INTRODUCTION

> There exists, then, a natural and unavoidable dialectic of pure reason – not one in which a bungler might entangle himself through lack of knowledge, or one which some sophist has artificially invented to confuse thinking people, but one inseparable from human reason, and which, even after its deceptiveness [*Blendwerk*] has been exposed, will not cease to play tricks with reason and continually entrap it into momentary aberrations ever and again calling for correction. (A298/B355)

The foregoing passage highlights the ostensible purpose of the Transcendental Dialectic in Kant's *Critique of Pure Reason* – to expose the illusion that presumably generates traditional attempts in metaphysics. Kant, of course, is well known as the philosopher who undermined the disciplines of traditional, rationalist, metaphysics. Despite the undeniable influence of his arguments on subsequent philosophical traditions, however, and despite the wealth of secondary literature devoted to these arguments, there remain serious difficulties in interpreting his claims. Part of the problem is that Kant's rejection of the metaphysical arguments is linked up with a unique theory of error. Kant refers to this unique kind of error as "transcendental illusion," and he clearly thinks that it provides an important insight in the propensities of the human mind to engage in speculative metaphysics.

Up to now there has been no sustained and detailed study devoted entirely to examining the role of the doctrine of transcendental illusion

Much of the material in this Introduction has already appeared in "Illusion and Fallacy in Kant's First Paralogism," *Kant-Studien* 83 (1993): 257–282.

in the variety of arguments in Kant's Transcendental Dialectic – an account that explains not only the nature of Kant's objections to the traditional metaphysical inquiries but also the connection between these criticisms and the more general theory of illusion. Superficially, the connection is obvious and well documented. The Dialectic itself is defined by Kant as the "logic of illusion" (A293/B350). The doctrine of illusion thus first emerges in the Dialectic along with Kant's attempt to introduce the third and presumably distinct activity of thought characteristic of "reason." Inherent in the very nature of reason, we are told, is the presumption that objects themselves conform to a rational demand for unconditioned unity. Very generally, Kant's claim is that we unavoidably move from a rational prescription to seek the ultimate explanation (and so a complete "systematic unity" of thought), to the assumption of an "unconditioned," which is given and which systematically unifies things in themselves. The identification of reason as "the seat of" (*als dem Sitze des*) transcendental illusion thus introduces a new and unique possibility for error, error that is distinct from the logical or judgmental error previously discussed in Kant's Transcendental Analytic. The doctrine of illusion thus appears throughout the Dialectic primarily in conjunction with Kant's attempt to undermine the disciplines of traditional (rationalist) metaphysics.

These considerations suggest that Kant's principle aim in the Dialectic is to critique the three central disciplines of "special" metaphysics in the rationalist tradition. Each of these (rational psychology, rational cosmology, and rational theology) attempts to obtain knowledge of a transcendent object by means of formal (transcendental) principles. So, for example, in rational psychology one wants to arrive at substantive metaphysical conclusions about the nature, properties, and constitution of the "soul"; in rational cosmology one wants to arrive at such metaphysical conclusions about the "world"; and in rational theology one wants to do so about "God." Central to Kant's arguments is clearly the view that the metaphysical conclusions in each case are grounded in the "transcendental illusion," which is itself implicit in the very nature of human reason. Kant states the problem in the following important passage:

> These conclusions are . . . to be called pseudo-rational . . . they are not fictitious and have not arisen fortuitously, but have sprung from the very nature of reason. They are sophistications not of men, but of pure reason itself. Even the wisest of men cannot free himself from them. After

> long effort he perhaps succeeds in guarding himself from actual error; but he will never be able to free himself from the illusion [*Schein*] which unceasingly mocks and torments him. (A339/B397)

As this quotation makes clear, Kant's concern is to refute a set of arguments whose conclusions we are in some sense *constrained* to draw. Given this, he needs to show not only that the metaphysical arguments are fallacious but also how they are "rooted" in the nature of reason itself. It is precisely on this latter point, however, that Kant's arguments are commonly thought to fail. The problem is that his subsequent diagnoses of the fallacies of the metaphysical arguments seem to have nothing to do with his more general claims about an "unavoidable illusion."[1] In all cases, rather, the arguments are dismissed on the grounds that they involve an erroneous (i.e., transcendental) application of concepts. Such an account seems incompatible with the claims about illusion, for although Kant repeatedly argues that the metaphysical doctrines are products of an unavoidable, inevitable, and indispensably necessary transcendental illusion, he clearly offers his own critique as a remedy to these very same errors. As we shall see, this complaint is by now commonplace in the secondary literature and reflects the common contention that Kant's general claims about transcendental illusion are inconsistent with his particular criticisms of the fallacies involved in the dialectical syllogisms.

