1
An Introduction to the Problem

The aim of this work is to provide both an interpretation and, where possible, a defense of Kant's transcendental idealism. Since this idealism is inseparable from Kant's views on the nature, conditions, and limits of human knowledge, as well as his critique of other philosophical positions, this project involves a discussion of many of the central topics of the *Critique of Pure Reason.*[1] Some familiar and important topics are omitted, however, both in order to keep the focus as much as possible on Kant's idealism and to allow room for a sufficiently detailed treatment of those topics that are covered. Thus the work also can be characterized as a more or less comprehensive study of Kant's theoretical philosophy, organized around the theme of transcendental idealism. It differs from other treatments of Kant in the recent literature first in its emphasis on the connection between Kant's idealism and his substantive claims, and second in the philosophical weight that it gives both to this idealism and to these claims. Unlike most writers on Kant, I take much of the *Critique* to be not only "interesting" or to "contain more of value than is sometimes supposed," but to be philosophically defensible. At the very least, I believe that with a bit of help from the sympathetic interpreter it can be defended against many of the familiar criticisms that are repeatedly presented as "devastating."

As a first step in this admittedly ambitious project, I shall briefly characterize what I take to be the standard picture of Kant's idealism (which is the source of the familiar criticisms) and attempt to indicate its inadequacy as an account of what Kant actually maintained. I shall then introduce and discuss in a preliminary way the conception of an epistemic condition. My claim is that this conception, although merely implicit in the *Critique,* is the real key to the understanding of transcendental idealism, and with it Kant's philosophical achievement. This will provide the basis for the more extended discussion, in the next two chapters, of transcendental idealism and of the frequently misunderstood argument that Kant advances in support of it in the Antinomy of Pure Reason.

I. THE STANDARD PICTURE AND ITS INADEQUACY

According to the standard picture, Kant's transcendental idealism is a metaphysical theory that affirms the unknowability of the "real"(things in

themselves) and relegates knowledge to the purely subjective realm of representations (appearances). It thus combines a phenomenalistic account of what is actually experienced by the mind, and therefore knowable, with the postulation of an additional set of entities which, in terms of the very theory, are unknowable. In spite of the obvious difficulties that it creates, this postulation is deemed necessary to explain how the mind acquires its representations, or at least the materials for them (their form being "imposed" by the mind itself). The basic assumption is simply that the mind can only acquire these materials as a result of being "affected" by things in themselves. Thus, such things must be assumed to exist, even though the theory denies that we have any right to say anything about them (presumably including the claim that they exist).

Although this picture, which can be traced back to Kant's own contemporaries,[2] has been repeatedly criticized, it is still widely accepted, especially in the Anglo-American philosophical world. To a considerable extent this acceptance is due to the influence of P. F. Strawson, who, echoing the standard picture, defines transcendental idealism as the doctrine that "reality is supersensible and that we can have no knowledge of it."[3] Starting with this understanding of Kant's idealism, Strawson sets as his avowed task the separation of what he terms the "analytic argument" of the *Critique* from the transcendental idealism with which he believes Kant unfortunately and unnecessarily entangled it.[4] In the latter respect he has been followed by numerous commentators who have tried to formulate and defend some vaguely Kantian "transcendental arguments" that are uncontaminated by any idealistic premises.[5]

But Strawson not only rejects transcendental idealism as incoherent and attempts, as it were, to save Kant from himself; he also provides an account of what led Kant to this "disastrous" doctrine. As Strawson sees it, transcendental idealism is the direct consequence of Kant's "perversion" of the "scientifically minded philosopher's" contrast between a realm of physical objects composed of primary qualities and a mental realm consisting of the sensible appearances of these objects (including their secondary qualities). This mental realm, like its Kantian counterpart, is thought to be produced by means of an affection of the mind, in this case by physical objects. Kant allegedly perverts this model by assigning the whole spatiotemporal framework (which according to the original model pertains to the "real," that is to say, to physical objects) to the subjective constitution of the human mind. The resulting doctrine is judged to be incoherent because, among other reasons, it is with reference only to a spatiotemporal framework that one can talk intelligibly about "affection."[6]

