

DERK PEREBOOM

SELF-UNDERSTANDING IN KANT'S TRANSCENDENTAL DEDUCTION

Synthese 103, April 1995, pp. 1-42.

ABSTRACT: I argue that §§15-20 of the B-Deduction contain two independent arguments for the applicability of a priori concepts, the first an argument from above, the second an argument from below. The core of the first argument is §16's explanation of our consciousness of subject-identity across self-attributions, while the focus of the second is §18's account of universality and necessity in our experience. I conclude that the B-Deduction comprises powerful strategies for establishing its intended conclusion, and that some assistance from empirical psychology might well have produced a completely successful argument.

Of the many intricate passages in the Critique of Pure Reason, none has proven to be more challenging to Kant's interpreters than the Transcendental Deduction. Scholarly controversy has persisted about the nature, the value, and the ultimate aims of the arguments it embraces. Nevertheless, the last several decades have been marked by significant progress in the understanding of this text, and in the articulation of interesting new perspectives from which to view its various strata. Here I hope to show that further progress can yet be made, particularly in understanding the structure and content of its arguments.

The arguments of the deduction form the centerpiece of Kant's transcendental philosophy. Transcendental philosophy is his vision of a new model for metaphysics, designed to overcome the problems of rationalist apriorism, and at the same time to fend off the anti-metaphysical

naturalism championed by Hume. As I have argued elsewhere, transcendental philosophy does not seek to illuminate fundamental features of reality as it is in itself, much less to do so by employing justifications completely independent of any sensory experience. Rather, it aims to delineate the contributions of the mind to experience, by justifications that are empirical only to the degree to which they employ information derivable from any experience possible for us.¹ Hence, by contrast to the rationalist's grand optimism about the extent our metaphysical knowledge may achieve, transcendental philosophy has a limited domain. To be sure, its justificatory procedure is a priori, but in a sense less restrictive than that exemplified by the metaphysical method of Leibniz and Wolff.

Since the deduction is Kant's showcase for transcendental philosophy, one might take its effectiveness to exhibit the potential of this new model for metaphysics. I hope to show that the deduction can realize a significant measure of its aims, and that this may well indicate that transcendental philosophy can be successful for a circumscribed set of metaphysical tasks. But we shall also see that the deduction fails at a critical point. I shall argue that this failure is especially instructive because it exposes the limitations of a method that is confined to a priori justifications, even if they are a priori in a comparatively liberal sense.

1

The transcendental deduction is Kant's attempt to demonstrate, against empiricist psychological theory, that an account of experience requires the legitimate applicability of a priori concepts, concepts whose source is the understanding of the subject. As Dieter Henrich has argued (1989), the German equivalent of 'deduction' is originally a legal term; a Deduktion is

a title search, an argument intended to provide a historical justification for a property claim.² In §7 of An Enquiry Concerning Human Understanding, Hume attempts to execute such a title search for the idea of causal power. He maintains that this deduction will only be successful when an original experience, an impression, of causal power is found. Since the search for an impression of causal power is fruitless, Hume concludes that the idea of causal power cannot legitimately be applied to objective phenomena in experience. In Kantian language, Hume's conclusion is that the concept of causal power has no objective validity.

In the transcendental deduction, Kant attempts a different sort of deduction of a concept of cause -- together with eleven others, all of which he believes to be a priori concepts (A80/B106). His strategy is not to unearth an original impression for these twelve categories; such a deduction would be empirical and not transcendental (A85/B117). In Kant's model, a transcendental deduction begins with a slender premise about any possible human experience, a premise to which any reasonable participant in the debate can agree. The argument then proceeds to establish a particular theory of mental processing, by way of providing an explanation for the truth of the slender premise. It then aims to demonstrate that certain a priori concepts have an essential role in this mental processing. The argument concludes by claiming that these a priori concepts are legitimately applicable to experience; if they were not, the truth of the slender premise would lack an explanation. Thus, by the arguments of the transcendental deduction Kant aims to secure a normative claim, that it is legitimate to apply certain a priori concepts to experience, by establishing a psychological theory.³

Perhaps in Kant's day the claim that experience requires processing of mental states had not been suitably justified, but it has come to be regarded as a well-established hypothesis. The

view that an account of experience should begin with unorganized sensation and should then proceed to explain organization by means of conceptual processing has become the standard position in cognitive psychology. According to current theories, mental processing begins with sensory input, and produces cognition by the organization and the application of concepts to this input. As Patricia Kitcher has convincingly argued, this is Kant's simple, yet original, account of mental processing.⁴ Several of the details of Kant's picture, including the view that some of the concepts employed in processing are not derived from experience, but are a priori in the sense that they originate in the structure of the mind, have also become widely accepted.

For Kant, the most important rival theory of mental processing is Hume's. Hume agrees that a theory of experience demands an account of the organization of mental states, but he does not believe that such an account requires the introduction of a priori concepts. According to Hume's view, associationism, the human mental repertoire consists solely of mental states (perceptions, in the Humean vocabulary) that are sensations or very much like sensations. He divides perceptions into two categories, the more vivid original sensations, the impressions, and their less vivid copies, the ideas, which function in imagination, memory, reasoning, and conceptualization. The impressions are passively received, and ideas arise from them without the operation of any real causal power, neither one belonging to an agent, nor a causal power of any other sort (Enquiry, §2, §7).

In Hume's account, association proper is the process by which these perceptions are organized. The hallmark of the theory is that mental processing requires no resources beyond what perceptions are able to provide. To the extent that perceptions are organized in a particular mode or way, this mode of organization must be provided by the perceptions themselves. There

is no possible source of organization outside of the perceptions themselves, simply because perceptions are the only things that exist. The subject, which for Kant is the source of modes of organization, cannot play this role in Hume's picture, since for Hume the subject is nothing other than a collection of perceptions (Enquiry, §§2-3; Treatise, pp. 251-3, 207).

In Kant's theory, organization or processing of mental states (Vorstellungen, translated as representations) is called synthesis. Synthesis is "the act of putting different representations together, and grasping what is manifold in them in one cognition (Erkenntnis)" (A77/B103); it is that which "gathers the elements for cognition, and unites them to form a certain content" (A78/B103). Hence, synthesis is the process by which mental states are connected to form a further state, a state with cognitive content.

Synthesis differs from association in three fundamental ways.

(i) First, the source of synthesis is to be found in a subject: "[b]eing an act of the self-activity of the subject, it cannot be executed save by the subject itself" (B130), and further, this subject is distinct from its states: "through the 'I', as simple representation, nothing manifold is given; only in intuition, which is distinct from the 'I', can a manifold be given" (B135).

Kitcher argues that Kant would be better off if his conception of the self were of a subject that consisted just in interrelated representations: "persons are not mere heaps of perceptions, but we are no more than contentually interconnected systems of cognitive states (at least so far as we can ever know)."⁵ She believes that Kant does not require, and in fact that he should not be positing a representation-independent subject as the source of synthesis. In her view, the best model for Kant's self is a bundle of representations, some empirical and some a priori, among which there are real connections. Synthesis, in her view, is the coming to be of such real

connections among representations, in such a way that the resulting content is dependent on the contents of these representations.

Kitcher's argument for the rejection of a subject independent of its representations as the source of synthesis is that

... acts or processes of synthesis could not be performed by agents. They are unconscious activities within agents that enable them to have cognitive capacities required for agency.

In Daniel Dennett's useful terminology, they are "subpersonal" processes, not acts performed by persons.⁶

Indeed, the view that synthesis is performed by an agent, just as ordinary actions are, is implausible. Synthesis does not involve deliberation, choice, and consciousness in the way that ordinary agency does. But a representation-independent subject might have causal roles, kinds of "self-activity" (B130), other than agency in the process of synthesis. As we shall see, for Kant, such a subject realizes the causal power responsible for synthesis, and it is the source of the concepts by which synthesis organizes representations. Neither of these roles of the subject need involve deliberation, choice, and consciousness in the way that agency does. Perhaps in the final assessment the view that the subject has a causal role in synthesis will need to be rejected. But since a subject's causal roles need not each involve ordinary agency, there may yet be room for the view that the subject plays a causal role in synthesis.

(ii) Secondly, synthesis can employ a priori concepts as modes of processing representations, whereas association never does. According to Kant, synthesis does not by virtue of its definition employ a priori concepts, or any concepts at all, for that matter: "By synthesis, in its most general sense, I understand the act of putting different representations together, and of

grasping what is manifold in them in one cognition" (A77/B103). Kant assigns synthesis, in this general sense, to the faculty of imagination. It is then the task of the understanding to make synthesis conceptual, to supply concepts as modes or ways by which synthesis takes place: "to bring this synthesis to concepts is a function which belongs to the understanding" (A78/B103). Kant is especially concerned to establish a kind of synthesis in which representations are organized by a priori concepts (A79/B105). These concepts, the most celebrated of which are the twelve categories, are a priori in the genetic (and not justificatory) sense that they have their source in the understanding of the subject: "the categories have their source in the understanding alone, independently of sensibility" (B144).⁷ In fact, Kant will eventually conclude that all synthesis takes place by means of the categories: "All synthesis, therefore, even that which renders perception possible, is subject to the categories" (B161).

One might want to designate Kant's a priori concepts innate, but he rejects this classification:

... the Critique admits of no divinely implanted (anerschaffene) or innate (angeborene) representations. It regards them all, whether they belong to intuition or to the concepts of the understanding as acquired (erworben). (On a Discovery, Ak VIII, 221)

Nevertheless, the categories (and the forms of intuition, space and time) are acquired in a special way, by "an original acquisition," by means of which the faculty of knowledge "brings them out of itself a priori" (Ak VIII, 221-2). Kant is not forthcoming about the specific nature of original acquisition, but he does add that "there must... be a ground in the subject which makes it possible for these representations to originate in this and no other manner, and which enables them to be related to objects which are not yet given," and that "this ground at least is innate" (Ak VIII, 221-

2). Thus, although Kant refuses to use the term 'innate' to describe the categories, he does state that an innate capacity or source accounts for the genesis and nature of these a priori concepts. But further, given the ordinary and contemporary use of the term 'innate,' it is in fact applicable to the categories as Kant conceives of them. Kant's use of the term 'innate' is idiosyncratic, since he seems to believe that a concept is innate only if it is not acquired, regardless of whether the acquisition is from sensory experience.⁸ But in the more common use of the term, a concept is innate just in case it is not acquired from sensory experience, but has its source in the structure of the mind, and Kant's categories meet this criterion.

