*,” though one has to worry whether it would really carry over to the level of judgmental components.

tation (whether it be a proper name, an indexical, a demonstrative, or a definite description). And representations in this category cannot play the grammatical role of predicates⁴³—though, of course, they may figure as parts of predicates. In the case of a term for a universal, by contrast, that term can occupy either subject or predicate position in a judgment. For example, both “Blue is a color” and “American mailboxes are blue” are grammatical. (Of course, the same goes *mutatis mutandis* for what really concerns us: the judgmental components these words express. Hereafter, this point will be left unstated.)

Kant’s key observation about the B-paralogism—the point on which his entire critical analysis of the argument turns—is that this fact about the grammatical distribution of the first-person singular personal pronoun entails nothing about how the self can or cannot be conceived. *A fortiori*, it entails nothing about whether or not the self is a substance₁. Thus, although the representation of the self cannot be used as the predicate of a judgment, this fact “discloses absolutely nothing about the manner of my existence” (B 412). This central point can be underscored by a comparison with the case of an ordinary proper name. Suppose I were to name my right fist “Fred.” Kant would accept that I cannot put the word “Fred” in predicate position; but this emphatically does not show that Fred is a substance₁. Indeed, since Fred is something whose existence depends on the configuration of other things, it must be deemed a *mode*; hence, a “determination” of other things.

On this reading, the ambiguity on which the paralogism trades can, as Kant suggests, be located in the word “thought.” In the major premise to “think” of something is to *conceive* of it. But in the minor premise to “think” of something is to assign its representation a certain role in a judgment. Thus, in the sense appropriate to the major premise, an entity that cannot be “thought” as a predicate is something that cannot be conceived as a property or mode of another thing; while in the sense appropriate to the minor premise an entity that cannot be “thought” as a

43. One might be inclined to doubt this on the grounds that in the judgment, for example, that “Allah is God,” it looks as though “God” occupies predicate position. However, in spite of appearances to the contrary, this judgment does not in fact contain a predicate position since it is not of subject-predicate form. Rather, it is an identity judgment, and so of relational form. In a Kantian setting, of course, this point may be less obvious than it is to us today since Kant failed officially to recognize relational judgments as having a distinctive logical form. Nonetheless, it remains the case that the fact that “God” is a singular term disqualifies it from predicate-position, even if Kant himself lacked some of the resources needed to defend this point against alleged counterexamples of this kind.

predicate is something whose representation cannot play the grammatical role of a predicate. Exploiting the ambiguity of “predicate,” which in Kant’s writings sometimes means a property (or mode) and sometimes a judgmental component in predicative role, we might characterize the ambiguity in question as that between the notion of something that cannot be *conceived* as a predicate and the notion of something that cannot be *deployed* as a predicate. Alternatively, sacrificing elegance to precision, we might say that the kernel of Kant’s critical insight amounts to the following observation: the fact that something’s representation cannot be deployed as a predicate in judgments doesn’t entail that that thing itself cannot be conceived as a property or mode of another thing, and, a fortiori, it doesn’t entail that it cannot exist in this manner. That observation, I take it, constitutes a philosophical insight of real depth.⁴⁴

44. Among the interpretations of the B-edition paralogism known to me, the one that comes closest to my own is that of Karl Ameriks (2000). According to Ameriks (2000, 71), the reason there can be no sound argument of the kind the first paralogism attempts to be is that “despite systematically misleading appearances, we don’t have sufficient evidence to make any subsumption under the first premise.” I agree with that assessment; but my interpretation of the B-paralogism differs from Ameriks’s on one substantive point and three points of emphasis. The substantive point is that the middle term (in the formally presented argument) can (and, if we are to make sense of Kant, must) be viewed as ambiguous in *two* places: in the word “thought” and in the word “entities.” The first point of emphasis is that the minor premise is not, in the first instance, as Ameriks would have it, a claim about how I must represent myself whenever I consider myself merely as thinking (that is, as a subject). It is more specifically a claim about where my representation—paradigmatically, the word “I”—can or cannot occur in a judgment. One might, of course, offer this as a gloss on “how I must represent myself whenever . . .”; so on this point, our readings are, I think, compatible. However, they differ insofar as Ameriks does not in fact offer this gloss, and so leaves it somewhat unclear what it is to represent the self as a subject. A second (very minor) difference of emphasis relates to the idea that the focus of the minor premise lies not, as Ameriks would have it, on the point that I must think of myself as subject, which, on my reading, would mean that the representation “I” must occur in subject position; it lies rather on the point that the representation “I” cannot go in predicate position. A third difference of emphasis relates to my employment of the idea that, “in relation to self-consciousness” (that is, in the epistemic state I’m in at the terminus of Cartesian doubt), my resources for thinking of myself are diminished in such a way that, under the influence of transcendental illusion, it will easily seem that I have nothing to appeal to but an inappropriate grammatical test when considering whether or not I can conceive of myself as a determination. In spite of these differences, I believe there is much to admire in Ameriks’s account, and, in particular, I agree with many of his criticisms of other interpreters. See Ameriks 2000, 68, 71–73.

5. The Paralogistic Fallacy in the A-edition

We are now equipped to attempt a reading of the A-edition version of the first paralogism.⁴⁵ I will argue that Kant's criticism of this version of the paralogism consists, at bottom, in the observation that it operates with the wrong conception of substance—"wrong" because it fails to correspond to any notion the rational psychologist might have in mind when claiming that the self is a substance. This conception turns out to be even weaker than the notion of substance₁ and to make an essential appeal to considerations having to do with where a judgmental component can (and cannot) occur in a judgment.

Kant formulates the A-edition version of the argument as follows:

That the representation of which is the **absolute subject** of our judgments, and hence cannot be used as the determination of another thing, is **substance**.