Although Strawson himself does not put it in quite this way, the usual manner of making essentially the same point is to claim that Kant is an inconsistent Berkeley.[7] The Berkeleian element consists of Kant's subjectivism, namely, the limitation of knowledge to appearances, with these being understood as "mere representations." The alleged inconsistency stems from Kant's combination of his essentially Berkeleian phenomenalistic idealism with his postulation of an inaccessible realm of things in themselves. This conception of what Kant was up to generates, in turn, the standard criticisms, many of which are reflected in Strawson's account. I shall deal with those criticisms that are directed against the Kantian conception of the thing in itself and the associated doctrine of affection in chapter 11. For the present I wish to consider merely those which concern the claim that we know only appearances.

Since it equates 'appearance' with 'mere representation', the standard picture takes Kant's claim that we know only appearances to mean that we know only the contents of our minds, that is, ideas in the Berkeleian sense. This reading of Kant is then sometimes used as the basis for a critique of the doctrine of the ideality of space and time, which Kant presents in the Transcendental Aesthetic. Simply put, the claim is that Kant's subjectivistic starting point forces him to choose between the following equally unpalatable alternatives: either (1) he must maintain that things only *seem to us* to be spatial (or temporal), a doctrine which entails that our consciousness of a world of objects extended and located in space is somehow illusory; or (2) he must claim that appearances, that is to say, representations, really are spatial, a doctrine which is absurd because it requires us to regard mental items as extended and as located in space.

Although this line of criticism has echoes in Strawson,[8] it has been developed most fully by H. A. Prichard, who concentrates much of his attack on the alleged incoherence of Kantian "appearance talk." According to Prichard's highly influential critique, Kant's whole conception of appearance is vitiated by a confusion of the claim that we know only *things as they appear to us* with the quite different claim that we know only a particular class of things, namely, *appearances.* He also suggests that Kant's tendency to slide from one of these claims to the other prevented him from confronting the dilemma posed by the abovementioned alternatives. Thus, on his reconstruction, what Kant really wished to claim is that we know things only as they appear to us; but since this, according to Prichard, entails that these things only seem to us to be spatial (the illusion thesis), in order to defend his cherished empirical realism, Kant is forced to shift to the doctrine that we know appearances and that they really are spatial.[9]

The most basic and prevalent objection stemming from the standard picture is that by limiting knowledge to appearance, that is, to the subjective realm of representations, Kant effectively undermines the possibility of any genuine knowledge at all. In short, far from providing an antidote to Humean skepticism, as was his intent, Kant is seen as a Cartesian

skeptic *malgré lui.* Some version of this line of objection is advanced by virtually every proponent of the standard picture, including Strawson.[10] Once again, however, the sharpest formulation is provided by Prichard, whose account can be taken as paradigmatic for the standard picture.[11] Prichard construes Kant's distinction between appearances and things in themselves in terms of the classic example of perceptual illusion: the straight stick that appears bent to an observer when it is immersed in water. Given this analogy, he has little difficulty in reducing to absurdity Kant's doctrine that we know only appearances. His analysis proceeds through various stages, but the main point is simply that this claim is taken to mean that we can know things only as they "are for us" or "seem to us" (in virtue of the distortion imposed by our perceptual forms), not as they "really are." Since to know something, according to Prichard, just means to know it as it really is, it follows that for Kant we cannot really know anything at all. Clearly, such a conclusion amounts to a *reductio* of the Kantian theory.

It seems obvious that, if this is how Kant's transcendental idealism is really to be understood, the Strawsonian project of trying to locate in the *Critique* a philosophical core that can be neatly separated from the idealistic trappings is very attractive. Indeed, it presents itself as the only philosophically fruitful way of dealing with Kant's thought. Nevertheless, in spite of the fact that it does seem to have some textual support, one can raise serious doubts about the adequacy of this interpretation, which is so frequently accepted as a matter of course. The root of the problem is that it tends to neglect altogether, or at the very least to minimize, certain distinctions that are central to Kant's whole transcendental enterprise.