(iii) The third difference between synthesis and association is causal. In his most careful characterization, Kant says that synthesis is the effect (Wirkung) of the power of imagination (Einbildungskraft) (A79/B104) and further, in the second-edition deduction Kant states that the transcendental synthesis of the imagination is an effect (Wirkung) of the understanding on the sensibility" (B152). 'Wirkung,' throughout Kant's writings, means the effect of a cause, and hence, he should be understood as claiming that synthesis is produced by a cause that is realized in a faculty of the subject.⁹ He embellishes his account of this cause by stating that in synthesis the subject is spontaneous; "combination... is an act of spontaneity of the power of representation (Vorstellungskraft)" (B129-30, cf. A546-7/B574-5, Groundwork of the Metaphysics of Morals, Ak IV, 452). Kant fails to explain clearly the nature of the understanding's power of spontaneity, although he does specify that it cannot be a property of the senses (B129-30, B132, A546-7/B574-5), and that it "is not, like sense, confined to ideas which arise only when we are affected by things (and therefore are passive)" (Ak IV, 452). This suggests that the spontaneity of the understanding is at least a real, non-Humean, causal power, that is, a power which cannot be

reduced to regularities among perceptions or events.¹⁰

Kitcher argues that "two states are connected by synthesis just in case they are related as synthetic product and synthetic progenitor, or they are progenitors of some further state."¹¹ She claims, in addition, that connections produced by synthesis are stronger than associative but weaker than real causal connections. If Kant assumed the existence of real causal connections, she contends, he would be begging the question against Hume, since he proves the existence of real causality only after the deduction, in the Second Analogy.¹² But, first of all, it is not clear how there could be room for a type of connection wedged between the kind that association produces, which is causal only in the stripped-down Humean sense, and the non-Humean real sort. Moreover, Kant would not be begging the question against Hume by specifying that synthesis produces real causal connections, since he does not assume the existence of synthesis, but rather aims to demonstrate it -- along with its controversial features. In particular, Kant's argument for real causality in the Second Analogy might be regarded, in part, as an attempt to establish the causal component of his theory of mental processing.

But, furthermore, as Robert Howell points out,¹³ in Kant's view, synthesis does not, at least by definition, produce a causal order or even a quasi-causal order among mental states. Howell argues that if one supposes synthesis to be a causal process -- and this, I believe, is the only way that causation enters into the definition of synthesis -- it does not follow that the resulting relations among mental states are causal. If I put two blocks next to each other, I do so by means of a causal process, but the resulting spatial relation is not a causal relation between the blocks, in the sense that the spatial position of neither block is an effect of the other. Perhaps we should leave it open at this stage of the discussion, as Kant himself does, precisely what type of

relations among mental states synthesis produces. In the final analysis, these relations will be characterized partially by the concepts which function as modes of synthesis.

We are now prepared to proceed to the arguments of the deduction. I shall limit the purview of this discussion to §§16-20 of the B-deduction, the first part of the deduction of the second edition of the Critique. This leaves out a discussion of the less streamlined deduction in the first edition, and of the second part of the B-deduction, found in §§21-27, in which Kant advances an argument for the comprehensiveness of the categories--the thesis that the categories legitimately apply to all of experience (§26), and a more detailed psychological account of the processes by which the categories apply to experience (§24 and §26).¹⁴ As Dieter Henrich has argued, the concern of the first part of the B-deduction is to demonstrate that the categories legitimately apply to any unified cognition, whereas the second part aims to establish that all of our experiences of objects in space and time are unified cognitions. Both parts of the project are significant, but I shall focus exclusively on the first part of the B-deduction.

In my reading of §§16-20, Kant employs a two-pronged strategy for defeating Humean associationism and establishing synthesis.¹⁵ The first argument, contained in §16, is designed to show that associationism lacks adequate resources for the explanation of pure self-consciousness, and hence, that synthesis is required in this explanation. This type of argument has become known as an argument from above.¹⁶ Correlatively, in §§17-20 we find an argument from below, by which Kant aims to demonstrate that synthesis is also needed to explain our representation of objects. Finally, in §20 he concludes that if synthesis is required to explain features of our experience, then we must employ the twelve a priori ways of synthesizing, the categories, and this, in turn, entails the legitimacy of the application of the categories in experience. The

division of §15-20 into two arguments is controversial; I shall defend my interpretation while discussing §17.

2

The argument from above in §16 best divided into two stages. The first aims to establish certain features of the principle of the necessary unity of apperception. The second advances to a priori synthesis by providing an explanation (which Kant believes to be the only explanation) of how we might grasp one prominent feature of this principle. Let us begin, therefore, by examining the principle of the necessary unity of apperception (PNUA). Apperception, first of all, is the apprehension of a mental state as one's own. According to Kant, my apperception has necessary unity because all of my representations, my "empirical consciousness" must be grounded "in pure apperception, that is, in the thoroughgoing identity of the self in all possible representations" (B131-2). By this he means that

(PNUA) It must be the case that each of my representations is such that I can attribute it to my self, a subject which is the same for all of my self-attributions, which is distinct from its representations, and which is or can be conscious of its representations (A116, B131-2, B134-5).

Let me first make four observations about the meaning of this principle, and subsequently consider Kant's argument for it.

(i) According to some interpreters, Kant not only forswears the existence of a representation-independent subject as the locus of the causal power that produces synthesis, and as the source of concepts employed by synthesis, but he denies even that there is an independent

subject to which my representations can be attributed. We have seen that in Kitcher's view, he should deny the existence of such a subject. In her interpretation, Kant's subject should be no more than a collection of representations among which there are real (but not fully causal) connections. Kant does indeed deny Kitcher's view of the subject. Furthermore, a representation-independent subject of self-attribution functions crucially in the argument from above in the second edition. Moreover, it is certainly not clear that the denial of such a subject of self-attribution is theoretically advantageous.

As we have seen, in §16 Kant says that "through the 'I', as simple representation, nothing manifold is given; only in intuition, which is distinct from the 'I', can a manifold be given" (B135, emphasis mine). If intuition is distinct from the 'I', then one's intuitions are not a part of this subject. It would be odd, moreover, to interpret Kant to hold that the 'I' consists merely of a collection of concepts. Thus, given that all mental representations are either intuitions or concepts, one should conclude that the 'I' does not consist of mental representations at all. Moreover, if this 'I' were a collection of representations, Kant would not deny, as he does in the above passage, that anything manifold is given through the 'I'. In addition, Kant frequently affirms that I have no inner intuition of the subject (e.g. B157), but this claim would cause trouble for him if the subject consisted of my mental states, since he also maintains that I can intuit my mental states by inner sense (e.g. A33/B49).

The thesis that the subject of consciousness is distinct from its representations is an important feature of Kant's psychological theory. One source of worries about positing such a subject is the Humean/positivist version of naturalism. According to this view, the notion of a representation-independent subject suggests a Cartesian soul, and the existence of a Cartesian

soul is not empirically verifiable. Moreover, this notion also suggests a locus of real causal powers, and real causal powers do not exist. But although worries of this sort move persons of radical empiricist or positivist sympathies, they fail to carry force even in other versions of naturalism. Psychology posits many structures that are not obviously reducible to something more basic, yet one need not admit that they are non-natural or immaterial (although Kant himself is not a materialist. cf. A379-83). Naturalists can safely maintain that a subject distinct from its representations is no more alarming than mental states and faculties. Real causal powers, and places for them to be realized, have been rehabilitated by naturalists in recent decades. Hence, there is no need to resist attributing to Kant a subject distinct from its representations for reasons of charity.¹⁷

(ii) Kant states that the ability to attribute my representations to a subject is pure, rather than empirical apperception. This means, in part, that one cannot attribute one's representations to a single subject just in virtue of Humean inner perception, or Kantian empirical inner intuition. Kant agrees with Hume that one cannot have inner perception, or in Kantian terms, empirical intuition, of this subject. The subject is implicit, rather than explicit, in experience.¹⁸ But in addition, Kant repeatedly affirms that the purity of this apperception does not imply that the subject to which one's representations can be attributed is intuited -- represented as an object -- in a purely rational or a priori way (e.g., B406-9).

(iii) Kant also states that pure apperception is original, since "it is that self-consciousness which, while generating the representation 'I think' ... cannot itself be accompanied by any further representation" (B132). Houston Smit has pointed out that for Kant 'original' generally indicates the primary or original source of something's

generation or justification.¹⁹ Accordingly, in my view, pure apperception is original since it is a type of self-consciousness that produces self-attributions, that is, 'I think...'- type thoughts, and I am not conscious of the self-consciousness -- the apperceiving I -- indicated in these thoughts in any manner independent of what is contained in these thoughts. I cannot have an intuition or any other type of representation of the apperceiving I independent of 'I think...'-type thoughts, and thus, these thoughts are the original representations of the apperceiving I. I can come to have further transcendental knowledge of the apperceiving I, but my representation of the I itself is limited to what is contained in the 'I think...'-type thoughts and what can be derived from them.

Kant maintains, further, that the apperceiving I is not intuited, i.e. represented as an object, in 'I think...'- type thoughts:

The 'I' is indeed in all thoughts, but there is not in this representation the least trace of intuition, distinguishing the 'I' from other objects of intuition. Thus we can indeed perceive that this representation is invariably present in all thought, but not that it is an abiding and continuing intuition, wherein the thoughts, as being transitory, give place to one another. (A350)

Although the thesis that the apperceiving subject is never represented as an object is somewhat obscure, we might adapt a suggestion of Sydney Shoemaker's, and take this claim to mean that I cannot represent the subject as having properties that are intrinsic to it, or even as having properties that appear to be intrinsic to it.²⁰ I can represent a physical object, such as a car, as having shape, size, and color -- properties which, Kant might well agree, are ordinarily experienced as intrinsic properties of the object, although he would argue that on a deeper analysis, these properties are actually relational properties of the object.²¹ But the only properties

I can represent the subject as having are its representations, which I in no sense represent as intrinsic to it.

Nevertheless, in virtue my capacity for apperception, I can have a type of propositional grasp of the apperceiving subject:

... in the synthetic original unity of apperception, I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am." (B157, emphasis mine)

In apperception, I am conscious that I exist as subject, but I have no intuition of this subject. In this same passage, Kant cautions against thinking of pure apperception as providing knowledge in the sense of 'Erkenntnis' of the self

for knowledge of myself I require, besides the consciousness, that is besides the thought of myself, an intuition [which my consciousness of myself via unity of apperception does not give me] of the manifold in me (B157)

Genuine Erkenntnis of the self as subject would require, in addition to pure apperception, an intuition of this subject.

(iv) A further aspect of the principle of the necessary unity of apperception is especially important for the argument from above; it is the claim that the subject of different self-attributions is the same. We shall see that for Kant, much turns on the thesis that it is "possible for me to represent to myself the identity of the consciousness in [i.e. throughout] these representations" (B133, emphasis mine), or that "I am conscious of the self as identical in respect of the manifold of representations that are given to me in an intuition" (B135, emphasis mine). In these passages, Kant is not claiming that I can come to know that the collection of introspectible representations preserves its identity over time. Rather, as these texts indicate, his thesis is that for different

instances of attributing representations to myself as their conscious subject, I can become conscious that this subject -- that which is conscious -- is the same.²²

Paul Guyer maintains that the argument of §16 begins with the claim that I have a priori certain knowledge of the numerical identity of the subject of all our representations, and to this he objects that Kant has failed to demonstrate that I have such a priori certain knowledge.²³ But although Guyer's objection is relevant to the first-edition deduction, where Kant claims that "the numerical identity... is a priori certain (ist die numerische Identität... a priori gewiß...) (A113), it fails to apply to the second-edition version, for there Kant no longer claims that we have such a priori certain knowledge. In the second edition he never even asserts that we have knowledge of subject-identity, only that we represent it (B133) and that we are conscious of it (B135), and a fortiori, he never claims that we are certain of this identity. Perhaps Kant, when he produced the second edition, still believed that he could show that our consciousness of subject-identity amounts to a priori certain knowledge, but he no longer committed himself to this task.