I, as a thinking being, am the **absolute subject** of all my possible judgments, and this representation of Myself cannot be used as the predicate of any other thing.

Thus I, as thinking being (soul), am **substance**. (A 348)

This argument diverges even farther than did the B-edition version from the strictly correct form of a categorical syllogism: the first premise is a definition rather than a universal claim stating a sufficient condition for being a substance; and the minor premise isn't general. Putting it into the form of a syllogism with a singular minor premise and simplifying by eliminating the redundant, defined predicate: "is the absolute subject of our judgments," we may reformulate the argument as follows:

Entities whose representation cannot be used as a determination of other things are substances.

The self is an entity whose representation cannot be used as a predicate of another thing.

Therefore the self is a substance.

On the assumption that "determination" in this context just means "predicate," the argument is valid. Moreover, because, as we have seen, the representation of the self cannot be used as a predicate, the minor premise is true. Since the major premise is just a consequence of a definitional truth, the argument is sound. The argument's only shortcoming is that

45. The ideas in this section have benefited greatly from a conversation with Kenneth Winkler, who convinced me that the conclusion of the A-edition first paralogism had to be *somehow* weaker than the proposition that the soul is a substance₁.

its conclusion is too weak to serve the rational psychologist's purposes. It establishes only that the self is a "substance" in the grammar-driven sense of something whose representation cannot occur as a predicate in a judgment. As Kant puts it, the argument establishes that the soul is a "substance only in the idea" (A 351).⁴⁶ In order to emphasize that this is the weakest of the conceptions of substance we have considered so far, it will be convenient to label it "substance₀."⁴⁷ As Kant observes, the argument doesn't entail anything metaphysically interesting—least of all the soul's natural immortality (A 350–51; A 400); for, as we've seen from our discussion of the case of "Fred," some substances₀ can cease to exist through the decomposition or rearrangement of their parts.

Because the A-edition first paralogism employs a weak and metaphysically neutral conception of substance, there is no need to posit a fallacy of equivocation to diffuse its force. But, equally, lacking such an equivocation, it fails to satisfy Kant's official description of a paralogism.⁴⁸

46. More precisely, he says that the *concept* of the soul *signifies* a substance only in the idea. This entails that the soul itself *is* a substance only in the idea—and that means that the soul's representation—the "I" of apperception—cannot be used as a predicate in judgments.

47. Kant's notion of substance₀ seems not to have any close precedent in the tradition. It might be thought to correspond loosely to Aristotle's notion of something that is not "said of" a subject from the *Categories* (sec. 5) (Aristotle 1984); for it's possible to read Aristotle here as offering a characterization of substance that rests, partly, on grammatical considerations. But that idea is only one element of Aristotle's notion of primary substance rather than the notion itself. It seems that Kant was led to the notion of substance₀ by his attempt in the A-edition to analyze the paralogisms quite generally as establishing nothing more than claims about the self "in the idea" or "in concept." The generality of this ambition is evident from his remark that mere apperception ("I")

is substance in concept, simple in concept, etc., and thus all these psychological theorems are indisputably correct. Nevertheless, one by no means thereby cognizes anything about the soul that one really wants to know, for all these predicates are not valid of intuition at all, and therefore cannot have any consequences that could be applied to objects of experience; hence they are completely empty. (A 400)

48. Karl Ameriks has suggested—quite plausibly—that Kant's A-edition criticisms of the first paralogism are directed not at the argument Kant presents there as the paralogistic syllogism per se but rather at an invalid extended argument for the permanence of the soul. Ameriks 2000, 68. According to Ameriks, the rational psychologist's error consists in passing from the conclusion that I cannot be a determination of another thing (that is, that I am a substance₁) to the further conclusion that I am a substance₂. Such a step involves a version of the fallacy Jonathan Bennett calls "inflating" the first paralogism. Bennett 1974, sec. 25. The B-edition improves on this diagnosis and criticism by locating the fallacy squarely within the paralogistic syllogism itself. I agree with the general style of this analysis, but I would claim that the conclusion that gets inflated

In order to present the belief in the substantiality of the soul as reached by a genuine paralogism, what is needed is a better definition of substance; and that is precisely what the B-edition's version of the paralogism works with—namely, substance₁.

As I see it, then, the main change between the A- and B-edition versions of the first paralogism—a change that enables Kant to portray the paralogism as invalid—is the change in the definition of substance from “that the representation of which cannot be employed as a determination of another thing”—substance₀ (compare A 348)—to “that which cannot exist otherwise than as subject”⁴⁹—substance₁ (B 149, B 288, B 289). This thought is confirmed by the absence of the latter notion from the A-edition and by Kant's omission of the former from the parts of the B-edition that he rewrote. I would therefore claim that it is his reformulation of the notion of substance that constitutes the main improvement in his presentation of the first paralogism.

6. Objections and Replies

The account I have given of the B-edition paralogism satisfies our main desiderata: it portrays the paralogism as a fallacious syllogism containing an ambiguous middle term; the fallacy is sufficiently subtle that it might have been committed by thinkers of some sophistication; and the equivocation is located precisely where Kant says it is—namely, in the word “thought.” So the reading has its merits. But what are its points of vulnerability?

One possible line of objection alleges that the interpretation would commit Kant to predicting the occurrence of dogmatism where it does not exist. If Kant is right that we are apt, when in the grip of transcendental illusion, to confuse the idea that the self's representation cannot occur in predicate position in a judgment with the idea that the self cannot be conceived as a determination (that is, as a property or mode), why, one might want to know, aren't we apt to make a similar mistake in other cases? Why, for instance, shouldn't we also be inclined to think of my right fist, whose representation, “Fred,” cannot occur in predicate position, as a substance₁?