Specifically, it fails to distinguish sharply between the empirical and the transcendental versions of two generally acknowledged and closely related distinctions. These are the distinctions between ideality and reality and between appearances and things in themselves. The issues here are complex, and at this point I can only attempt to provide a rough sketch of what these distinctions involve. I believe, however, that even this rough sketch should suffice to demonstrate the inadequacies of the standard picture as an interpretation of Kant's actual teaching.[12]

'Ideality', in the most general sense in which Kant uses the term, signifies mind dependence or being in the mind (*in uns*); while 'reality' (*Realität*), in the sense in which it is opposed to 'ideality', signifies independence of mind or being external to the mind (*ausser uns*).[13] In both the Transcendental Aesthetic and the Transcendental Dialectic, Kant distinguishes between an empirical and a transcendental sense of 'ideality', and, by implication at least, of 'reality'. Taken in its empirical sense, 'ideality' characterizes the private data of an individual mind. This includes ideas in the Cartesian-Lockean sense or, more generally, any mental content in the ordinary sense of 'mental'. 'Reality', construed in the empirical sense, refers to the intersubjectively accessible, spatiotemporally ordered realm of objects of human experience. At the empirical level, then, the ideality-reality distinction is essentially between the subjective and the objective aspects of human experience. When Kant claims that he is an empirical realist and denies that he is an empirical idealist, he is really affirming that our experience is not limited to the private domain of our own representations, but includes an encounter with "empirically real" spatiotemporal objects.

The transcendental version of the distinction is quite another matter. At the transcendental level, which is the level of philosophical reflection upon experience (transcendental reflection), 'ideality' is used to characterize the universal, necessary, and, therefore, a priori conditions of human knowledge.[14] In the Transcendental Aesthetic, Kant affirms the transcendental ideality of space and time on the grounds that they function as a priori conditions of human sensibility, that is, as subjective conditions in terms of which alone the human mind is capable of receiving the data for thought or experience.[15] He terms these conditions "forms of sensibility." Things in space and time (empirical objects) are ideal in the same sense because they cannot be experienced or described independently of these sensible conditions. Correlatively, something is real in the transcendental sense if and only if it can be characterized and referred to independently of any appeal to these same sensible conditions. In the transcendental sense, then, mind independence or being external to the mind (*ausser uns*) means independence of sensibility and its conditions. A transcendentally real object is thus, by definition, a nonsensible object or noumenon.[16]

The transcendental conception of ideality provides the basis for the transcendental conception of appearance and for the transcendental version of the contrast between appearances and things in themselves. Thus, to speak of appearances in the transcendental sense is simply to speak of spatiotemporal entities (phenomena), that is, of things insofar as they are viewed as subject to the conditions of human sensibility. Correlatively, to speak of things in themselves transcendentally is to speak of things insofar as they are independent of these conditions. In several places Kant insists upon the importance of not confusing this distinction with its empirical counterpart. One of the clearest of these is in "On the Progress of Metaphysics," where, in a discussion of the transcendental ideality of space and time, Kant writes:

> Furthermore, it is to be noted that appearance, taken in the transcendental sense, wherein it is said of things that they are appearances (phenomena), means something completely different than when I say, this thing appears to me in some manner or other, which should designate appearance in the physical sense, and which can be called semblance [*Apparenz*] and illusion [*Schein*]. For although these objects of the senses are mere appearances, since I can only compare them with other sensible objects . . . by the lan-

guage of experience they are nevertheless thought as things in themselves. Thus, if it is said of such a thing that it has the look [*Anschein*] of an arch, in this context the seeming refers to the subjective aspect of the representation of a thing, which can be a cause for it to be falsely taken in a judgment as objective. And, therefore, the proposition that all sensible representations only yield knowledge of appearances is not at all to be equated with the claim that they contain only the illusion [*Schein*] of objects, as the idealist will have it.[17]