The problems generated by Kant's attempt to link the rejection of metaphysics to a doctrine of transcendental illusion are compounded when we turn to the Appendix to the Transcendental Dialectic. There Kant claims that the very same illusion that was presented to us as the ground of error is nevertheless "indispensably necessary" (A645/B673). More specifically, he argues that the "illusory" demand that an "unconditioned" is actually given, as well as the transcendental ideas of reason which express that demand (the "soul," the "world," and "God"), are required not only for morality but, indeed, for empirical investigations into nature. With this claim, Kant moves from a "nega-

1 This complaint is formulated by Patricia Kitcher, "Kant's Paralogisms," *Philosophical Review* 91, no. 4 (1982): 518; W. H. Walsh, *Kant's Criticisms of Metaphysics* (Edinburgh: Edinburgh University Press, 1975); Norman Kemp Smith, *A Commentary to Kant's Critique of Pure Reason,* 2nd ed., rev. and enlarged (New York: Humanities Press, 1962), p. 457. P. F. Strawson makes essentially the same charge in connection with his discussion of the ideas of reason in *The Bounds of Sense, an Essay on Kant's Critique of Pure Reason* (London: Methuen, 1966); see, e.g., pp. 155–161. This list is by no means exhaustive.

tive" or critical project of "limiting the pretensions to reason" to a "positive" or "constructive" effort to secure for reason some legitimate theoretical function. The problem, once again, is to make sense of Kant's position. Although many commentators have attempted to provide an interpretation that makes sense of Kant's "positive" claims about the role of reason, there continues to be no general agreement on this issue. Moreover, none of these current discussions is concerned to draw a serious connection between the principle of "systematic unity" and the doctrine of transcendental illusion.

My aim here is to elucidate the way in which the doctrine of transcendental illusion simultaneously accommodates Kant's desire to limit the metaphysical "pretensions of reason" and his attempt to defend the necessary (albeit merely regulative) role in empirical knowledge played by this illusion. I begin with a more detailed discussion of the common criticisms offered against Kant. In so doing, I hope to illustrate some of the deeper theoretical reasons Kant has for arguing for a connection between the dialectical attempts of dogmatic metaphysics and his theory of illusion. This connection, in turn, motivates my own attempt to articulate Kant's criticisms in terms of such a theory.

The Inevitability Thesis

One common complaint with respect to Kant's position has to do with what I call his "inevitability thesis," that is, his view that the fallacious inferences involved in each of the dialectical syllogisms are themselves (and because of a transcendental illusion) somehow "natural," "inevitable," and "grounded in the nature of human reason" (A341/B399, A407/B434, A570/B598). In just what sense Kant considers the dialectical inferences to be unavoidable is not immediately clear; as a result, he has been accused of succumbing to hyperbole,[2] of historical prejudice,[3] of indulging in armchair psychology,[4] and of lapsing into incoherence.[5] To be sure, these complaints are not entirely unfounded.

2 Kitcher, "Kant's Paralogisms," p. 518.

3 Walsh, *Kant's Criticism of Metaphysics*, p. 173.

4 Strawson suggests this when he argues (with respect to the idea of God) that there is no reason to think that the idea arises "naturally" in the way that Kant claims (Strawson, *The Bounds of Sense*, p. 222; see also pp. 215–231). Although I cannot go into this issue here, the complaint issues from the attempt to evaluate Kant's claims about the origin of the ideas of reason on "psychological" grounds. On this, see Allen Wood, *Kant's Rational Theology* (Ithaca: Cornell University Press, 1978), esp. p. 62.

5 I take it that this is essentially the point that Jonathan Bennett wishes to make in con-

Prima facie, there seems to be no basis for Kant's claim that the metaphysical conclusions are arrived at "necessarily." Certainly, it is not unreasonable to suggest that the traditional metaphysical doctrines are based simply on erroneous or "bad" argumentation, and not on some mysterious and deep-seated "unavoidable illusion." Although the fallacious arguments may have gained widespread acceptance from Kant's rationalist predecessors, it simply does not follow from this historical fact that the arguments themselves, or the conclusions drawn from these arguments, are always or necessarily encountered.