(and which constitutes a lemma in the invalid, extended argument) is that the soul is a substance₀ rather than, as Ameriks seems to suppose, a substance₁.

49. Recall that I am treating the omission of modal force in the predicate of the major premise in the B-version as just a slip.

The reply, I take it, should run as follows: we are not tempted to take “Fred”’s inability to occupy predicate position as a sign that what it represents is incapable of being conceived as a determination of another thing because what it represents is *intuitable*. In intuiting Fred, I can perceive that it is complex and thereby come (perhaps tacitly) to understand that it is something whose existence consists in the configuration of other things. In other words, sensible intuition can assure me that Fred is a mode. In the case of the self, by contrast, all I have to go on in considering whether or not it can be conceived of as a determination of another thing are the features that could be known to attach to it at the terminus of Cartesian doubt. But, crucially, among these features is nothing that would indicate that the self is not a substance₁. They include only the general property of having thoughts, along with the particular properties of believing that *p*, doubting that *q*, willing that *r*, and so forth.⁵⁰ It is because in considering the self we have nothing to go on beyond these meager resources that in this case the fallacy is so easy to commit.

So much, then, for the first line of possible criticism. A second line challenges my attribution to Kant of the idea that the judgmental component expressed by “I” cannot occur as a predicate. Some grounds for this possible objection are mentioned by Patricia Kitcher (1982, 526). Although she herself wants to attribute something like this claim to Kant as part of her interpretation of the A-edition first paralogism, she does so only with reservations. “Kant,” she says, “never says why the ‘I’ cannot occur as a predicate. Further he himself points out that any concept can occur in either the subject or the predicate position in a sentence” (ibid.). As support for the second of these claims, Kitcher cites the texts: A 349, B 128–29, and A 242–43/B 300–301. Kitcher’s contention, if correct, would pose a difficulty for my reading only if what she means by a “concept” is a judgmental component in general. For the sake of argument, I shall suppose that that is so.

The reply to this objection simply involves observing that the passages cited do not in fact support the conclusion Kitcher draws from them. Consider A 349. Far from suggesting that any judgmental component can play the role of a predicate, this passage actually suggests that the judgmental component expressed by “I” cannot do so. It contains

50. As I noted earlier, it is clear that Kant takes us to know at least this much about the self, for he says, “Through this I, or He, or It (the thing), which thinks, nothing further is represented than a transcendental subject of thoughts = *x*, which is cognized [*erkannt*] only through the thoughts that are its predicates.” A 346/B 404, emphasis added, translation my own.

the claim that “in all our thinking the **I** is the subject, in which thoughts inhere only as determinations, *and this I cannot be used as the determination of another thing*” (A 349, emphasis added). If I am correct that when Kant speaks of “using the **I**” he means employing the self’s representation in a judgment, this would have to be taken to mean that the judgmental component expressed by the word “**I**” *cannot* occur as a predicate in a judgment. At any rate, there is no hint of a suggestion at A 349 that this representation *can* function as a predicate. Nor, so far as I can see, is there any hint of this idea at A 242–43/B 300–301. The only text cited by Kitcher that might seem to support the attribution of such a view is the passage from B 128–29. Here Kant offers a characterization of the categories:

[The categories] are concepts of an object in general, by means of which its intuition is regarded as **determined** with regard to one of the **logical functions** for judgments. Thus, the function of the **categorical** judgment was that of the relationship of the subject to the predicate, e.g., “All bodies are divisible.” Yet in regard to the merely logical use of the understanding it would remain undetermined which of these two concepts will be given the function of the subject and which will be given that of the predicate. For one can also say: “Something divisible is a body.” Through the category of substance, however, if I bring the concept of a body under it, it is determined that its empirical intuition in experience must always be considered as subject, never as mere predicate; and likewise with all the other categories. (A 94/B 128–29)

As I read this passage, Kant is not claiming that any representation whatsoever can occur as a predicate in a judgment. Rather, he is claiming that prior to the application of the category of substance, any *general* term might play either subject- or predicate-role in a subject-predicate judgment. As we have already noted (with our example of “blue”), this flexibility is typical of general terms. That idea, however, is far from entailing that “**I**” can occur in predicate position, for “**I**” is a representation that purports to stand for a *particular*.

A third possible objection would charge that my interpretation is flawed because it is *incomplete*. The problem is that immediately after presenting the first paralogism, Kant offers an explanation of the inference’s deceptiveness that appears to differ from the one in the footnote on which we have been focusing so far:

The major premise talks about a being that can be thought of in every respect, and consequently even as it might be given in intuition. But the

minor premise talks about this being only insofar as it is considered as subject, relative only to thinking and the unity of consciousness, but not at the same time in relation to the intuition through which it is given as an object for thinking. (B 411)

The present objection simply charges that we haven't yet done justice to this apparently rather different explanation of the equivocation. The "reply" to this objection—such as it is—is just to concede its correctness and to show that the needed further explanation is compatible with the one given so far.

Let's begin by considering the contrast in the second sentence. Kant says that the minor premise talks about a being—namely, the self—in a way that contrasts with thinking of it "in relation to the intuition through which it is given as an object for thinking." Kant repeatedly insists that the self as it is in itself is not an object of intuition; so he cannot mean that the minor premise neglects to consider the self relative to the intuition through which it is given as an object for thinking. What he must mean, rather, is just that if, contrary to fact, the minor and major premises were to mesh into a valid argument, the self would have to be the kind of thing that could be given in intuition. He might have said that the minor premise speaks only of entities that are related to the unity of consciousness, not of entities that *might be* given as objects.