The "language of experience," of which Kant speaks here, includes both ordinary and scientific experience. Both involve a distinction between those properties that a given object actually possesses and those it merely seems to possess for a particular observer under certain empirically specifiable conditions. The object as it "really is" (with its actual properties) is the thing in itself in the physical or empirical sense, while the representation of the object possessed by a particular observer under given conditions is what is meant by the appearance or semblance of the object. The main point here is simply that at the empirical level, or in "the language of experience," 'appearances' and 'things in themselves' designate two distinct classes of entity with two distinct modes of being. The members of the former class are "mental" in the ordinary (Cartesian) sense and the members of the latter are "nonmental" or "physical" in the same sense. At the transcendental level, however, things are quite different. There the distinction between appearances and things in themselves refers primarily to two distinct ways in which things (empirical objects) can be "considered": either in relation to the subjective conditions of human sensibility (space and time), and thus as they "appear," or independently of these conditions, and thus as they are "in themselves." Indeed, as Gerold Prauss has pointed out, when Kant is concerned with articulating the transcendental sense of his distinction, he usually does not use such expressions as *Ding an sich, Ding an sich selbst,* or *Sache an sich;* rather, he uses locutions, such as *Ding* or *Sache an sich selbst betrachtet.*[18]

It is certainly possible to detect a dim grasp of the distinction between the transcendental and the empirical conceptions of appearance in Prichard's contrast between things as appearing and appearances. Transcendental-level talk about appearances can be described as talk about things as appearing. Similarly, talk about appearances belongs naturally to the "language of experience." The problem here lies in Prichard's contention that Kant slides from one notion to the other. Given the preceding analysis, this is equivalent to the claim that Kant systematically confuses the transcendental and the empirical versions of his basic distinction. This is itself highly implausible, especially in light of Kant's frequent efforts to distinguish between these two senses of 'appearance'. Even apart from this, however, it can easily be shown that Prichard is guilty of the very

confusion of which he accuses Kant. We have seen that part of Prichard's basic objection to what he views as Kant's empirical realism is that it involves the absurd notion that appearances (mental contents) are spatial (extended). Kant is thus judged guilty of spatializing sensations, a charge that with much greater propriety can be directed against Hume. But obviously this "absurdity" arises only if Kant's claim about the spatiality of appearances is taken in the empirical sense. If, as Kant clearly wishes us to do, we construe claims about the spatiality of appearances in the transcendental sense, the absurdity disappears; for then spatiality (together with temporality) can be seen as a defining characteristic of things considered as they appear, not as a property mysteriously attributed to sensations.

The objection to Kant's alleged skepticism can be dealt with in a similar fashion.[19] It is clear from his use of the bent stick analogy that Prichard construes the distinction between appearances and things in themselves in the empirical sense. This, in turn, enables him to take Kant to be claiming that we can know only how things seem (appear) to us, which entails the skeptical conclusion. It is by no means certain, however, that this follows if we construe Kant's claim about the limitation of knowledge to appearances in the transcendental rather than in the empirical sense. Understood in this sense, which is the sense in which Kant intended it, it is an epistemological claim about the dependence of human knowledge on certain a priori conditions which reflect the structure of the human cognitive apparatus. These conditions do not determine how objects "seem" to us or "appear" in the empirical sense; rather, they express the universal and necessary conditions in terms of which alone the human mind is capable of recognizing something as an object at all. Thus the doctrine that we can know things only as they appear, not as they are in themselves, can be regarded as equivalent to the claim that human knowledge is governed by such conditions. If, in fact, there are such conditions, and if they function in the ways in which Kant contends, then it hardly makes sense to accuse him of being a skeptic because he denies the possibility of knowledge of things as they are independently of them, that is, of things as they are in themselves.

To say this is not, of course, to endorse Kant's account. We will not be in a position to evaluate Kant's claims regarding the a priori conditions of human knowledge until we have examined the arguments of the Transcendental Aesthetic and Transcendental Analytic. Nevertheless, it is not necessary to do so in order to realize the inappropriateness of the skepticism objection as formulated by Prichard and other proponents of the standard picture. The problem with this objection is that it fails completely to come to grips with Kant's intent, and thus to see what his transcendental claims actually involve. Instead, these claims are routinely interpreted as empirical or quasi-empirical. Similarly, Kant's talk about

the "conditions" of human knowledge is taken in a psychological sense. The inevitable consequence of this is that Kant is seen as a proponent of the very empirical idealism which he took such great pains to repudiate.