Indeed, Kant himself is ambiguous on the issue of the inevitability of the metaphysical conclusions. Such ambiguity is apparent in the previously cited passage; Kant wants to hold *both* that the erroneous metaphysical conclusions are somehow inescapable *and* that it is possible to avoid succumbing to the "actual errors" that are involved in accepting such conclusions. In making this last claim, Kant would appear to undermine his own position. *Either,* it would seem, the metaphysical conclusions are "inevitable," in which case the accompanying errors are unavoidable, *or* it is possible to correct, or avoid altogether, such errors. In the first case, the inevitability of the metaphysical conclusions would seem to "undo" Kant's entire critique, which is committed to the possibility of correcting the mistakes of traditional metaphysics through the critical procedure of transcendental reflection. In the latter case, there is little to the suggestion that the erroneous conclusions are themselves inevitable. Considerations such as these make it difficult to understand in exactly what sense, if any, the dialectical conclusions are to be regarded as unavoidable, or what Kant means by the notion of a "transcendental ground," which somehow constrains us to draw such conclusions.[6] Here, then, it appears that the only reasonable interpretation is one that downplays Kant's inevitability thesis.

Nevertheless, problems arise if we fail to consider seriously Kant's claims about the inevitability and necessity of drawing the metaphysical conclusions. In line with this, it should be noted that in all cases these conclusions involve claims about "objects" (the soul, the world, God) that, according to Kant, we must in some way think in order to achieve

nection with his rejection of Kant's attempt to explicate the fallacies of the Dialectic in terms of a general theory of reason. See *Kant's Dialectic* (Cambridge: Cambridge University Press, 1974), pp. 267–288.

6 Kant refers to a "transcendental ground" as necessitating the dialectical inferences in his discussion of the paralogisms (A341/B399). This issue is discussed in Chapter 5, in connection with Kant's rejection of rational psychology.

a completeness and systematicity of knowledge. As Allen Wood notes, the thought of such objects appears to give a certain "completeness" to our knowledge by furnishing the "unconditioned" in relation to various sets of objects that are experienced by us in some sense as "conditioned."[7] The concepts or ideas of such objects, then, function as the epistemological grounds of our knowledge of the actual objects encountered in experience.[8] Indeed, Kant goes so far as to suggest that these transcendental "maxims" of reason are necessary if we are to secure "a criterion" of even empirical truth (A651/B679). This suggests that the transcendental concepts of pure reason play some important role in the domain of empirical knowledge, and this despite Kant's frequent denial that such ideas themselves provide knowledge of anything whatsoever.

This point has been noted by Reinhard Brandt. According to him, the transcendental ideas and their "associated principles" appear to be offered in the *Critique* as "indispensable elements of the possibility of experience."[9] Brandt correctly notes that, as such, these principles must be construed as having a definite transcendental, and not merely logical, status.[10] Although Kant's specific arguments on this score cannot be evaluated until later on, it seems clear even at this point that the criticisms in the Dialectic against metaphysics cannot be read as any straightforward rejection of the conclusions that ground the postulation of transcendent objects.[11]

Kant himself does not want to reject the postulation of such "objects." His views on this issue are presented in the context of his theory

7 Wood, *Kant's Rational Theology*, pp. 17–18.

8 Kant is ambiguous on the issue of whether the transcendent objects represented through the ideas of reason provide ontological or merely epistemological grounds for empirical objects. See Chapters 3 and 8. For a good discussion of Kant's use of both epistemological and ontological senses of the "thing in itself," see Bernard Rousset, *La Doctrine kantienne de l'objectivité* (Paris: Librairie Philosophique J. Vrin, 1967), chap. 6.

9 Reinhard Brandt, "The Deductions in the Critique of Judgment: Comments on Hampshire and Horstmann," in *Kant's Transcendental Deductions*, ed. Eckhart Förster (Stanford: Stanford University Press, 1989), p. 178. Note that the necessary status of the ideas and principles of reason is an issue usually discussed in connection with Kant's philosophy of science. See Thomas Wartenberg, "Order through Reason," *Kant-Studien* 70 (1979): 409–424, and Gerd Buchdahl, *Metaphysics and the Philosophy of Science* (Cambridge, Mass: MIT Press, 1969), esp. pp. 523–530. This topic is discussed in Chapter 8.