Another reason, then, why the argument must be recognized as invalid is that the major premise speaks of things that have the right ontological status to be intuited—that is, things that can be *given* as objects—while the minor premise speaks of nothing of this kind. This does not bring out an ambiguity in the actual wording of Kant's formulation of the paralogism, but, applying a little charity, we might see the envisaged ambiguity as being located in the word "entities" as it figures in the more formally correct formulation of the argument given in section 4 above. Recall that this argument runs:

All entities that cannot be thought otherwise than as subjects are entities that cannot exist otherwise than as subjects, and therefore (by definition) are substances.

All entities that are thinking beings (considered merely as such) are entities that cannot be thought otherwise than as subjects.

Therefore,

All entities that are thinking beings (considered merely as such) are entities that cannot exist otherwise than as subjects, and therefore (by definition) are substances.

According to the line I am suggesting, we should say that in the major premise “entities” is understood to mean “objects,” while in the minor it is understood to mean things of another kind—“subjective unities,” perhaps—understood as the pseudo-objects imagined to account for the unity attaching to each person’s thoughts in virtue of those thoughts’ belonging to one consciousness.

There is no avoiding the conclusion that Kant effectively offers two complementary accounts of where the middle term—the whole phrase: “entities that cannot be thought otherwise than as subjects”—is infected with ambiguity, namely, in the word “thought” and in the word “entities.” But because he does not himself formulate the argument in its strictly correct form, the term “entities” does not occur in the formulation he offers. That being so, the account of the footnote must be adjudged the more satisfactory of the two, and one can suppose that Kant added the footnote in order to remedy this deficiency in the main text.⁵¹

The fourth objection raises a problem for our claim that the major premise of the B-edition paralogism would be true by Kant’s lights.⁵² I have argued that Kant understood the term “paralogism” in such a way that the idea that the premises should be true follows trivially from the very notion of a paralogism. However, one might reasonably wonder how

51. This account of the formal fallacy in the B-edition paralogism differs markedly from Grier’s, which I find to be less successful than her account of transcendental illusion. Grier suggests that the argument should be reconstructed as the invalid inference: “ $\forall x (Ox \supset (Sx \supset Ex))$, Sa; so Ea,” where “Ox” is “x is an object of possible experience”; “Sx” is “x cannot be thought otherwise than as subject,” and “Ex” is “x does not exist otherwise than as subject.” Grier 2001, 163. Grier doesn’t indicate what she takes “a” to be, but one supposes it must be the singular term “the self,” or, perhaps, “the thinking subject (considered merely as such).” This reconstruction has several drawbacks. First, and most obviously, it fails to portray the paralogism as a syllogism. Second, since it portrays the minor premise as a singular judgment, it would seem to fit the first-edition version better than the second. Third, the very suggestion that the first premise incorporates a restriction that is absent from the second seems to run counter to Kant’s words. He describes the first premise as talking of a being that can be thought of “in every respect” and the second as speaking of this being only “relative to thinking and the unity of consciousness.” This suggests that the second premise is in fact more restricted than the first.

52. I am grateful to an anonymous referee for pressing this question and to Kenneth Winkler for discussion of some possible responses. The response I finally settled upon was, in its essentials, suggested to me by Ivan Heyman.

Kant could have taken himself to know the major premise to be true; for while it seems to be intended to be read as a necessary claim, it is, at first sight, neither analytic nor synthetic a priori. I believe this objection can be answered by arguing that Kant would have viewed the major premise as analytic. Since this is far from obvious, however, it will be worth going into this matter in some detail.

Consider the “core” of the major premise, namely: “What(ever) cannot be thought otherwise than as a subject does not exist otherwise than as a subject.” I argued earlier that in order to be consistent, Kant ought to replace “does not” with “cannot.” Applying this emendation and replacing the phrase “otherwise than as a subject” with the (in this context) equivalent phrase “as a determination,” we obtain: “Whatever cannot be thought as a determination cannot exist as a determination.” To appreciate why Kant should have regarded such a claim as analytic it is necessary to attend to the modalities involved. For each occurrence of “cannot,” we must ask whether the modality in question is real or logical impossibility. The answer, I believe, is that the first impossibility is logical, the second real. Thus the claim should be read: “Whatever is such that there is always a contradiction in representing it as a determination of another thing cannot, as a matter of real possibility, exist as a determination of another thing.” Such a claim is analytic because it is analytic of the notion of “real possibility” that whatever is really possible is logically possible, so that whatever is logically impossible is really impossible (compare 29:811). (We might say that x is really possible just in case x is logically possible and x involves no real repugnancy.)

In support of such a reading of the relevant modalities, we may observe that sometimes when Kant speaks of “what can be thought” he clearly has in mind what can be represented as logically possible. So much is clear from, for example, Kant’s remark in the B-edition preface to the first *Critique* that “[Although I cannot cognize freedom as a property of my soul], nevertheless, I can *think* freedom to myself, that is, the representation of it at least contains no contradiction in itself” (B xxviii). This provides some support for the idea that the first impossibility in the major premise is logical impossibility. (Or, more carefully, it suggests that the whole phrase “whatever cannot be thought otherwise than as subject” means “whatever is such that there is always a contradiction in representing it as a determination of another thing.”) That the second impossibility is real is only to be expected since the notion of impossibility that features in the definition of substance (as that which *cannot* exist as a determination of another thing) is plausibly real impossibility.

Something, after all, would seem to still count as a substance₁ if the obstacle to its existing as a determination of another thing were merely real repugnancy rather than logical inconsistency.

Since the major premise is plausibly interpreted as analytic, we can suppose that Kant would have taken it to be true. We can also suppose that he would have taken the rational psychologist to have regarded it as true.