II. THE CONCEPT OF AN EPISTEMIC CONDITION

The interpretation of transcendental idealism which I hope to develop in this study will, in contrast to the standard picture, emphasize its connection with Kant's claims regarding the conditions of human knowledge: I shall argue that the claim that human knowledge has such conditions is the distinctive, indeed, the revolutionary thesis of Kant's philosophy, and that transcendental idealism is at bottom nothing more than the logical consequence of its acceptance. So far, however, the crucial notion of 'condition' has remained undefined, even unexamined. Any number of things—for example, the brain, the central nervous system, sense organs, and so forth—could legitimately be described as conditions of human knowledge; yet none of these would have very much to do with Kant's central claim or with transcendental idealism. In an effort to clarify and to pinpoint the relevant sense of the term, I propose to introduce the notion of an *epistemic condition.*[20]

Even though this notion is central to Kant's whole transcendental enterprise, the fact that he never explicitly deals with it makes it difficult, if not impossible, to define in any very precise way. For present purposes, then, it must suffice to characterize an epistemic condition simply as one that is necessary for the representation of an object or an objective state of affairs. As such, it could also be called an "objectivating condition"; for it is in virtue of such conditions that our representations relate to objects or, as Kant likes to put it, possess "objective reality." In this respect epistemic conditions are to be distinguished from what Kant terms "logical conditions of thought," for example, the principle of contradiction. The latter serves as a rule of consistent thinking, but not for the representation of objects. Thus it is not an epistemic condition in the sense in which this notion is taken here. Roughly speaking, the distinction between logical and epistemic conditions reflects Kant's own distinction between general and transcendental logic. In fact, the main business of transcendental logic is to establish a set of epistemic conditions, namely, the pure concepts of the understanding.

In addition to the pure concepts of the understanding, which Kant defines as "concepts of an object in general," space and time (the forms of human sensibility) must also be regarded as epistemic conditions. Although together these two types of condition constitute what Kant himself terms "necessary conditions of the possibility of experience," there are two reasons for believing that the broader notion of an epistemic condition better captures the essential thrust of Kant's thought. The first is simply that Kant is not solely, or even primarily, concerned with experiential knowledge. Epistemic conditions must, therefore, also figure in the Kantian account of nonempirical knowledge, that is, of mathematics and metaphysics. Indeed, as we shall see in chapter 6, the pure concepts in their "logical use" can even be regarded as epistemic conditions of analytic judgments.

The second and main reason for talking about epistemic conditions, rather than about conditions of the possibility of experience, is that this term makes it easier to grasp the difference between this central Kantian conception and other senses of 'condition' with which it is frequently confused. Such confusions are reflected in many of the stock criticisms of Kant, including those discussed in the previous section, as well as in the standard picture as a whole. Moreover, many of Kant's own criticisms of other philosophical positions can be seen to turn on the claim that they involve confusions of what are here called epistemic conditions with conditions of other sorts. One of these is, of course, the confusion of merely logical with epistemic conditions, which figures largely in Kant's polemic with Leibnizian rationalism. This distinction has already been noted and needs no further comment at this point. For the present, it is important to distinguish epistemic conditions from psychological conditions on the one hand and from ontological conditions on the other.

By a psychological condition I mean some mechanism or aspect of the human cognitive apparatus that is appealed to in order to provide a genetic account of a belief or an empirical explanation of why we perceive things in a certain way. This can be understood to include physiological as well as narrowly psychological factors. Custom or habit, as used by Hume in his account of causality, is a prime example of such a psychological condition. As is well known, Kant was insistent in claiming that, although the appeal to such factors may be necessary to explain the origin of our beliefs and perceptions, or even of our knowledge "in the order of time" (*der Zeit nach*), it cannot account for its objective validity. In Kant's terms it can answer the *quaestio facti* but not the *quaestio juris*. The latter is the proper concern of the *Critique,* and this requires an appeal to epistemic conditions.[21] In fact, Kant's basic charge against Hume is that he confuses these questions and thus, implicitly at least, these two kinds of conditions. The clearest example of this is Kant's claim in the *Prolegomena* that, in his analysis of causality, Hume "mistook a subjective necessity (habit) for an objective necessity arising from insight."[22] Ironically enough, this very same line of criticism is frequently raised against the *Critique* by critics who find in it a dangerous subjectivism.