10 Brandt, "The Deductions in the Critique of Judgment: Comments on Hampshire and Horstmann," p. 178.

11 Robert B. Pippin recognizes this problem and discusses it in chapter 7 of his book, *Kant's Theory of Form* (New Haven: Yale University Press, 1982), esp. pp. 193–215.

of the ideas of reason, which provides the framework within which Kant assesses and rejects the various disciplines of special metaphysics. Although Kant wants to argue against the attempt to acquire metaphysical knowledge of these objects, he continues to maintain the necessity of postulating them in thought – an approach particularly evident in the Appendix to the Transcendental Dialectic. Having argued in relation to each of the ideas that the transcendent employment of reason is dialectical, Kant undertakes to secure for reason some "good and proper" employment (A643/B671). Here the "natural tendency" to transgress the limits of possible experience is at issue. When such an attempt is undertaken with a view to yielding knowledge of an object (i.e., transcendently), then reason can be shown to defy the very conditions required for its success, for the ideas do not have any real object corresponding to them. If, however, in passing beyond the sphere of possible experience, the ideas are deployed as devices for directing the proper employment of the understanding (i.e., regulatively), then the use of reason has positive results (it provides unity of the understanding) and is deemed "indispensably necessary" (A645/B673).

Once again, as Brandt notes, Kant seems to hold that without the ideas of reason, the acts of the understanding – and, indeed, the categories themselves – are "incoherent and useless."[12] Presumably, the usefulness of the ideas and principles of reason issues from their legislating capacity. Kant's view is that reason itself prescribes that we seek knowledge in accordance with certain goals and interests, which in turn define what will count as knowledge in the first place. Although the ideas that express these interests of reason may be "illusory," they are nevertheless taken to be necessary presuppositions in the acquisition of knowledge. This last view, which is considered in Chapter 8, reflects Kant's view that the body of knowledge is to be understood as an active "project" undertaken in light of the "subjective" interests definitive of human reason.

Given these considerations, we may distinguish between the *negative* critique of the particular metaphysical arguments and the *positive* account of the principles and maxims of reason. But it is imperative to see that these two undertakings are, for Kant, inextricably bound up with one another. His position that the ideas of reason are necessary and unavoidable means that we will forever be tempted to regard them as ob-

12 Brandt, "The Deductions in the Critique of Judgement: Comments on Hampshire and Horstmann," pp. 178–179.

jects of possible knowledge. In this case, the doctrine of transcendental illusion is absolutely central to his account of metaphysical error. Not only does the doctrine provide the framework within which Kant first introduces the problems of the Dialectic, but each of the disciplines subsequently criticized is repeatedly held to involve such illusion. The transcendental paralogism, for example, is defined as a fallacious syllogism that is "grounded in the nature of human reason, and which gives rise to an illusion which cannot be avoided" (A341/B399). Both the A and B edition versions of the paralogisms end with a general exposition of the "transcendental and yet natural illusion in the paralogisms of pure reason" (A396–397 and also B427). Lest this be assumed to be peculiar to the paralogisms, it should be noted that equal emphasis is placed on the role of transcendental illusion in Kant's criticisms of the cosmological and theological arguments as well.[13] It seems strange, then, that one is hard-pressed to find in the secondary literature any detailed, "full-scale" investigation into Kant's doctrine of transcendental illusion.[14]

Transcendental Illusion

Kant generally identifies transcendental illusion with the propensity to take the subjective or logical requirement that there be a complete unity of thought to be a requirement to which "objects" considered independently of the conditions of experience (things in themselves) must conform (A297/B354). In accordance with this, Kant suggests, we move from the subjective or logical requirement for systematic unity of thought to the assumption of an "unconditioned," which is given independently of the subjective conditions of experience. It is precisely this assumption that, Kant will hold, generates metaphysical error. Central to his position are two claims: the requirement for systematic unity of

13 See, e.g., A422/B450, A484/B512, A582/B610. This list is by no means exhaustive.

14 I by no means wish to suggest that the topic of transcendental illusion has escaped discussion altogether, but only to point out that it has not received the kind of detailed attention that has succeeded in making its role in Kant's philosophy clear. I am indebted to many previous helpful discussions. See Robert Theis, "De l'illusion transcendentale," *Kant-Studien* 76 (1985): 119–137; Robert Butts, "Kant's Dialectic and the Logic of Illusion," in *Logic and the Workings of the Mind,* ed. Patricia Easton, North American Kant Society Studies in Philosophy, vol. 5 (Atascadero, Calif.: Ridgeview, 1997); Karl Ameriks, "The Critique of Metaphysics: Kant and Traditional Ontology," in *The Cambridge Companion to Kant,* ed. Paul Guyer (Cambridge: Cambridge University Press, 1992). My indebtedness to these and other works will become clear in what follows.

thought is a necessary rational requirement; and such a requirement is projected by us as holding "objectively," of objects themselves. This last feature presumably accounts for the illusory nature of the error. Although Kant is rarely taken seriously on this issue, the coherence of his position would seem to depend on the claim that an "inescapable illusion" somehow necessitates the attempt to move from the conditioned to the unconditioned. Although reason's attempt to pass from the "conditioned" to the "unconditioned" is said by Kant to be "unavoidable," he maintains that the transition also generates an illegitimate application of the categories, an application that is illegitimate because it moves beyond all possible experience.