Let us now turn to the first stage of the argument in §16, which aims to secure certain especially important features of the principle of the necessary unity of apperception. Kant begins by saying:

It must be possible for the 'I think' to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me. (B131-2)

On one reading, the sense in which a representation would be impossible or nothing to me if it were not attributable to a subject is that I could not become conscious of it.²⁴ It might well be

uncontroversial that for any representation of which I am conscious, I can attribute it to myself as subject, assuming my mental faculties are not defective, and as long as no particular account of the nature of the subject is presupposed.²⁵ One might even understand why Kant would think a claim like this to be an analytic truth.²⁶ But the assertion that, even supposing normal mental functioning, I can become conscious of each of my representations, and that I can thus attribute each of them to myself as subject, is not an analytic truth, and may well be false.²⁷ Some representations are deeply buried in my subconsciousness, while they should still be classified as mine in virtue of the types of causal relations they bear to my perceptions, my behavior, and mental states that are uncontroversially mine. But is arguable that I cannot attribute each of these subconscious representations to myself, and it is certainly not an analytic truth that I can. As we shall see, however, the premise that each of my representations is such that I can attribute it to myself is not required for the first stage of the argument from above. Instead, the crucial claim here is for the identity or sameness of the subject of different self-attributions, since it is from this premise that Kant intends to demonstrate that the subject is representation-independent. Whether each of my representations is self-attributable is not especially significant; what is important is the identity of the subject across different self-attributions.²⁸ But failure to secure the premise that each of my representations is such that I can attribute it to the same subject might detract from the success of the second stage of the argument. For without this premise Kant might be barred from concluding that a priori concepts legitimately apply to all of experience, since the representations I cannot self-attribute would seem to be excluded from their purview. Yet the legitimate applicability of a priori concepts to a significant part of experience is not thereby undermined, and this is a dramatic anti-empiricist result well worth pursuing.

Although Hume might well agree that each of my representations is such that I can become conscious of it, and that I can attribute it myself, he would reject the notion of a subject distinct from one's representations:

For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe any thing but the perception... They are the successive perceptions only, that constitute the mind. (Treatise, pp. 252-3; cf. pp. 207, 251ff)

In this view, for me to self-attribute a perception is to think of it as being a member of the bundle of perceptions that is me. A challenge to Hume's position is that when I become conscious of a representation, there must be a subject, separate from all of my representations, which is conscious of it, for although one is not explicitly aware of such a subject, there must nevertheless be something "entering most intimately in what I call myself." Something must function as the agent of the introspection. To this a Humean might reply that the awareness of a perception is just another perception, and hence, such awareness need not involve a subject that is distinct from perceptions. Such awareness is indeed different from any type of perception that Hume explicitly countenances in his account; it is a perception that is an awareness of a perception.²⁹ Yet one might imagine Hume arguing that although he did overlook perceptions with this structural feature, he might easily admit their existence, and thus avoid positing a Kantian subject.

But even if we allow Hume the peculiarity of perceptions that are awarenesses of perceptions, Kant's argument is not thereby deflected. Kant's complaint is not that Hume cannot account for conscious awareness of representations. Rather, the difficulty he sees for "empirical

consciousness," that is, for consciousness according to Humean psychological theory, is that "it is in itself dispersed (an sich zerstreut) and without relation to the identity of the subject" (B133). One implication of this critical passage is that Hume's theory of the subject lacks the resources required to explain how various of my representations can be attributed to a subject that is both conscious of them and the same subject for each act of self-attribution. This objection does not beg the question against Hume, for it assumes only a claim that is difficult to deny, that the conscious subject of different self-attributions is the same, and not that this subject is distinct from its representations.

Hume's account cannot explain this identity, because Humean perceptions of perceptions are wholly distinct from one another; they are "dispersed" (B133), and share no common element (Treatise, p. 252). Hume might propose to explain the identity of the conscious subject of different self-attributions by the perceptions of perceptions being elements of a single causally coherent bundle, which is the subject. But the bundle is not conscious of perceptions; in our embellished Humean account, consciousness of perceptions is a feature of certain individual perceptions. This proposal therefore fails to explain how that which is conscious for different self-attributions can be the very same element of one's psychological structure.

Thus in the Humean theory, the only element of psychological structure that is the same for different self-attributions of perceptions is the bundle of perceptions. But the bundle is not itself conscious of perceptions, rather, such consciousness would have to be possessed by individual perceptions of perceptions. Hence, in the Humean picture, nothing is both conscious of perceptions and the same entity for different self-attributions of perceptions, and this, according to the Kantian view, provides sufficient reason to reject it. Consequently, explaining how the

conscious subject of different self-attributions can be the same entity requires that this subject be distinct from its representations.

In the first stage of the argument from above, then, Kant has argued for certain features of the principle of the necessary unity of apperception. The argument for one such feature, the representation-independence of the subject of self-attribution, required the premise that the subject of my different self-attributions is the same, which is a further aspect of the principle of the necessary unity of apperception, and indeed, the aspect on which the argument of §16 turns. In the second stage, Kant aims to explain how we might grasp the truth of this crucial premise. More precisely, he attempts to demonstrate that Hume's psychological theory cannot explain how I might represent or become conscious of the identity of the subject -- which we now understand not to be made up of its states -- and that a priori synthesis must be embraced to provide this explanation. Accordingly, a second implication of Kant's claim that "the empirical consciousness, which accompanies different representations, is dispersed and without relation to the identity of the subject" (B133) is that Hume's theory cannot account for my representation-relation to the identity of the subject; it cannot explain how I can "represent to myself the identity of the consciousness in [i.e. throughout] these representations" (B133).³⁰ But the theory of a priori synthesis, Kant intends to show, has the resources to provide the account required.

One might entertain two types of explanation for my representation of the identity of the subject of different self-attributions. The first is introspective, and depends on conceiving introspection as analogous to external perception. In this view, I have inner intuitions that enable me to represent the identity the subject for each of my various self-attributions, and the way I represent the identity of the subject is similar to the way I represent the identity of ordinary

objects. Just as I represent two car-experiences to be of the same car by making reference to apparently intrinsic properties I perceive the car to have, I represent two instances of self-attribution to involve the same subject on the basis of apparently intrinsic properties I perceive the subject to have. But Kant and Hume would agree that I cannot represent the identity of the subject in this way. Following Shoemaker's suggestion, they would both maintain that since I have no experience of the subject as possessing intrinsic properties, I cannot represent the identity of the subject in the way I represent the identity of ordinary objects.³¹ For Kant, introspection reveals only my representations, which are not experienced as intrinsic properties of the subject. Hence, this introspective account of my representation of subject-identity is excluded.

The second type of explanation, which Kant endorses, is that I have an indirect representation-relation to the identity of the subject:

That relation comes about, not simply through my accompanying each representation with consciousness, but only in so far as I conjoin one representation with another, and am conscious of the synthesis of them. Only so far, therefore, as I can unite a manifold of given representations in one consciousness, is it possible for me to represent to myself the identity of the consciousness in [i.e. throughout] these representations. (B133, cf. B135)

As Henry Allison points out, Kant is claiming that if I cannot represent the identity of the subject of different self-attributions by intuiting the subject, such representation must depend on my apprehension of a feature of my mental states.³² Kant suggests that the relevant feature is some type of unity of these representations. If the representations that I can attribute to myself are provided with a unity of the right sort, and if I recognize that they have this unity, then I can

represent the subject of the self-attribution of any one of them as identical with the subject of any other.

Association and synthesis are the possible explanations for the production of this unity or, less ambitiously, for my ability to recognize this unity. But, Kant seems to suppose, since we have already ruled out Hume's psychological theory, synthesis is the only possible explanation. Consequently, in order to explain how I represent the identity of the subject of different self-attributions, I must produce or recognize a unity among these representations, and synthesis -- indeed, a priori synthesis -- must be embraced to explain this recognition. Kant argues that this combination "is an affair of the understanding alone, which itself is nothing but the faculty of combining a priori" (B134-5).³³ Since, as we have seen, the understanding provides concepts for synthesis, and since for synthesis to be a priori in this context is for it to employ a priori concepts, Kant is contending, in effect, that synthesis by means of a priori concepts is required to account for my production of or recognition of unity, which in turn explains how I can represent the identity of the subject of different self-attributions.

Kant maintains, then, that my representation of the identity of the subject comes about "only in so far as I conjoin one representation with another, and am conscious of the synthesis of them" (B133). I believe that this claim is best interpreted to mean that my representation of the identity of the subject requires the consciousness not of the act of synthesis, but rather of the product of synthesis, that is, of the requisite unity among my representations produced by synthetic mental processing. For first, when he formulates his general characterization of synthesis, Kant affirms that it is a function of the soul "of which we are scarcely ever conscious" (A78/B103). And later, in the first edition deduction, he remarks

For this unitary consciousness is what combines the manifold, successively intuited, and thereupon also reproduced, into one representation. This consciousness may often be only faint, so that we do not connect it with the act itself, that is, not in any direct manner with the generation (Erzeugung) of the representation, but only with the effect (nur in der Wirkung). (A103-4)³⁴

Consciousness of the combination or synthesis may be faint, and when it is, one is conscious of the effect or product of the act, and not of the act of combination itself. Second, to interpret Kant's argument against Hume as dependent on the claim that we are conscious of this sort of process would undermine its force, because it seems quite clear that if there is such a process as synthesis, it is typically subconscious rather than conscious.

According to Kitcher, in §16 Kant invokes synthesis to construct an account of personal identity.³⁵ In her interpretation, one's personal identity consists in the existence of synthetic connections among one's mental states. But if, as I contend, Kant claims the existence of a subject independent of one's representations, and he is concerned to explain how one represents the identity of this subject for different self-attributions, then in his view the identity of this subject does not consist in the existence of synthetic connections among one's mental states. Rather, such synthetic connections are required to account for one's representation of this subject's identity.³⁶ In my interpretation, therefore, Kant invokes synthesis not to explain what personal identity consists in, but rather to explain how one represents the identity of the subject of one's self-attributions.

Let us examine Kant's contentions more closely. What sort of unity must I produce or recognize among my representations in order to explain how I might represent the identity of the

subject? Regrettably, Kant does not provide a specific answer to this question. On Allison's reading, the requisite unity among representations is that they simultaneously be objects of conscious awareness:

The awareness of the identity of the I that thinks A with the I that thinks B obviously requires an awareness of both A and B. This is because the I of the 'I think' has no determinate content, and thus cannot be characterized apart from its representations.

Consequently, unless I can become aware of both representations together, I cannot become aware of the identity of the I that thinks the one with the I that thinks the other.³⁷

In Allison's view, the kind of unity among my representations that I must recognize in order to be aware of the identity of the subject of different self-attributions is that the self-attributed representations are actually objects of awareness at the same time. But Allison's view is unlikely to capture the truth of the matter, since there are many representations whose subject of self-attribution I represent as identical that do not have the requisite co-consciousness relation. For example, I represent the subject of the representation I had at 2 P.M. yesterday as identical to the subject of the representation I am having now, but it is not the case that these representations are simultaneously objects of my awareness. Moreover, since some of the representations I had yesterday have since gone out of existence, they cannot become objects of awareness simultaneously with the one I am having now, even though I am conscious that they can be attributed to the same subject. Fortunately, however, that I "unite a manifold of given representations in one consciousness" (B133) need not be interpreted to mean that I am aware of these representations at the same time, but only that these representations are provided with and are recognized as having a certain unity, of which I, as a single conscious subject, am aware.