7. A Closer Look at Transcendental Illusion

Before closing it will be useful to address some apparent problems for applying to the Paralogisms chapter the account of transcendental illusion outlined in the opening sections of this essay. That account, I believe, is broadly correct; but it requires some further elaboration and refinement. Two problems need to be addressed: First, although there is much in the first *Critique* to support the idea that the illusion that tempts us to commit the fallacy of the first paralogism is the illusion that *D*—or rather an instance of it⁵³—is true, there are also numerous passages that, on the face of it, suggest a rather different story. Second, the picture we found helpful from the *Prolegomena*—namely, that of a regressive series of conditions whose final member appears to be an ultimate subject—is expressly repudiated at one point in the first *Critique*. The present section is devoted to showing that, in spite of these problems, the essential elements of the account of transcendental illusion given so far do nonetheless apply to the first paralogism.

In the course of defending the completeness of his table of cosmological ideas (stated at B 443) Kant claims:

The category of substance and its accidents is not suited to a transcendental idea, that is, in regard to this category reason has no ground to proceed regressively toward conditions. (A 414/B 441)⁵⁴

53. The instance is: "If a *predicate* is given, the whole series of *subjects* subordinated to one another, a series which is therefore itself unconditioned, is likewise given."

54. Kant goes on to concede that we do have one concept of substance that might "still seem" to be an idea of transcendental reason, namely, the concept of the "substantial." In this context, he understands this to be the concept of a subsisting object that lacks any predicates. A 414/B 441. But since this would not, if it did exist, be an appearance (that is, an object of possible experience), Kant thinks it plain that it cannot be thought of as a final member in a regressive series of *appearances*, which is what would be required in the present context. Note, incidentally, that we may think of the "substantial" as yet a fourth conception of substance recognized by Kant.

Does this mean that the *Prolegomena*'s story is only a temporary aberration, a view held fleetingly between the two editions of the first *Critique*? This seems doubtful since the B-edition retains a number of remarks from the first edition in which the *Prolegomena*'s account is plausibly present. These remarks all occur in the chapter of the first *Critique* dealing with the general notion of a transcendental idea.

First, Kant describes the "totality of conditions and the unconditioned" as the "common title of all concepts of reason" (A 324/B 380). Since one of these "concepts of reason" is the transcendental idea of the self, this idea would seem to have as part of its content the notion of an unconditioned totality of conditions. Second, in the course of arguing that there are just as many species of deceptive syllogistic inference as there are ideas, Kant says that in each species of [dialectical] syllogism "prosyllogisms proceed to the unconditioned: one, *to a subject that is no longer a predicate*, another to a presupposition that presupposes nothing further" (A 323/B 379–80, emphasis added). Here an absolute subject is cited as an "unconditioned," and the idea of a series of conditions is suggested by the idea of prosyllogisms "proceeding" to the unconditioned. Third, in the course of introducing "the system of the transcendental ideas," Kant explains that "all transcendental ideas will be brought under **three classes**, of which the **first** contains the absolute (unconditioned) **unity of the thinking subject**" (A 334/B 391). If the thinking subject's unity is *unconditioned*, it sounds as though its unity might be the stopping point of a regress of conditions.

Together these remarks create the impression that in both editions of the first *Critique* Kant takes seriously at least *some* parts of the *Prolegomena*'s picture of the self as the unconditioned ultimate condition of a regressive series. The passage from the Antinomies chapter also suggests that he takes some of this picture back. So which parts of the *Prolegomena*'s story are supposed to go, and which are supposed to stay? It is not easy to say, but a clue is provided by the prominent position accorded by both editions to the following passage:

There will be as many concepts of reason as there are species of relation represented by the understanding by means of the categories; and so we must seek an **unconditioned, first**, for the **categorical** synthesis in a **subject**, **second** for the **hypothetical** synthesis of the members of a **series**, and **third** for the **disjunctive** synthesis of the parts in a **system**. (A 323/B 380)⁵⁵

55. Compare a partial quotation from reflection 5553:

This remark constitutes the fourth paragraph of the section dealing with the transcendental ideas. It thus occurs in a part of the book in which Kant is laying out the grosser features of the Dialectic's architectonic. Here Kant is trying to forge a connection between, on the one hand, the various kinds of synthesis represented by the ideas and, on the other, the three species of dialectical syllogism—categorical, hypothetical, disjunctive—he takes to be associated with them. The three kinds of “unconditioned” under discussion here are intended to match up with the transcendental ideas of the self, the world, and God.

Kant's words at A 323/B 380 suggest that he conceives of the transcendental idea of the self as the idea of an “unconditioned” associated specifically with *categorical* synthesis. But now he implies that this unconditioned is *not* to be thought of as the first member of a *series*. Instead, that notion is reserved for the unconditioned as it relates to hypothetical synthesis, a form of synthesis that Kant elsewhere implies is associated with the Antinomies (A 406/B 432–33). Given the prominence of this remark, it seems wisest to assume that Kant's more considered strand of thought involves the rejection of the *Prolegomena*'s picture of the self as the final member of a regressive *series*. In the Antinomies Kant explains the reason for this rejection as follows: “Accidents (insofar as they inhere in a single substance) are coordinated with one another, and do not constitute a series” (A 414/B 441).

So, to answer our question about the application of Kant's model of transcendental illusion to the first paralogism: what *goes* (or, at any rate, *should* go) is the idea of a *series* of conditions, while what *stays* (or should stay) is the idea of a subject that is not (and cannot be) a predicate of something else, and is, in that sense, unconditioned. Admittedly, Kant is not entirely consistent on this point,⁵⁶ but this seems like the best way to try to make him consistent.