Even an abbreviated account of this illusion reveals a number of important points. First, Kant wants to use the doctrine of transcendental illusion to provide a unified account of the way in which the misapplication of the categories generates metaphysical (synthetic a priori) claims about transcendent objects. Second, despite this connection, Kant may not identify the illusion with the fallacious application of the categories. This point is made in the opening sections of the Dialectic, when Kant explicitly distinguishes between the misemployment of the categories and transcendental illusion. The former is characterized as an error in judgment, issuing from a certain misemployment of the *understanding* (cf. A296/B353); transcendental illusion, on the other hand, involves the use of the transcendent ideas, maxims, and principles of *reason* (A297/B354).

Most commentators overlook the distinction between the illusions that ground the fallacies of the Dialectic and the actual fallacies themselves.[15] Indeed, only on the assumption that the two are the same, or are supposed to explain the very same error, do the charges of inconsistency make any sense. Accordingly, I hope to show that Kant's arguments require that we draw a distinction between transcendental illusion and the fallacies that presumably emerge in conjunction with it. Such a distinction, in turn, suggests an obvious resolution to the prob-

15 Thus, as we have seen, Kant is oftentimes accused of inconsistency precisely because his description of the illusion is not the same as his account of the fallacies. I have already mentioned Patricia Kitcher (see *Kant's Transcendental Psychology* [Oxford: Oxford University Press, 1990], p. 185) and Kemp Smith (*Commentary*, p. 457). However, the failure to distinguish the illusions and the fallacies is fairly widespread, even among Kant's defenders. See Henry Allison, *Kant's Transcendental Idealism* (New Haven: Yale University Press, 1983) pp. 282–283; Wood, *Kant's Rational Theology*, p. 76; Karl Ameriks, *Kant's Theory of Mind* (Oxford: Clarendon Press, 1982), pp. 55–57.

lems encountered in connection with Kant's inevitability thesis; for although the illusions of the Dialectic are "inescapable," "inevitable," and "necessary,"[16] the fallacies or judgmental errors inherent in the metaphysical arguments are not. That Kant wants to distinguish between the illusions and the fallacies of the Dialectic is further evidenced by his division of the Dialectic into two books. Whereas the first of these identifies and accounts for the transcendental concepts of pure reason (the ideas), the second is explicitly concerned to critique the dialectical (fallacious) inferences drawn in accordance with such "necessary ideas" (A309/B366).

The interpretation presented here operates on the assumption that, for Kant, transcendental illusion is not necessarily or in itself deceptive, although, in combination with a transcendental misapplication of the categories, it grounds certain fallacious inferences that are.[17] That Kant does not consider the unavoidable illusion to be in itself or necessarily deceptive is clear from his frequent use of optical analogies. Consider the following:

> Transcendental Illusion [*Schein*] . . . does not cease even after it has been detected and its invalidity clearly revealed by transcendental criticism. . . . This is an illusion [*Illusion*] which can no more be prevented than we can prevent the sea from appearing higher at the horizon than at the shore; . . . or to cite a still better example, than the astronomer can prevent the moon from appearing larger at its rising, although he is not deceived [*betrogen*] by this illusion. (A297/B354)

> The transcendental dialectic will therefore content itself with exposing the illusion [*Schein*] of transcendent judgments, and at the same time take precautions that we be not deceived [*betruge*] by it. (A298/B355)

Note that while Kant considers the illusion that grounds the metaphysical move to the idea of the unconditioned to be both unavoidable and epistemologically necessary, he refers to this idea as a "*focus imaginarious,*" suggesting that it functions merely as a theoretical "point" toward which our inquiries are to be directed (A645/B673). Accordingly,

16 At A298/B355 Kant explicitly claims that it is the illusion that is natural and inevitable. I hope to show that this is his consistently held view.

17 Meerbote distinguishes between deceiving and nondeceiving semblance in his introduction to the translation of Kant's "Concerning Sensory Illusion and Poetic Fiction." See *Kant's Latin Writings, Translations, Commentaries and Notes,* ed. L. W. Beck (New York: Peter Lang, 1986), pp. 193–201.