A more plausible account is that the unity among my representations I recognize consists in certain intimate ways in which mental states in a single subject are typically related. These intimate relations must be readily accessible to the subject, for example, by introspection or by inference from introspective information. One relation of this sort is that a single subject's mental states are inferentially integrated to a significant degree, more so than are mental states of different subjects. Another is that the mental states of a single subject can and often do causally interact with each other without a non-mental intermediary. Accordingly, I suggest the following embellishment of Kant's account: my recognition that my mental states stand in these intimate relations is what accounts for my ability to represent the identity of the subject of different self-attributions.³⁸

One might object that this Kantian account fails because it is more plausible that my consciousness of the identity of the subject is basic, and thus not dependent on any further recognition. In our account, I am conscious that my representations are not attributable to more than one subject because I recognize that they are related to each other in certain intimate ways. But for what reason do I believe that the states of two distinct subjects cannot be so intimately related? I might say that I have never experienced the states of two distinct subjects to be intimately related in these ways, but how can I rule out the possibility that, for example, the states I call mine are attributable to two subjects and are so related? Without accepting my consciousness of the sameness of the subject of my self-attributions as basic, it is hard to see how to get off the ground.

This objection misses the best way to construe the Kantian account. Although Kant does not do so himself, it is preferable to cast his view as underlain by an argument to the best

explanation. In general, Kant's strategy, according to which the deductions are arguments to the necessary conditions of (certain features of) any possible human experience, does not realize the potential of the considerations he advances as well as the best explanation approach does. So far we have suggested that what explains my consciousness of the identity of the subject of different self-attributions is my recognition that my representations are intimately related to each other in certain ways. The plausibility of this explanation, in turn, is undergirded by the fact that the identity of the subject is part of the best explanation of one's representations being related in these intimate ways.³⁹ There are different explanations for these representations being so related, and it is very plausible that the Kantian candidate, which incorporates the thesis that they are attributable to the same subject, is better than the others. It would indeed seem to be logically possible that the states I call mine are intimately related in the ways we have discussed, but that there are two subjects who hold these states in common. But there exist analogous logically possible hypotheses for entities that we posit for best explanations in the physical sciences. We rule out such hypotheses because we consider them to belong to inferior explanations, and for the same reason we can reject the two-subject thesis.

Furthermore, by considering hypothetical cases we can see that the identity of the subject is indeed plausibly part of the best explanation for one's representations being intimately related in the ways we have indicated. Suppose, first, that a human body comes to be, and I begin to notice that the representations that originate in this new body causally interact without non-mental intermediary with my representations, and that those that originate in this new body are inferentially integrated with mine. For example, when this body visits Egypt, I have visual experiences of the pyramids, and these experiences cause additions and revisions to my

representations in just the way my more conventional visual experiences do. In these circumstances, I have evidence that I am the same subject as the subject of the experiences produced in this body, and this may well be the best explanation for the phenomena. Second, imagine explorers discovering an unusual being on the planet Mars. The only way information taken in by its right half can causally interact with information taken in by its left half is by hand-signing. Both the right and the left halves of this being possess faculties of mental processing sufficient to allow gathering of information and subsequent transmission by hand-signing, and no causal interaction between right- and left-half information can take place without a non-mental causal intermediary. Such circumstances provide evidence that there are two subjects in the Martian being, one for each half, and perhaps the evidence is strong enough for the two subject-hypothesis to count as the best explanation.

In our Kantian model, then, I represent the identity of the subject of different self-attributed representations in virtue of recognizing that they are related to each other in certain intimate ways, for example, that they are inferentially integrated in a special way, and that they can and often do causally interact without a non-mental intermediary. The plausibility of this account is underlain by the identity of the subject being part of the best explanation of these various ways one's self-attributed representations are intimately related. In addition, I believe that this type of account is a definite contender, although the claim that consciousness of subject identity is basic is perhaps an attractive alternative.

But what of the claim that recognizing that my representations are intimately related in certain ways requires synthesis, and moreover, synthesis by means of a priori concepts, and that neither association nor synthesis by empirical concepts is sufficient? As Guyer argues, demon-

strating the need for synthesis by means of a priori concepts would require ruling out the possibility that empirical information alone could account for the recognition of unity of this sort.⁴⁰ Yet it might be, for all Kant has shown, that recognizing that, for example, my representations are intimately related is explainable only in virtue of my awareness of information derived from experiences of interactions among my representations. He does not attempt to argue against such rival empiricist hypotheses, and hence, he has not confirmed the need for a priori synthesis. I shall argue that this type of problem also affects the argument from below, but in addition, that this is might well be the only major flaw in the strategy of the deduction.

Let us take stock. The argument of §16 seems to falter in its aim to demonstrate that a priori synthesis is the process required for the explanation of one's knowledge of the identity of the subject of different self-attributions. Yet it supplies a weighty argument for establishing the existence of a subject distinct from its representations, and it provides the blueprint for a contending account of one's knowledge of the identity of the subject of different self-attributions. Moreover, in view of the purposes of the deduction as a whole, the failure of the main aim of §16 by no means renders its successes wasted effort. The existence of a representation-independent subject, in particular, is a crucial element of Kant's theory of mental processing, and it functions as an important prolegomenon to the argument from below. Finally, one might view the argument of §16 as introducing what Kant believes to be the most fundamental role for a priori synthesis, the part it plays in pure apperception, while he intends to establish the details of his theory only in later sections.

In my view, the main function of §17 is to provide a characterization of an object, or more significantly, of a representation of an object, that incorporates a challenge to Humean associationism, and thereby initiates the argument from below. In this section, Kant proposes that an object is "that in the concept of which a manifold of a given intuition is united" (B137). This notion of an object is very broad, in the sense that it is not restricted to bounded, persisting objects, but comprises all objective phenomena, including, for example, events, states of affairs, non-solid substances, and mental states. It is important that this characterization is stated in terms of the notion of a concept of an object, since this indicates that Kant's overriding concern is with objects as we represent them, rather than with objects simpliciter. The type of unification of a manifold Kant refers to in this characterization of an object is clearly synthetic, and requires the subject and the powers it needs for self-attribution, for immediately following the characterization of an object he claims that "all unification of representations demands unity of consciousness in the synthesis of them" (B137). The main upshot of Kant's account is that my representation of this house, for example, is of an object just in case this representation is made up of a group of sensations synthesized under a concept such as 'this house.' This depiction is not best regarded as providing a definition of an object, but rather as a characterization designed to present Kant's view about the processing required to produce a representation of an object, and hence, one intended to provide a challenge to Humean associationism. In addition, this characterization does not set out a position Kant expects us to accept without argument, but rather, one he aims to establish in §§18-20. In these sections, Kant's arguments proceed by adducing characteristics of our representations of objects, about which both he and Hume can agree, and by showing that the theory of synthesis must be embraced to account for those

features.

Allison's reading of §17 differs from mine. According to his interpretation, this section does not simply contain a challenge to Hume, but it also includes a demonstration of the existence of representations of objects, on the basis of the theory of self-consciousness developed in §16. This picture is part of Allison's broader schema, in which Kant establishes that the transcendental unity of apperception, that is, the attributability of all of my representations to a single subject, entails the representation of objects, and, conversely, that the representation of objects entails the transcendental unity of apperception:

The essential move in the first part of the Deduction is the attempt to establish a reciprocal connection between the transcendental unity of apperception and the representation of objects. I shall call this the "reciprocity thesis." It is the specific concern of §17.⁴¹

The crucial claim for Allison's interpretation is that the transcendental unity of apperception is not only a necessary but also sufficient condition for the representation of objects. In my view, by contrast, the transcendental unity, and more precisely, the synthesis that explains our consciousness of the identity of the subject, is only a necessary condition for the representation of objects, and, moreover, a condition which Kant only proposes in §17, and does not aim to demonstrate until §§18-20. Elsewhere in the Critique, for example in the first-edition deduction (A108) and in the Refutation of Idealism (B276), Kant indeed endorses the view that some form of self-consciousness entails representations of objects.⁴² But I contend only that establishing such a thesis is not an element in the argument of the second-edition deduction, and in particular, not a part of the project of §17.

Perhaps Allison's aim in proposing the proof that we represent objects is to show that §§15-20 comprise a single argument whose only assumptions are uncontroversial premises about self-consciousness discussed in §16. Demonstrating that we represent objects would then have a place in this schema, since then no mere assumption about the existence of representations of objects would be required for the stages of the argument that take place in §§18-20. This interpretation is unconvincing, mainly because, as we shall see, in §§18-20 Kant makes crucial assumptions about features of our representations of objects that exceed anything that he might be thought to have argued for in §17. In particular, he assumes that our representations of objects exhibit certain kinds of necessity and universality, and Kant never attempts to establish by argument that our representations of objects have such characteristics, either in §17 or anywhere in the Critique.

Moreover, in the summary statement of the deduction in §20 Kant does not include premises from §§15-16. What we actually find in §20 provides evidence that Kant intends §§17-20, with some help from §13, to constitute a single, self-contained argument. Perhaps the summary statement fails to mention §§15-16 because Kant conceives these sections to function mainly as a prolegomenon which introduces the notion of synthesis and its relation to pure apperception, while he intends to secure the main aims of the deduction only in the argument from below.

Furthermore, the text of §17 fails to support Allison's claim. In his view, the argument from the transcendental unity of apperception for the existence of representations of objects is compressed into this single paragraph:

(A) Understanding is, to use general terms, the faculty of cognitions (Erkenntnisse). They

consist (bestehen) in the determinate relation of given representations to an object: and an object is that in the concept of which the manifold of a given intuition is united. Now all unification of representations demands unity of consciousness in the synthesis of them. Consequently it is the unity of consciousness that alone constitutes the relation of representations to an object, and therefore their objective validity and the fact that they are cognitions (Erkenntnisse): and upon it therefore rests the very possibility of the understanding. (B137)

Allison himself presents a problem for his interpretation of this passage. In his view, the reciprocity thesis is encapsulated in this sentence:

(1) it is the unity of consciousness that alone constitutes [ausmacht] the relation of representations to an object, and therefore their objective validity...⁴³

and Kant presents (1) as a direct consequence of the premise that

(2) all unification of representations demands unity of consciousness in the synthesis of them.

But in this picture, Allison points out, Kant is apparently guilty of a gross non sequitur, since (2) would clearly license only the conclusion that the unity of consciousness is a necessary condition for the representation of an object, and not also a sufficient condition.

Allison's way out of this problem is to suggest that in §17 Kant employs a minimal sense of 'object,' according to which an object is the result of any synthetic process of unification.⁴⁴ In §16, Allison contends, Kant has argued that the unity of consciousness is impossible apart from a synthetic unity of representations, and that this synthetic unity can be achieved only by uniting these representations under concepts. But since (by definition) any such synthetic unity counts as

an object, it also follows that the representation of an object is a necessary condition for the unity of consciousness. But this is equivalent to saying that the unity of consciousness is a sufficient condition for the representation of an object, and thus the controversial part of the reciprocity thesis is vindicated.

Perhaps the most problematic feature of Allison's interpretation is the definition of an object he attributes to Kant. According to Kant,

an object is that in the concept of which the manifold of a given intuition is united.

(B137)

Allison claims from this characterization it follows "that whatever can be represented by means of the unification of a manifold of intuition under a concept counts as an object."⁴⁵ But Kant does not say that unification under any concept results in the representation of an object. His characterization leaves open the possibility that there are unifications of representations that do not employ concepts of objects, and which therefore do not result in representations of objects. An object is that in the concept of which the manifold of a given intuition is united, not the result of the unification of the manifold of a given intuition under any concept. Kant's characterization entails only that unification under a concept of an object is both a necessary and sufficient condition of the representation of an object, but in §17 he does not attempt to demonstrate from principles established in §16 that we employ such concepts of objects when we synthesize representations.