One says that something is a mere idea if one cannot even approximate to it. 1. The unconditioned of inherence (or of the aggregate). 2. That of dependence or of the series. 3. That of the concurrence of all possibility in one and of one for all. (R 5553; 18:228)

56. The account of transcendental illusion in the introduction to the Dialectic and in the chapter on Ideas is not altogether consistent with Kant's remarks in the Antinomies. For in the two former places he maintains that *all* ideas give rise to regressive *series* of conditions. See especially A 307–8/B 364 and A 336/B 393. It seems likely that in the B-edition Kant simply failed to correct fully the parts of the text he did not wholly redraft.

Getting clearer about these subtle architectonic considerations also helps with our second problem. This, recall, is the worry that in the first *Critique* Kant sometimes characterizes the illusion underlying the paralogisms in terms that make no reference to the apparent truth of *D*—or even to the apparent truth of an instance of *D*. The following remark is typical of those moments:

Nothing is more natural and seductive than the illusion of taking the unity in the synthesis of thoughts for a perceived unity in the subject of these thoughts. (A 402)

What is in question here is an illusion prompting us to mistake one kind of unity for another. This characterization of the illusion corresponds closely to the one given in an important reflection, from which the following two partial quotations are drawn:

The paralogism of pure reason is properly a transcendental subreption [*Subreption*],⁵⁷ where our judgment about objects and the unity of consciousness in them is held to be a perception of the unity of the subject. (R 5553; 18:223)

The first illusion [that is, the one operative in the Paralogisms, as opposed to in the Antinomies or the Ideal of Pure Reason] is that where the unity of apperception, which is subjective, is taken for the unity of the subject as a thing. (R 5553; 18:224)⁵⁸

These remarks suggest, first, that “the unity in the synthesis of thoughts” mentioned at A 402 is the unity of consciousness (“the unity of apperception”), and, second, that the illusion prompting the paralogisms consists

57. By a “subreption” Kant means an instance of the fallacy “*vitium subreptionis*” (the vice of theft or pilfering). One commits this fallacy when one mistakes a thing of one kind for something of quite another kind. Kant’s examples of this fallacy are often cases in which something subjective is mistaken for something objective, for example, a judgment of perception is mistaken for a judgment of experience (24:767), or a concept—“the subjective of thinking”—is mistaken for a thing—“the objective of what is thought.” *Real Progress*, 20:349. Kant also describes as a “*vitium subreptionis*” the fallacy from which arises “the mixing of concepts of experience and of reason” (*Blomberg Logic*, sec. 255; 24:254–55), and here, by a “concept of reason,” he means a pure concept of the understanding. Thus one commits this fallacy when, for example, one takes the concept of cause and effect to be derived from experience. *Ibid.* The reverse error—that of taking a “sensitive concept” for a pure concept of the understanding—also counts as an instance of the subreptive fallacy. See *Inaugural Dissertation*, 2:412. In the *Dissertation*, however, Kant is explicit that he means to be using the term only *by analogy* with its accepted meaning. All of this suggests that a *transcendental* subreption is the error of taking a thing of one kind for a thing of another kind, where this error is grounded in transcendental illusion.

58. Erich Adickes dates this fragment to one of the periods 1778–79 and 1780–83.

in the appearance that this unity is “the unity of the subject as a thing” (R 5553; 18:224).

What does “the unity of the subject as a thing” consist in? I would suggest that, *in the case of the first paralogism*, it is the unity apparently imposed on experience in virtue of the subject’s apparently instantiating the category of “inherence and subsistence.” The notion of “the unity of the subject as a thing” must, however, be interpreted differently in the different paralogisms. That much is clear from Kant’s table at A 404, which sums up the ways in which each of the special categories⁵⁹—namely, the categories of subsistence, reality, unity (as opposed to plurality), and existence (A 403)—is supposed to secure its own kind of “unconditioned unity.”⁶⁰ The table makes plain that each of these categories is supposed to be paired with one of the attributes the rational psychologist ascribes to the self. Thus, while the category of subsistence (and inherence) is deployed in dialectical inferences to the substantiality of the self, that of reality is deployed in dialectical inferences to its simplicity. (Just why this category goes with simplicity is not immediately apparent, as Kant concedes in a footnote [A 404].) For its part, the category of unity (as opposed to plurality) is deployed in inferences to the self’s identity over time, while the category of existence is deployed in inferences to the self’s independence from other things.

This style of diagnosis survives into the B-edition:

Rational psychology has its origin in a mere misunderstanding. The unity of consciousness, which grounds the categories, is here taken for an intuition of the subject as an object, and the category of substance is applied to it. But this unity is only the unity of **thinking**, through which no object is given; and thus the category of substance, which always presupposes a given **intuition**, cannot be applied to it, and hence this subject cannot be cognized at all. (B 421–22)

How does this story about our tendency to confuse one kind of unity with another connect up with the account of illusion given in terms of

59. Kant identifies these four categories as special because each of them is supposed to ground the unity of the other categories in its class. A 403.

60. Rolf-Peter Horstmann sees Kant as countenancing several “senses” of unity: the relational, qualitative, quantitative, and modal, corresponding, respectively, to the substantiality, simplicity, numerical identity, and plain existence of the soul. Horstmann 1993, 417. This is certainly along the right lines. However, because I find it doubtful that Kant would have supposed the German word “unity” to be fourfold ambiguous, I prefer to think of these as four distinct *species* of unity—four ways of being an unconditioned condition.

the appearance that *D*, or an instance of it, is true? The connection is not obvious, but a hint may be gleaned by considering two remarks from the A-edition paralogisms chapter. "Reason," Kant tells us, "represents [the categories of subsistence, reality, and so forth] as conditions of the possibility of a thinking being which are themselves unconditioned" (A 403). He also claims that the self is represented in transcendental illusion as enjoying the "unconditioned unity of **relation**," and he glosses this as meaning "not as inhering but rather **subsisting**" (A 404). Taken together, these remarks suggest that what it is for somethings to seem to have categorially imposed "unconditioned unity" is for it to seem to be unconditioned in some respect. The relevant respect will depend on which paralogism is under consideration. In the second paralogism, for example, the imagined unconditioned unity of the self amounts to its imagined simplicity. In this case, transcendental illusion inclines us to confuse the lack of complexity in the representation "I" with a lack of complexity in a supposed objective self (A 381; compare A 340/B 398; A 355).