It is equally important to distinguish epistemic from ontological conditions. By the latter I mean conditions of the possibility of the being of things. Since the being of things is here contrasted with their being known, an ontological condition is, by definition, a condition of the possibility of things as they are in themselves (in the transcendental sense).

Newtonian absolute space and time are clear examples of conditions of this sort. Kant describes them as "two eternal and infinite self-subsistent [*für sich bestehende*] non-entities . . . which are there (yet without there being anything real) only in order to contain in themselves all that is real" (A39/B56). In a Second Edition addendum to the Transcendental Aesthetic, Kant indicates the dangerous theological consequences of such a view by pointing out that "as conditions of all existence in general, they must also be conditions of the existence of God" (B71). Even apart from theology, however, Kant believes that this conception of space and time leads to absurdities:

> For if we regard space and time as properties which, if they are to be possible at all, must be found in things in themselves, and if we reflect on the absurdities in which we are then involved, in that two infinite things, which are not substances, nor anything actually inhering in substances, must yet have existence, nay, must be the necessary condition of the existence of all things, and moreover, must continue to exist, even although all existing things be removed,—we cannot blame the good Berkeley for degrading bodies to mere illusion. Nay, even our own existence, in being made thus dependent upon the self-subsistent reality of a non-entity, such as time, would necessarily be changed with it into sheer illusion—an absurdity of which no one has yet been guilty. [B70–71][23]

Kant's point is that, for all its absurdity, Berkeley's idealism, which he interprets as involving the denial of the reality of material objects ("degrading bodies to mere illusion") makes a certain amount of sense when it is viewed as a response to Newton.[24] This is because once the empirical reality of material objects and persons is seen to depend on the absolute (transcendental) reality of space and time, the absurdities connected with the latter make it plausible to deny the former. But this consequence can easily be avoided if, instead of viewing space and time with Newton as conditions of the possibility of things in themselves, we view them as conditions of the possibility of our knowledge or experience of things. In Kantian terms, instead of being "two eternal and infinite, self-subsistent non-entities," they now become two "sources of knowledge" (*Erkenntnisquellen*) (A38/B55). By analyzing the situation in this way, Kant claims to be able not only to distinguish his idealism from that of Berkeley, but also to provide a critical alternative to both the Newtonian and the Leibnizian conceptions of space and time.

Consequently, just as Kant's strategy with Hume is to show that the skeptical consequences of his analysis stem from a confusion of psychological and epistemic conditions, his strategy with Newton is to show that the untenable consequences of the latter's theory of space and time are the results of a confusion of epistemic with ontological conditions. We shall also see that this mode of analysis can be applied to the Kantian critique of many other thinkers. For the present, however, the key point to note is the connection between these two distinctions and the consequent confusions. Although the above account may have suggested otherwise, it is not simply the case that there are some philosophers who just happen to be guilty of one of these confusions and others who just happen to be guilty of the other. The fact of the matter is rather that, from a Kantian standpoint, the two kinds of confusion represent two sides of the same coin, namely, the failure to recognize the role in human knowledge of a set of distinctively epistemic conditons.

Indeed, one can claim that the fundamental issue raised by the *Critique* is whether it is possible to isolate a set of conditions of the possibility of knowledge of things (in the sense already indicated) that can be distinguished from conditions of the possibility of the things themselves. Since the former kind of condition would count as a condition of things as they appear and the latter of things as they are in themselves, an affirmative answer to this question entails the acceptance of the transcendental distinction, and with it of transcendental idealism. If, on the other hand, the question is answered in the negative, as it is by the standard picture, then any purportedly "subjective" conditions are inevitably construed in psychological terms. The subjectivistic, psychologistic, phenomenalistic reading of Kant, which is characteristic of the standard picture, is thus a direct consequence of its negative answer to this question. The real problem with the picture, however, is not that it answers this question in the negative, for Kant's position may very well turn out to be incoherent or otherwise untenable. It is rather that, by presupposing a negative answer, it never really addresses itself to the question at all. An explicit focus on this question should, therefore, at the very least lead to a more accurate interpretation of Kant's intent and of the nature of transcendental idealism. Beyond that, I also hope to be able to indicate that it makes it possible to regard transcendental idealism as a powerful philosophical position rather than as a curious anachronism or a mass of confusions.