In addition, there exists no convincing reason to believe that Kant's claim

(1) Consequently, it is the unity of consciousness that alone constitutes [ausmacht] the relation of representations to an object, and therefore their objective validity...

should be read as a statement of the reciprocity thesis. Consider again the entire paragraph of which (1) is a component:

(A) Understanding is, to use general terms, the faculty of cognitions (Erkenntnisse). They consist (bestehen) in the determinate relation of given representations to an object: and an object is that in the concept of which the manifold of a given intuition is united. Now all unification of representations demands unity of consciousness in the synthesis of them. Consequently it is the unity of consciousness that alone constitutes the relation of representations to an object, and therefore their objective validity and the fact that they are cognitions (Erkenntnisse): and upon it therefore rests the very possibility of the understanding. (B137)

To be precise, Kant does not say here that the unity of consciousness constitutes the representation of an object, but rather that it constitutes the relation of representations to an object. Cognitions of objects, he argues, consist in some determinate relation of representations to object, and here he suggests that this relation is produced by the same process of unification, synthesis, that explains knowledge of the unity of consciousness. But then (1) does not entail that the synthesis that explains knowledge of the unity of consciousness cannot take place without producing a relation of representations to objects. By analogy, smelting and molding are the processes which produce steel girders, but it does not follow that the processes of smelting and molding cannot take place without the production of steel girders. Just as the production of steel girders requires particular molds in addition to the processes of smelting and molding, so the production of representations of objects may well require concepts of objects in addition to the synthesis that explains knowledge of the unity of consciousness. Furthermore, the term

'alone' in (1) need not be taken to indicate sufficiency. Its function may well be to point out that the synthesis that explains knowledge of the unity of consciousness is the only process that constitutes the relation of representations to an object, not that any such synthesis produces a representation of an object without any further conditions being fulfilled.

Richard Aquila cites the first sentence from §18 as confirmation of the reciprocity thesis:

The transcendental unity of apperception is that unity through which all the manifold given in an intuition is united in the concept of an object. (B139, emphasis Aquila's)⁴⁶

But there is no reason to take the emphasized 'is' as the 'is' of identity, or as indicating sufficiency in some other way. In my reading, Kant is saying that the synthesis that explains knowledge of the transcendental unity is also the process that provides representations of objects, although in this passage he is silent on whether the process, independent of any other conditions, is sufficient for representations of objects.

In addition, Kant does not articulate the reciprocity thesis in the summary statement of §20, where one would expect to find it if he indeed intended it to have a critical role in the deduction. The summary of section §17 is simply:

The manifold given in a sensible intuition is necessarily subject to the original synthetic unity of apperception, because in no other way is the unity of intuition possible (§17).

(B143)

This sentence entails only that the synthetic unity of consciousness is necessary for intuitions, that is, for representations of objects, and not that this unity of consciousness is also sufficient for such representations.

Moreover, suppose that in §17 Kant were intending to show that the unity of

consciousness is a sufficient condition of representation of objects in some minimal sense of 'object.' Then the sufficiency claim could at best be regarded merely as an interesting aside, a component of an analysis of self-consciousness, and not as an essential element of the argument of the B-deduction. For as I have pointed out, in §§18-20, Kant argues for a priori synthesis on the basis of contentions about the necessary and universal character of our representations of objects, claims that far exceed the thesis that we have representations of objects in a minimal sense of 'object.' Or worse, if the argument of §18 required Kant to have demonstrated the claims about the necessary and universal character of our representations of objects solely on the basis of premises about the unity of consciousness, then the deduction would be an obvious failure, for he neither offers such a demonstration, nor does one seem possible. Consequently, since the goal of demonstrating, by means of premises about unity of consciousness alone, that we have representations of objects has no clear function in the project of the B-deduction, and since there is no textual evidence that Kant intended this goal to have such a function, Allison's reading should be rejected. The main role of §17 is best interpreted as characterizing our representations of objective phenomena in a way that issues a challenge to Humean associationism. Kant's aim in this section, then, is to provide an exposition of the view that our representations of objects require synthesis, the same type of synthesis needed to explain knowledge of the unity of consciousness, and this he intends to establish in §§18-20, to which we now turn.⁴⁷

4

In §18, then, Kant continues the argument from certain features of our representations of objective phenomena against association and for synthesis.

The transcendental unity of apperception is that unity through which all the manifold given in an intuition is united in a concept of the object. It is therefore entitled objective, and must be distinguished from the subjective unity of consciousness... Whether I can become empirically conscious of the manifold as simultaneous or as successive depends on circumstances and empirical conditions. Therefore, the empirical unity of consciousness, through association of representations, itself concerns an appearance, and is wholly contingent... Only the original unity is objectively valid: the empirical unity of apperception,... which... is merely derived from the former under given conditions in concreto, has only subjective validity. One person connects the representation of a certain word with one thing, the other [person] with another thing; the unity of consciousness in that which is empirical is not, as regards what is given, necessarily and universally valid. (B139-40)

Kant here characterizes the empirical unity of consciousness as non-universal, contingent, and subjectively valid, and distinguishes it from the transcendental unity of apperception, which he describes as universal, necessary, and objectively valid. Since the empirical unity of consciousness is an organization of representations produced by association, Kant is maintaining that association can achieve only an organization that is non-universal, contingent, and lacks objective validity. The transcendental unity of apperception, by contrast, is linked to an organization of representations produced by synthesis, and thus Kant is claiming that synthesis can generate an organization that is universal, necessary, and objectively valid.

In this argument, a salient feature of certain representations of objects is their objective validity. As Guyer points out, representations are objectively valid when they represent or apply

to entities or states of affairs that are independent of the way they are perceived for their characteristics and existence.⁴⁸ In the argument of §§17-18, Kant assumes that the representations that make up experience are objectively valid. He then argues that association is an inadequate account of psychological processing because it can yield only a result that is subjectively valid, only an experience of "objects" that are not independent of the way they are perceived for their characteristics and existence. This case for the deficiency of association appears, in turn, to depend on the claim that Hume's theory cannot account for the universality and necessity that are present in our representations of objective phenomena.

Here we must be careful. The objection Guyer raises most persistently against the arguments of the deduction attributes to Kant an illegitimate assumption of knowledge of necessity. As it applies to §§17-18, Guyer's objection has two stages. The first is that Kant makes a supposition that a Humean would reject, that the mental organization or connection required for representation of objects is in some sense necessary. The second is that Kant assumes without argument that any necessary connection would have to be known a priori, and therefore demands a priori concepts.⁴⁹ I shall argue that on the most preferable interpretation, the argument from below does not succumb to Guyer's criticisms.

The crucial premise in Kant's argument against the adequacy of association in the above passage is:

(3) Whether I can become empirically conscious of the manifold as simultaneous or as successive depends on circumstances or empirical conditions.

This leads him to conclude that

... the empirical unity of consciousness, through association of representations, itself

concerns an appearance, and is wholly contingent. (B139-40)

In these passages, Kant is invoking considerations that foreshadow the discussion of the Second Analogy, although the details of the argument of that section differ from those of the argument of §18. In the Second Analogy, Kant argues that human mental representation, considered independently of its content, is always successive. The states of affairs that are represented, however, are not always represented as successive. For example, in the case of an experience of a boat floating downstream, the states of affairs are indeed represented as successive, while in the case of the viewing the parts of a house, scanning it from the roof to ground, they are represented as simultaneous.⁵⁰ How do we account for the difference in the way the contents of these successive mental representations are represented? In §18, Kant implies that an important clue for answering this question is that the simultaneity or successiveness of the content of our representations is, in a sense we must yet determine, universal and necessary.

A first approximation of the import of 'universal' in this context is:

(4) Any human experience of the parts of the house is an experience of an these parts as objectively simultaneous.

The addition of necessity has the following effect on (4):

(5) Necessarily, any human experience of the parts of the house is an experience of these parts as objectively simultaneous.

Kant's argument, then, would be that given only the resources of association, the truth of (5) cannot be explained, for "whether I can become empirically conscious of the manifold as simultaneous or as successive depends on circumstances or empirical conditions" and, therefore, "the empirical unity of consciousness... is wholly contingent" (B139-40).

On one reading, Kant is claiming that if association exhausts our capacity for mental processing, then it is a real possibility that unusual circumstances, resulting in mistaken perceptions of objective simultaneities and objective successions, arise in experience. In this interpretation, he is arguing that if the associationist account is true, then it is possible that circumstances arise in experience that entail

(6) Possibly, some human experiences of the parts of the house are not experiences of these parts as objectively simultaneous.

and (6) is inconsistent with (5). And since (5) is true, associationism is false. For example, one might envision repeatedly observing different parts of the house in the same temporal order. These circumstances of experience might result in one's seeing the parts of a house as an objective succession, and not as an objective simultaneity.

According to this interpretive strategy, Kant is here advancing the "accidental constant conjunction" argument against Hume's theory of association. In general, unusual patterns in experience might well give rise to unusual associations and causal beliefs. But whereas this type of argument does raise problems for Hume's theory of what causation is, it fails to undermine his account of mental processing. For Hume could reply that such an account need only show why the usual organization normally results, and that an acceptable theory need not preclude the odd deviant organization. The Humean reply is, in effect, to deny the necessity in (5), and to explain (5)'s apparent plausibility by the fact that under normal circumstances the usual organization is produced by association. Guyer's criticism is, I believe, effective for this first reading of the argument. In this interpretation, Kant assumes the existence and knowledge of a necessity that Hume would resist.

But let us now consider a second, more satisfying interpretation of the premise

(3) Whether I can become empirically conscious of the manifold as simultaneous or as successive depends on circumstances or empirical conditions.

On this second reading, Kant is not asserting that divergent, and in fact, impossible experiences of objective simultaneities and successions would result if associationism were true and empirical conditions were unusual. Rather, he is pointing out that the only resources for explanation possessed by associationism are empirical conditions, and empirical conditions are insufficient to explain the organization that our experience must display, when such conditions are normal.

According to Humean associationism, the material for psychological explanation includes only passively received, sensation-like mental representations. The mind consists just in such passively received representations, and hence there are no resources in the mind beyond this sort of representation. In a word, the resources of association to account for mental organization do not reach beyond empirical conditions. In this interpretation, the problem for Humean associationism is not that it fails to produce normal organization when empirical conditions are abnormal. Rather, supposing that empirical conditions are normal, associationism is insufficient to account for the types of organization of representations that must result.

We saw that Kant's assumption that

(5) Necessarily, any human experience of the parts of the house is an experience of these parts as objectively simultaneous.

can be denied by Hume if we take the necessity operator to range over all possible circumstances, because his theory can absorb the odd deviant organization in unusual conditions. Suppose, then,

we recast (5) as

(5') Necessarily, if empirical conditions are normal, any human experience of the parts of the house will be an experience of these parts as objectively simultaneous.

In this interpretation, Kant's view is that for all human beings, given certain specific empirical conditions, one particular type of organization of mental representations must result.

Significantly, Hume would not deny the existence of some type of necessity in such contexts.