In general, then, reason is supposed to generate the illusion that the self instantiates each of the special categories (subsistence, reality, unity [as opposed to plurality], and existence) in virtue of its being an unconditioned condition of the relevant kind (for example, an unbearable bearer of properties, a partless part, and so forth). Owing to transcendental illusion, we are inclined to think of the self not just as unifying its representations but, further, as unifying them in virtue of standing to them as their unconditioned condition. This idea perhaps makes the most sense in connection with the "unconditioned unity of **relation**" since we can see why Kant might have taken transcendental illusion to tempt us to view the self as an unbearable bearer of "the thoughts that are its predicates" (A 346/B 404). In this case transcendental illusion tempts us to confuse the unity of consciousness with the unconditioned unity of relation, and so to suppose that the "I" of apperception represents something that is indeed an unbearable bearer of properties, or substance₁. Thus "the soul cognizes⁶¹ . . . itself, not as inhering but rather as **subsisting**" (A 403–4).

Because the idea of the self as an imagined unconditioned condition remains even after the notion of its being the imagined unconditioned condition *in a series* is downplayed, it is possible to see Kant's various remarks about mistaking one kind of unity for another as indeed

61. This is, of course, a nonfactive use of "*erkennen*."

in keeping with the spirit (if not, admittedly, the letter) of our original account. This is possible because the illusion operative in the first paralogism can now be thought of simply as the illusion that there is an unconditioned condition for any mental predicate in a particular consciousness—that is, a noninhering subject in which they all inhere. Kant supposes that, when we are in the grip of this illusion, we will be inclined to mistake the subjective unity of our thoughts—the unity of apperception—for the objective unity of a self. And this imagined “objective unity” will in this case just boil down to the self’s apparent property of failing to be capable of inhering in another thing. However, because principle *D* makes an essential reference to a *series* of conditions, the present account is compatible only with a modified version of our original account. We must now say that *P* gives rise to the illusion not that *D* is true but that another principle, *D'*, is:

D': If the conditioned is given, then so is its unconditioned condition.

Kant recommends the *Prolegomena*’s account of the paralogisms as superior to that of the first edition of the first *Critique* (4:381); but it is clear from his remarks in the Antinomies that he also entertained doubts about the idea of characterizing transcendental illusion specifically as the illusion that a regressive *series* of conditions has an unconditioned last member. Why should he have entertained these doubts?

Again, Kant does not say; but one possible reason is that such a conception would invite the question: Why isn’t there an *antinomy of the self*? This question would arise because, if the relation of inherence were to generate a regressive series of conditions, the situation would be structurally parallel to the case of an antinomy: we would be able to seek the unconditioned *either* in a final member of the regress *or*, supposing the regress unending, in an infinite totality of members subsequent in the regress to a given conditioned member. Kant in fact indicates his awareness of these two options in the parenthetical remark in the passage from the *Prolegomena*. He says, “Pure reason demands that for each predicate of a thing we should seek its own subject, but that for this subject, which is in turn necessarily only a predicate, we should seek its subject again, and so forth to infinity (*or as far as we get*)” (*Prolegomena*, sec. 46; 4:333, emphasis added). The availability of these two options, each equally recommended by reason, would generate what Kant calls a “two-sided illusion,” and so an antinomy. And yet in the first *Critique* he is keen to maintain that “the transcendental paralogism effects a merely

one-sided illusion regarding the idea of the subject of our thought, and for the opposite assertion there is not the least plausibility forthcoming from concepts of reason" (A 406/B 433). Kant is wedded to this view for architectonic reasons. It is important to him to maintain that there are only four antinomies, one for each "heading" of the table of categories (A 415/B 442; compare A 462/B 490). Kant sees this as important because he wishes to claim that his critical project proceeds systematically. He takes the neat division of antinomies—into two mathematical antinomies and two dynamical ones—as a sign that he has succeeded in identifying *all* the instances of dialectical inference associated with the hypothetical syllogism.

Kant's actual view in the Paralogisms avoids the appearance of a two-sided illusion by dispensing with the picture of a regressive series; but when fully thought through, this change demands a corresponding change in the general account of transcendental illusion. Principle *D'* rather than *D* must now be adopted as the most general statement of the content of transcendental illusion.

Conclusion

I have argued that Kant's diagnosis of the rational metaphysician's dogmatic belief in the substantiality₁ of the self in the B-edition paralogism has a hitherto unappreciated persuasiveness. Through his account of transcendental illusion, Kant has explained why sophisticated philosophers might well have been attracted to this position. And he has, in addition, offered a suggestive diagnosis of the fallacy that constitutes the first paralogism. This diagnosis has two parts: the first turns on the (correct) idea that certain grammatical (or quasi-grammatical) considerations do not, after all, function as a guide to conceivability. That, I take it, amounts to an insight of lasting value. The second turns on the idea that we are apt to confuse the unity of our perceptions and thoughts with the perception of a unity. That idea can be seen as starting from Hume's observation that we never do "catch" ourselves without a perception or idea. But, it goes beyond this in offering a novel, non-Humean diagnosis of why we might nonetheless imagine that we do so.⁶²

More questionable, I believe, is Kant's attempt to portray transcendental illusion, and the attendant doctrine of the transcendental ideas,

62. Hume's own account of why we imagine ourselves to perceive a numerically identical, enduring, simple self is offered in book 1, part 4, section 6 of the *Treatise*. See Hume 2000, 164–71.

as an *exhaustive* account of the origins of speculative metaphysics. Even if we understand that discipline as narrowly as Kant does, such a claim can seem hard to defend. Just confining ourselves to the topic of the self, we might wonder: How can Kant be taken to have diagnosed, for instance, Spinoza's belief that the finite mind is merely a mode of the infinite mind of God? Wasn't Spinoza led by purely rational considerations to the doctrine—a "dogma" by Kant's lights—that the human soul is *not* a substance₁? Or, again: How can he be taken to have diagnosed the Cartesian doctrine that the essence of the self is to think? Kant believes that the self can be known to think but not that it can be known to be incapable of existing without thought. One searches the paralogsms in vain for any diagnosis of how such doctrines might arise.

It seems likely that the pressures of Kant's architectonic play a role here. In respecting the constraints laid down by that architectonic, Kant simply lacks the flexibility to offer diagnoses of the full range of rationalist doctrines. Nonetheless, I think it fair to say that Kant has at least provided an elegant and somewhat plausible account of how in principle philosophers might arrive at the belief that the self is a substance₁. Whether that account accurately describes the actual reasoning of individual proponents of that view is, of course, a further question; but one, unfortunately, that lies beyond the scope of this essay.

References

- Ameriks, Karl. 2000 [1982]. *Kant's Theory of Mind: An Analysis of the Paralogsms of Pure Reason*. 2nd ed. Oxford: Oxford University Press.
- Aquinas, Thomas. 1999. *Commentary on Aristotle's "Physics,"* trans. R. J. Blackwell, R. J. Spath, and W. Edmund Thirlkel. Rev. ed. South Bend, IN: Dumb Ox Books.
- Aristotle. 1984. *The Complete Works of Aristotle: The Revised Oxford Translation*, ed. Jonathan Barnes. Vol. 1. Princeton: Princeton University Press.
- Baumgarten, Alexander. 1902. *Metaphysica*. Reprinted in Immanuel Kant, *Gesammelte Schriften*, vol. 15, part 1 (5–54) and vol. 17 (5–226). Berlin: Königlich-Preussischen Akademie der Wissenschaften zu Berlin, 1902–. Orig. pub. in Halle, 1739, and Frankfurt, 1757. English translation of selected passages in Eric Watkins, *Kant's "Critique of Pure Reason": Background Source Materials*. Cambridge: Cambridge University Press, 2009.
- Bennett, Jonathan. 1966. *Kant's Analytic*. Cambridge: Cambridge University Press.
- . 1974. *Kant's Dialectic*. Cambridge: Cambridge University Press.
- Boehm, Omri. 2009. "Kant's Critique of Spinoza." PhD diss., Yale University.

- Chignell, Andrew. 2007. "Belief in Kant." *Philosophical Review* 116: 323–60.
- Erdmann, Benno. 1881. *Nachträge zu Kants Kritik der reinen Vernunft*. Kiel: Lipsius and Tischer.
- Grier, Michelle. 1993. "Illusion and Fallacy in Kant's First Paralogism." *Kant Studien* 83: 257–82.
- . 2001. *Kant's Doctrine of Transcendental Illusion*. Cambridge: Cambridge University Press.
- Guyer, Paul, ed. 2005. *Notes and Fragments*. The Cambridge Edition of the Works of Immanuel Kant. Trans. Curtis Bowman, Paul Guyer, and Frederick Rauscher. Cambridge: Cambridge University Press.
- Guyer, Paul, and Allen Wood, eds. and trans. 1998. *Critique of Pure Reason*. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press.
- Horstmann, Rolf-Peter. 1993. "Kant's Paralogismen." *Kant-Studien* 84: 408–25.
- Hume, David. 2000. *A Treatise of Human Nature*, ed. D. F. Norton and M. J. Norton. Oxford: Oxford University Press.
- Kant, Immanuel. 1902–. *Gesammelte Schriften*. Berlin: Königlich-Preussischen Akademie der Wissenschaften zu Berlin (now de Gruyter).
- Kemp Smith, Norman, trans. 1933. *Kant's Critique of Pure Reason*. 2nd ed. London: Macmillan. Repr., Boston: Bedford Books, 1969.
- Kitcher, Patricia. 1982. "Kant's Paralogisms." *Philosophical Review* 91: 515–47.
- Longuenesse, Béatrice. 2008. "Kant's 'I think' versus Descartes' 'I am a thing that thinks'." In *Kant and the Early Moderns*, ed. Daniel Garber and Béatrice Longuenesse, 9–31. Princeton: Princeton University Press.
- Pinder, Tillmann. 1987. "Zu Kants Logik-Vorlesung um 1780, anlässlich einer neu aufgefundenen Nachschrift." *Kant-Forschungen* 1: 79–114.
- Spinoza. 1992. *The Ethics; Treatise on the Emendation of the Intellect; Selected Letters*, ed. S. Feldman, trans. S. Shirley. Indianapolis: Hackett.
- van Cleve, James. 1999. *Problems from Kant*. Oxford: Oxford University Press.
- Wilson, Margaret. 1974. "Leibniz and Materialism." *Canadian Journal of Philosophy* 3: 495–513.
- Wolff, Christian. 1747. *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt* [Rational Thoughts on God, the World and the Soul of Human Beings, also All Things in General]. Halle: Renger. Orig. pub. 1720.
- Young, J. Michael, ed. and trans. 1992. *Lectures on Logic*. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press.