For example, Hume believes that it is impossible, in some sense, given an experience of constant conjunction, that the mind not be carried from the experience of the first conjunct to the thought of the next:

... having found, in many instances, that any two kinds of objects, flame and heat, snow and cold, have always been conjoined together; if flame or snow be presented anew to the senses, the mind is carried by custom to expect heat or cold, and to believe, that such a quality does exist, and will discover itself upon a nearer approach. This belief is the necessary result of placing the mind in such circumstances. It is an operation of the soul, when we are so situated, as unavoidable as to feel the passion of love, when we receive benefits; or hatred, when we meet with injuries. (Enquiry, §5, p. 30)

In this passage Hume claims that given the types of being we are, and certain specific empirical circumstances, a particular type of mental organization necessarily or unavoidably comes about.

Thus Hume and Kant might well agree on propositions of the form of (5'), and consequently, the first stage of objection that Guyer raises fails to dislodge the argument of §18. Neither Hume nor Kant would claim that necessarily, empirical conditions are always normal, or that necessarily, one particular mode of organization of passively received representations will

result whether empirical conditions are normal or abnormal. They may well agree, however, that in some particular perceptual situation, necessarily or unavoidably, when empirical conditions are normal, one particular mode of organization of passively received representations will result.

One might object that my interpretation illegitimately employs the concept of necessity, because this concept is one of the categories whose objective validity Kant is attempting to demonstrate. But claiming that premise (5') is true does not amount to presupposing that a category legitimately applies to experience. Although the premise does make a claim to necessity or unavoidability, this claim does not presuppose that the 'necessity' in the premise is an a priori concept. Moreover, the objection that no premise for the deduction may employ concepts whose apriority and legitimate applicability to experience Kant aims to establish is much too demanding. 'Existence' and 'negation,' for example, are among the categories, and to claim that the deduction may not employ premises involving these concepts is to submit to an impossible standard.

The argument also does not fall prey to the second part of Guyer's objection, that Kant merely assumes that all knowledge of necessity must be rooted in a priori concepts. One need not interpret him as simply making a direct leap from the truth of propositions of the form of (5') to the existence and legitimate applicability of a priori concepts. Rather, one should see him as advancing his claim for the objective validity of the categories by ruling out association as an explanation for such truths. Let us divide the next point he hopes to reach into three steps:

(7) In order to explain the truth of (5'), there must exist a mental capacity for processing representations.

(8) This mental capacity cannot be constituted solely of sensation-like mental states.

(9) This mental capacity must employ concepts that are a priori in the genetic sense, i.e. concepts that are innate.

Kant's challenge is to explain why, under certain conditions, the very same type of organization is inevitably produced in everyone. Part of the best explanation, he believes, is (7), that there must exist some type of mental capacity for producing this organization by processing representations.

Hume might conceivably agree with this conclusion, but he would certainly deny (8), that this mental capacity cannot be constituted solely of sensation-like perceptions. In Hume's account, there can be no mental capacity not made up of sensation-like perceptions, for the simple reason that the mind consists exclusively of such perceptions. Kant argues that Hume's proposal for a mental capacity that consists solely of such perceptions, association, cannot account for the truth of propositions such as (5'). Kant asks us to consider an activity, word association, which functions as a paradigm for the Humean mental capacity. Word association, familiarly, does not yield universal and necessary patterns; "one person connects the representation of a certain word with one thing, the other [person] with another thing..." (B140). Indeed, Hume's own primary example for association in §3 of the Enquiry is the connections among parts of a conversation. Kant's point is that in such conversations, people make different associations in the same circumstances. If the very paradigms for association never exhibit the necessity and universality, then the hypothesis that association is powerful enough to produce organizations with such features is on shaky ground at best. Hume might reply that such variability in associations made with a particular word comes about as a result of the variability in patterns of experiences connected with that word, but if experiential histories were to exhibit strict uniformity, patterns of association would display the requisite necessity and universality. Hume may

well be right here, but the deeper problem for his view is that although association may be subject to causal laws, it is responsive to representations of too many background features of a situation -- it is too "noise"-sensitive. The process that produces experiences as of objective simultaneities and successions yields the same type of organization in circumstances in which the represented background features of the situation are very different, and this fact is a significant component of the necessity and universality of a type of organization. And hence, the process that produces such experiences must be capable of ignoring or filtering out representations of background features of a situation. Association, as is evident from the paradigm of word association, is not, to the requisite degree, "noise"-independent.

One might supplement Kant's argument with an additional worry about Humean associationism. The central element in Humean association is custom, which results from repetition. Familiarly, if B's always follow A's in one's experience, (where A's and B's are perceptions), a custom will be produced that leads one to expect B's upon the appearance of A's. But how is this capacity realized? Besides the vivid original impressions, the only resource Hume possesses are ideas, the faded perceptions which serve, among other things, as memories. The Humean mental capacity, it would seem, must consist solely in such faded perceptions. Perhaps a whole series of faded A's and B's in conjunction form a part of a mind. But what is it about this set of perceptions that would produce an expectation of a B upon the appearance of a new A? Does the new A interact with the faded A's and B's to generate an expectation of a B? What would it be about this group of faded perceptions that would produce expectation of a B? It would seem that any mental capacity Hume can construct from perceptions alone is insufficiently explanatory. Hence, one is led to assume the existence of some sort of perception-

independent mental capacity.

Suppose Kant can establish (8), that there must be a mental capacity not constituted solely of sensation-like perceptions. Nevertheless, the argument for (9), that this capacity must employ a priori concepts, does not fare as well. To show that some necessary and universal mode of organization requires a priori concepts, Kant must first demonstrate that the available empirical information is insufficient to establish a mental capacity that might produce this organization. In the case of the perceptions of the house, for example, Kant must appraise the hypothesis that information such as that derived from the sense of head or eye motion, from the perception of absence of motion of the object in relation to a background, and from past experiences of houses whose parts do not succeed one another, provides sufficient evidence for everyone, under normal circumstances, inevitably to come to perceive the parts of a house as objectively simultaneous. Similarly, there is much empirical evidence for certain events being objectively successive, and Kant must rule out the view that this evidence is sufficient to account for the way in which we organize our experiences of such events.

Such considerations give rise to doubts about whether Kant can hope to execute the arguments of the deduction by justifications that are a priori in his comparatively liberal sense, that is, by justifications that are empirical only to the extent that they depend on information derivable from any possible human experience, and that thus do not rely on thoroughly empirical experiment and observation. Contemporary arguments for the innateness of concepts, like Chomsky's argument for the innateness of deep grammatical structure, go beyond Kant's type of a priori methodology in at least two ways.⁵¹ First, in order to demonstrate that the relevant type of psychological processing which employs the concepts is (potentially) present in all normal

human subjects, one needs thoroughly empirical evidence, and not just evidence that is empirical to the extent to which it is derivable from any possible human experience. To show, for example, that no children make linguistic errors of a particular form, special empirical observations must be made. Moreover, it is quite clear that universality claims of the sort Kant employs, such as

(10) If empirical conditions are normal, any human experience of a boat floating downstream will be an experience as of an objective succession.

cannot be justified without appeal to special empirical observations, despite their obviousness independent of any overt empirical investigation. Empirical information is required for one to come to know whether someone else's experience of a boat floating downstream is an experience as of an objective succession. I also suspect that any necessity claim that can be established in such contexts is based on the correlative universality claim, is causal in character, and demands additional empirical evidence. Once the universality claim is secured, one infers that it is undergirded by a mental capacity that necessitates a type of mental organization. The justification of the inference to a mental capacity plausibly requires experience, perhaps of correlations of regularities and causal mechanisms of various sorts.

The second way contemporary innateness arguments reach beyond Kant's a priori methodology is in ruling out rival empiricist hypotheses. The contention that a concept is innate can often be challenged by a serious rival hypothesis that the concept can be acquired through experience, and it is difficult to see how such rival empiricist hypotheses can be undermined without thoroughly empirical evidence. For a contemporary illustration of special relevance to Kant's aims in the deduction, consider a recent article by Nancy Soja, Susan Carey, and Elizabeth

Spelke, in which they argue for the innateness of the ontological categories of object and substance.⁵² According to a Quinean hypothesis, categories like that of a bounded, persisting physical object (table), and that of a non-solid substance (peanut butter), are learned from experiences of the quantificational syntax of the language (e.g. 'this' indicates an object, and 'some of' indicates a non-solid substance), and do not constrain learning from the beginning.⁵³ In Quine's view, young children who have not yet mastered the quantificational syntax of language use words to pick out salient properties of perceptual experience, not objects or non-solid substances. To undermine Quine's position, Soja, Carey, and Spelke performed a series of experiments with children of 2 to 2½ years, at a level of development prior to mastery of quantificational syntax. These young children were introduced to words in the presence of a bounded, persisting objects of certain types, and then tested to see whether they would generalize by applying the words to new instances of objects of the same type, or to things that were not objects of the same type but produced experiences that shared salient properties with experiences of the original objects. The same types of trials were run for non-solid substances. All the children expressed a very strong predilection to apply the words to similar objects and non-solid substances, and not to shared salient properties of perceptual experiences. Whether or not words for objects were introduced with quantificational syntax, the children always manifested a strong predilection in favor of objects and non-solid substances.

Such an argument rules out an empiricist hypothesis that is a serious rival to an innateness claim. It rules out such a hypothesis by means of thoroughly empirical evidence, and it is difficult to imagine how it could do so without it. One might reasonably conjecture, then, that Kant's inability to demonstrate the innateness of the concepts required for knowledge of

certain claims to necessity and universality, especially the innateness of concepts as weighty as cause and substance, can be traced back to his adherence to an a priori methodology for justification. It may well be that no a priori method can yield justifications for relevant universality claims, and for refutations of rival empiricist hypotheses, essential elements of arguments for the innateness of concepts, especially for the innateness of concepts of the sort Kant has in mind.

5

Kant has thus argued that a priori synthesis accomplishes two tasks: it explains one's knowledge that the subject of different self-attributions is the same, and it accounts for the psychological processing required for representations of objective phenomena. These two explananda are features of any possible human experience, and thus, if successful, Kant's arguments show that any human experience requires a priori synthesis, and consequently, also the concepts employed by this synthesis. The final conclusion is that these concepts legitimately apply in experience. In §§19-20, Kant attempts to establish that judgment is the medium that brings about synthesis, and that this medium demands certain forms of judgment which are connected to the twelve categories. Hence, by connecting synthesis to judgment, Kant proposes to attain the legitimate applicability of the categories in experience.

In section §19, Kant suggests that there must be a certain way in which each of my representations is unified in the subject, and he identifies this way with judgment: "I find that a judgment is nothing but the manner in which given cognitions (Erkenntnisse) are brought to the objective unity of apperception" (B141). In §18 Kant has argued that only synthesis can explain

the objective validity of our representations, their application to objects that are independent of our subjective states for their nature and existence, for only synthesis can account for the universality and necessity on which objective validity depends. In §19, Kant contends that judgement is the vehicle of this type of synthesis. Judgment, he claims, is objectively rather than subjectively valid, and hence exhibits the kind of universality and necessity that underlies objective validity (B142). He then alleges that without synthesis and judgement as its vehicle, an organization of representations might reflect what appears to be the case, but would not allow us to make distinctions between reality and mere appearance.

In §20, Kant links this notion of judgment to his table of forms of judgment, and he in turn ties the forms of judgment to the categories (B143). The claim has often been made that the connections he specifies between synthesis and judgment, judgment and the forms of judgment, the forms of judgment and the categories are underargued.⁵⁴ Guyer argues that Kant has not adequately established the last of these links, that although Kant asserts that the categories are simply the forms of judgment as they are employed in the synthesis of representations in an intuition (A79/B104-A80/B105, B143), he has failed to make this claim plausible. For example, why should the hypothetical form of judgment be linking with the category of cause, and the disjunctive form of judgment with the category of community? Kant's position would seem to require a more thorough justification.⁵⁵

A further problem arises for Kant's attempt to connect synthesis and judgment on the one hand, and judgment and the categories on the other. Suppose, first, that Kant had indeed established the need for synthesis by a priori concepts in §18. Then the crucial function of §§19-20 would be to specify exactly which a priori concepts synthesis employs. But here we

encounter the difficulty. If Kant wishes to conceive of judgment as employing precisely his twelve categories, then he has not established that judgment conceived in this way is the vehicle of the sort of synthesis he has established in §18. And if Kant wants to conceive of judgment as the vehicle of the sort of synthesis he has established in §18, then he has not shown that judgment conceived in this way must employ his twelve categories. We are missing a substantive argument that the type of synthesis established in §18 requires precisely Kant's twelve categories. These difficulties, however, are mitigated by the fact that one would not reasonably have expected Kant to demonstrate the legitimate applicability of precisely his twelve categories by arguments as general as those of the deduction. Rather, it seems appropriate to count the deduction as successful were it to establish the legitimate applicability of some a priori concepts or other, but no particular ones. After all, it is only later in the Critique, in the System of All Principles of Pure Understanding, that Kant genuinely focusses his attention on each of the categories individually.

But secondly, as we have seen, in §18 Kant does not show that the synthesis for which he has argued demands a priori concepts. This problem is not remedied by identifying judgment as the vehicle of this type of synthesis, and then claiming that judgment require the categories. For if Kant wants to conceive of judgment as employing a priori concepts, then he has not established that judgment of this kind is the vehicle of the synthesis established in §18. And if he wishes to conceive of judgment as the vehicle of the synthesis argued for in §18, then he has not shown that judgment of this sort must employ a priori concepts. Kant must supply a substantive demonstration that the synthesis argued for in §18 requires a priori concepts, and again, such an argument would seem to demand evidence more empirical than that allowed by Kant's apriorist

methodology.

6

Transcendental philosophy, Kant's revisionary model for metaphysics, may well be successful for a specific set of philosophical tasks. The structuring of the argument from above is ingenious, and the particular claims Kant makes for a representation-independent subject, and for the causal explanation of the knowledge of the identity of the subject are weighty. Furthermore, the structuring of the argument from below has lasting significance, and its challenge to associationism may well be successful. Moreover, the method for all of these results arguably lies within the epistemological restrictions Kant places on transcendental philosophy. Against this last claim an empiricist might contend that we can know which strategies provide good explanations in transcendental philosophy only on the basis of knowledge of standards for successful explanations in the mature natural sciences, and that the justification for our acceptance of such standards may well be thoroughly empirical. But since this objection is sufficiently difficult to assess, Kant's apriorist position remains a genuine option.

Nevertheless, it would seem that one crucial aim of the deduction, establishing the need for concepts that are a priori, especially the need for a priori concepts as weighty as those of cause and substance, cannot be met solely by the type of justification to which transcendental philosophy is limited. One cannot justify universality claims, or rule out empiricist rivals to innateness hypotheses, without appealing to information derived from thoroughly empirical experiment and observation. Yet if Kant were willing to allow an appeal to results in empirical psychology, the arguments of the deduction might well be successful. We might, accordingly,

submit a positive but qualified assessment of the Transcendental Deduction. The structure of its arguments, beginning with slender premises about any possible human experience, then proceeding to a synthesis by a priori concepts by way of an explanation for these slender premises, and finally arriving at the conclusion that some a priori concepts are legitimately applicable to experience, stands as a model for philosophical strategy. In addition, Kant provides us with two impressive schemes for realizing this structure. Perhaps he failed to appreciate only one requirement for completing these arguments successfully: some assistance from empirical psychology in establishing the need for a priori rather than merely empirical concepts.⁵⁶

NOTES

1. I argue for this characterization of transcendental philosophy in Pereboom (1990, 1991). In my interpretation, justification in transcendental philosophy is a priori in the sense that it is dependent on information derivable from any possible human experience. For example, any possible human experience will yield the information that one can self-attribute one's representations, and that experience is of objective phenomena distinct (in some way) from one's representations of them. Such information is non-empirical in the sense that it is independent of empirical experiments and observations that are not a feature of any possible human experience. This account of a priori justification in transcendental philosophy is inspired by Philip Kitcher (1980, 1981, 1984ab).
2. The term "deduction," therefore, does not denote a deductive argument.
3. Patricia Kitcher (1990, pp. 3-29) argues convincingly that the psychological side of the Critique of Pure Reason has been thoroughly undervalued by commentators for various ideological reasons. One should keep in mind, however, that in the Transcendental Deduction, Kant advances his psychological theory primarily to support a normative claim. (Thanks to Arthur Kuflik for stating the point in this way.)
4. Patricia Kitcher (1990), e.g. pp. 84-6.
5. Patricia Kitcher (1990), pp. 121-2.
6. Patricia Kitcher (1990), p. 122; Daniel Dennett (1969), pp. 93-96.
7. For accounts of the distinction between the genetic and the justificatory senses of apriority, see Patricia Kitcher (1990), pp. 14-21, and Pereboom (1990), p. 26.

8. For an intriguing discussion of the notion of innateness in Kant, see Zoeller (1989). Zoeller argues that Kant's distinction between innate, unacquired representations and representations acquired from an innate source is philosophically significant, but nothing he says impugns the more common use of the term 'innate.'

9. Although at A79/B104 Kemp-Smith renders 'Wirkung' as 'result,' and at B152 as 'action,' it is better translated as 'effect,' in the sense of the correlate of cause. Kemp-Smith correctly translates the various occurrences of 'Ursache und Wirkung' as 'cause and effect' (e.g. A80/B106, B232).

10. For further discussion of this issue see Pippin (1987), Allison (1989), Aquila (1989), pp. 31-2, Patricia Kitcher (1990), p. 253, note 5. Robert Howell, in correspondence, questions whether synthesis can be a causal process, since Kant indicates that it is non-temporal. Kant does maintain, however, that there may be non-empirical causes, and he argues for this view in the Antinomy of Pure Reason (A532-58/B560-86), and in the third section of the Groundwork of the Metaphysics of Morals, Ak IV, 446-46.

11. Patricia Kitcher (1990), p. 117.

12. Patricia Kitcher (1990), p. 103.

13. In correspondence.

14. This is Patricia Kitcher's view of the second step (1990, pp. 142-80), which is indebted to, but also importantly different from Dieter Henrich's (1969). Henrich denies that the second step has an important psychological component, and Kitcher, correctly I believe, disagrees.

15. Edwin McCann first suggested to me this way of dividing up the B-Deduction.

16. This terminology derives from A119.

17. Patricia Kitcher suggests (in correspondence) that I might be somewhat more wary about a representation-independent subject of self-attribution because cognitive psychology doesn't make use of such an entity. But this is so because cognitive psychologists have not focussed their attention on the notion of a subject. Most recent work in cognitive psychology has focussed on the topic of representation. Perhaps the notion of a subject will be in line for discussion soon. Kant's discussion of self-attribution is compatible with recent work in cognitive psychology, although that particular topic has been largely ignored.

18. See Van Gulick (1989), especially pp. 226-228.

19. Smit points out a passage in the Discipline in which Kant it is clear that Kant is using the term in this way:

Consequently, mathematics is the only science that has definitions. For the object which it thinks it exhibits a priori in intuition, and this object certainly cannot contain either more or less than the concept, since it is through the definition that the concept of the object is given -- and given originally, that is, without its being necessary to derive the definition from any other source. (A730/B758)

Smit is developing a view of Kant's methodology in which the notion of 'originality' plays a central role.

20. Shoemaker (in 1986, especially p. 105) analyses Hume's view that we can have no introspective awareness of the self as the claim that we cannot represent the self as having intrinsic (i.e. non-relational) properties.

21. The size, shape, and color of the car, in Kant's view, are relational properties of the car, since the car's having them involves a relation to the faculties of the subject, faculties such as sensation and pure intuition. But yet the car's color and size are ordinarily experienced as intrinsic properties of the car, in contrast with, for example, its being in the garage, which is experienced as a relational property. Hence, to adapt Shoemaker's suggestion to Kant's view, we might make a distinction between properties which appear as or are experienced as intrinsic to a thing and properties that appear as or are experienced as non-intrinsic or relational.

22. Patricia Kitcher suggests (in correspondence) that my interpretation of the argument from above has Kant begging the question against Hume because the basic premise in the argument is that the subject of different self-attributions is the same, while Hume denies that there is such a single subject. But first, Hume does not simply assert that there is no such single subject. Rather, he makes this claim on the basis of finding no impression that corresponds to it. In fact, he agrees that our ordinary notion of the self is "that to which our several impressions and ideas are supposed to have a reference," but then, because he finds no impression of such a self, he concludes there is no self that matches this notion (Treatise, pp. 251-3). As we shall see, in §16 Kant agrees with Hume that there is no impression that can secure this ordinary notion of the self. However, there he also argues against Hume that there is another way in which the self can be represented as identical, a way that involves a priori synthesis. Thus Kant and Hume agree on the nature of our ordinary notion of the self, but while Hume contends that there is no such self because we have no impression corresponding to it, Kant argues that we have another way of representing the self as identical. The best case for Kant's begging the question against Hume here is that Kant assumes without argument that Hume's empiricist theory of representation is

false. But Kant's argument in §16 can itself be viewed as an argument directed against this theory (and as a part of Kant's broader strategy against this view). We all start with the notion of a self that is the same for different self-attributions. Hume's empiricist theory of representation cannot account for how we represent this identity, whereas Kant's can, and this provides a reason to reject Hume's view and to accept the Kantian alternative.

23. Guyer (1987), pp. 138, 146.

24. The alternative is that if I could not become conscious of a representation, it would not exist. See Paul Guyer's (1987, pp. 139ff.) fine discussion of this issue.

25. Robert Howell has pointed out (in correspondence) that it does not follow from the fact that I am conscious of a representation that I can self-ascribe it. There might, for example, be mental disease which does not affect one's capacity for consciousness of one's representations but impairs the ability to self-ascribe. But supposing normal human mental functioning, it would seem that any representation of which I am conscious is also one I can self-ascribe.

26. At B135, Kant claims that "The principle of the necessary unity of apperception is itself, indeed, an identical, and therefore analytic, proposition." Edwin McCann (1985, pp. 73-4) provides an enlightening discussion of this claim. Henry Allison argues (1983, p. 171) that a plausible interpretation of the first part of the second-edition deduction would not only preserve the analyticity of this principle, but it would also render its entire argument analytic. I argue in Pereboom (1990) that Kant's view that this principle is analytic cannot be justified, but I do not see how this weakens or changes the character of the argument of the deduction. One should note that in the first edition (A118), Kant claims that a very similar principle is synthetic. I argue, in addition, that Kant never

claims that the argument as a whole is analytic, that he frequently expresses his disdain for analytic procedure, and that the argument is implausibly reinterpreted as analytic. Patricia Kitcher also favors the view that the argument of the deduction is synthetic (1990, p. 172). The analytic interpretation, I believe, is an attempt to read the program of logical positivism into Kant's views (see Philip Kitcher (1992, pp. 53-56).

27. See Guyer (1987), pp. 139-144, for a discussion of these issues.

28. The language of B133, the core of the argument from above, indicates that Kant implicitly acknowledges that the most significant claim for this argument is for the identity of the subject for different self-attributions, and not for the identity of the subject for all my self-attributions. For example, in the summary sentence Kant asserts "Only in so far , therefore, as I can unite a manifold of representations in one consciousness, is it possible for me to represent to myself the identity of the consciousness in [i.e. throughout] these representations (B133). He does not say, for example "Only in so far as I can unite all of my representations in one consciousness..."

29. See Pike (1967).

30. Thus, in my view, the crucial sentence "the empirical consciousness, which accompanies different representations, is dispersed and without relation to the identity of the subject" (B133) means (at least) that Hume's psychological theory cannot account for our knowledge or representation (or consciousness) of the identity of the subject of different self-attributions. The sentence is intended to have two important implications for Hume's theory. First, without a representation-independent subject, Hume's theory cannot account for the fact that the subject of different self-attributions is identical, which must indeed be a fact if I am to know or represent it.

Second, without a priori synthesis, my ability to know or represent this identity has no explanation. It is important to note that accounting for my representation of the identity of the subject is not the same task as explaining how I might take each of various distinct representations of mine to be mine.

31. See Shoemaker (1986).

32. Allison (1983), pp. 142-4; see also Guyer (1987), pp. 133-139.

33. Guyer claims that in the argument of §16, Kant does not "pursue the special promise of a direct connection between apperception and the synthetic nature of all thought," and instead "makes no legitimate use of the concept of apperception", and "relapses into a form of the deduction from the conceptions of object and judgment which is not significantly different from that offered in the Prolegomena" (1987, p. 114), a version Guyer argues to be a failure. As we have just seen, however, in §16 Kant does conclude that a priori synthesis must be what produces the unity of representations required to explain my knowledge of the identity of the subject for different self-attributions; "synthetic unity of the manifold of intuitions, as generated a priori, is thus the ground of the identity of apperception itself" (B134). But in §16 Kant does not conclude that precisely his twelve categories apply to experience; that he reserves for §§19-20.

34. A passage which more strongly suggests consciousness of the act of synthesis occurs at A108. See Allison's (1983, pp. 142-3) discussion of these issues.

35. Patricia Kitcher (1990), pp. 117-121; cf. McCann (1885), pp. 75-6.

36. This identity of the subject should not be understood as the identity of a substance, an entity

that is, among other things, a fundamental causal power, for Kant does not believe that we have knowledge of the subject as a substance, or for that matter, of its identity conditions (A361ff, B408; Pereboom (1991)). Rather, this subject, at this stage, should be conceived simply as a subject of mental representation, and in particular, of self-attribution. Given Kant's terminology, therefore, he might well not want to construe this discussion as concerned with personal identity (e.g. A363). Whether this discussion actually is about personal identity, supposing the contemporary understanding of this notion, is a matter of controversy.

37. Allison (1983), p. 143.

38. Thanks to Hilary Kornblith for helping me formulate this issue more clearly.

39. This situation realizes the following schema: Recognizing or knowing that A explains recognizing or knowing that B, because B is part of the best explanation of A. Another example: my knowledge that there are dinosaur bones in Alberta explains my knowledge that there were once dinosaurs in Alberta, because that there were once dinosaurs in Alberta is part of the best explanation for there being dinosaur bones in Alberta.

40. Guyer (1987), pp. 146-7.

41. Allison (1983), p. 144.

42. At A108 Kant says:

The original and necessary consciousness of the identity of the self is thus at the same time a consciousness of an equally necessary unity of the synthesis of all appearances according to concepts, that is, according to rules, which not only make them necessarily

reproducible but also in so doing determine an object (Gegenstand) for their intuition, that is, the concept of something wherein they are necessarily interconnected.

See also Richard Aquila's (1989, pp. 158-176) fine discussion of self-consciousness.

43. Allison (1983), p. 145.

44. Allison (1983), p. 146; cf. pp. 27, 118-9, 135.

45. Allison (1983), p. 145.

46. Aquila (1989), p. 159.

47. It is important to distinguish metaphysical/causal claims about relations among synthetic processes from claims about the structure of the arguments regarding these processes. I agree with the following metaphysical/causal claim: a type of process -- a priori synthesis -- which accounts for the unity of apperception also accounts for the objectivity in experience that Kant discusses in §§17-19. This is one implication of the first sentence of §18: "The transcendental unity of apperception is that unity through which all the manifold in an intuition is united in the concept of an object. Two further passages in the second-edition deduction, in my view, make the same point. (Howell argues, in correspondence, that these passages support something like Allison's general schema of the argument.) In §19 Kant discusses "...principles of the objective determination of all representations, in so far as knowledge can be acquired by means of these representations -- principles which are all derived (abgeleitet) from the fundamental principle of the transcendental unity of apperception" (B142), and in §27 Kant describes the "determination of appearances in space and time" as "ultimately out of (endlich dieser aus) the original synthetic unity of apperception" (B168-9). I

suggest that these passages convey the view that the principles which govern objectivity in experience are metaphysically dependent on ((out of (aus), derived (abgeleitet) from) the principle of the unity of apperception (Kant uses abgeleitet in this metaphysical/causal sense at B140), meaning that the fundamental aim of a priori synthesis is to provide unity of apperception, and in doing so a priori synthesis also produces objectivity in experience, or that synthesis produces objectivity in experience in order to provide unity of apperception. These relations of metaphysical priority do not entail that in the second-edition deduction the principles governing objectivity in experience are demonstrated solely on the basis of the principle of the unity of apperception. One might read these passages as making this claim about the process of demonstration in the second-edition deduction, but the plausibility of this reading is dependent on whether there is a respectable argument in §§17-20 which fits this model.

48. Guyer (1987), pp. 11-24.

49. Guyer (1987), p. 105. One should also note that in the deduction Kant does not argue for a priori knowledge of certain concepts, but rather for the objective validity of concepts that are a priori in the genetic sense that their origin is not in experience but in the structure of the mind.

50. The type of viewing Kant is concerned with here is a continuous scanning of the different parts of an assembled house. If, for example, photographs of the different parts were shown to me one by one, or if I were discontinuously shown the parts of a disassembled house over a period of a year, then I might well not represent these parts as simultaneous.

51. See, for example, Chomsky (1965).

52. Soja, Carey, and Spelke (1991). For an example of a contemporary argument that provides evidence for the innateness of certain causal concepts, see Gelman and Wellman (1991).

53. Quine (1960, 1969).

54. See Guyer's incisive discussion of these issues, (1987), pp. 94-102.

55. Guyer (1987), pp. 98-9.

56. I wish to thank Robert Adams, Henry Allison, Tyler Burge, David Christensen, Mike Jacovides, Amy Kind, Alan Kim, Patricia Kitcher, Arthur Kuflik, Hilary Kornblith, William Mann, Houston Smit, Brian Van Norden, and Carol Voeller for instructive commentary and discussion on drafts of this paper. I am especially indebted to Robert Howell, who supplied very helpful and detailed written comments on two occasions.

REFERENCES

Kant's Critique of Pure Reason is cited in the traditional way: 'A' indicates the first edition, 'B' indicates the second edition. All citations are taken, with some alterations, from Immanuel Kant's Critique of Pure Reason, tr. Norman Kemp Smith, London: Macmillan, 1929. Other abbreviations used are:

Ak Immanuel Kant, Kant's gesammelte Schriften, edited by the Königliche Akademie der Wissenschaften and its successors, George Reimer, subsequently W. de Gruyter, 1902- , Berlin.

Ak VIII Quotations from Ak VIII are from the translation of On a Discovery in The Kant

Eberhard Controversy by Henry Allison, The Johns Hopkins University Press, 1973, Baltimore.

Treatise David Hume, A Treatise of Human Nature, L.A. Selby-Bigge, ed., Oxford University Press, 1962, Oxford.

Enquiry David Hume, An Enquiry Concerning Human Understanding, Eric Steinberg ed., Hackett, 1981, Indianapolis.

Allison, Henry: 1983, Kant's Transcendental Idealism, Yale University Press, New Haven.

Allison, Henry: 1989, 'Kant's Refutation of Materialism', The Monist 72, 190-208.

Aquila, Richard: 1989, Matter and Mind, Indiana University Press, Bloomington.

Chomsky, Noam: 1965, Aspects of the Theory of Syntax, MIT Press, Cambridge, Mass.

Dennett, Daniel: 1969, Content and Consciousness, Routledge and Kegan Paul, London.

Gelman, Susan A, and Henry M. Wellman: 1991, 'Insides and Essences: Early understandings of the non-obvious', Cognition 38, 213-244.

Henrich, Dieter: 1969, 'The Proof-Structure of Kant's Transcendental Deduction', Review of Metaphysics 22, 640-59

Henrich, Dieter: 1989, 'Kant's Notion of a Deduction and the Methodological Background of the First Critique', in Kant's Transcendental Deductions, Eckart Förster, ed., Stanford University Press, Stanford, pp. 29-46.

Guyer, Paul: 1987, Kant and the Claims of Knowledge, Cambridge University Press, Cambridge.

Kitcher, Patricia: 1990, Kant's Transcendental Psychology, Oxford University Press, Oxford.

Kitcher, Philip: 1980, 'A Priori Knowledge', Philosophical Review 89, 3-23.

Kitcher, Philip: 1981, 'How Kant Almost Wrote `Two Dogmas of Empiricism"', Philosophical Topics, v. 12, n. 2, 217-249.

Kitcher Philip: 1984a, 'Kant's Philosophy of Science', in Self and Knowledge in Kant's Philosophy, Cornell University Press, Ithaca, pp. 157-73.

Kitcher, Philip: 1984b, The Nature of Mathematical Knowledge, Oxford University Press, Oxford.

Kitcher, Philip: 1992, 'The Naturalists Return', Philosophical Review 101, 53-114.

McCann, Edwin: 1985, 'Skepticism and Kant's B Deduction,' History of Philosophy Quarterly 2, pp. 71-89.

Pereboom, Derk: 1990, 'Kant on Justification in Transcendental Philosophy', Synthese 85, 25-54.

Pereboom, Derk: 1991, 'Is Kant's Transcendental Philosophy Inconsistent?', History of Philosophy Quarterly 8, 357-72.

Pike, Nelson: 1967, 'Hume's Bundle Theory of the Self: A Limited Defense', American Philosophical Quarterly 4, 159-165.

Pippin, Robert: 1987, 'Kant on the Spontaneity of Mind', Canadian Journal of Philosophy 17, 449-76.

Quine, W. V. O: 1960, *Word and Object*, MIT Press, Cambridge, Mass.

Quine, W. V. O: 1969, Ontological Relativity and Other Essays, Columbia University Press, New York.

Shoemaker, Sydney: 1986, 'Introspection and the Self', Midwest Studies in Philosophy X, University of Minnesota Press, Minneapolis, pp. 101-120.

Soja, Nancy, Susan Carey, and Elizabeth Spelke: 1991, 'Ontological categories guide young children's inductions of word meanings: Object terms and substance terms', Cognition 38, 179-211.

Van Gulick, Robert: 1989: 'What Difference Does Consciousness Make?' Philosophical Topics XVII, 211-230.

Zoeller, Guenter: (1989), 'From Innate to A Priori: Kant's Radical Transformation of a Cartesian-Leibnizian Legacy', The Monist 72, 222-235.