

## KANT'S METAPHYSICAL AND TRANSCENDENTAL DEDUCTIONS

in *A Companion to Kant*, Graham Bird, ed., Oxford: Blackwell Publishers (2006), pp. 154-68.

Derk Pereboom

### Introduction

The *transcendental deduction* (A84-130, B116-169) is Kant's attempt to demonstrate, against empiricist psychological theory, that a priori concepts correctly apply to objects in our experience. What makes a concept a priori is that its source is the understanding of the subject and not in sensory experience (A80/B106). Dieter Henrich (1968-9) points out that '*Deduktion*' is originally a legal term; it denotes an argument intended to provide a historical justification for the legitimacy of a property claim. In Kant's derivative epistemological sense, a deduction is an argument intended to provide a justification for the legitimacy of a concept, one that shows that the concept correctly applies to things.

David Hume attempts a deduction for various metaphysical ideas that he thinks are suspect -- the idea of causal power, for example (1748: §7). In his view, a deduction will only be successful when a sensory experience, an impression, of causal power is found. Since the search for an impression of causal power is fruitless, Hume concludes that the idea does not truly apply to things. In Kantian language, Hume here attempts an *empirical* deduction (A85/B117), and from its failure he concludes that the concept of causal power has no *objective validity*, that is, it fails to apply to the objects of our experience.

In the transcendental deduction, Kant attempts a different sort of deduction for twelve concepts, one of which is the concept of cause, all of which he believes to be a priori. These a

priori concepts are the *categories*. A transcendental deduction begins with a slender premise about any possible human experience, a premise to which reasonable participants in the debate can agree. Kant's argument attempts to establish a particular theory of mental processing by showing that its truth is a necessary condition for the truth of such a slender premise. As I shall contend, in his strategy involves employing two such premises; one concerns self-consciousness, the other certain characteristics of our representations of objective features of reality. Kant then aims to demonstrate that the categories have an essential role in this sort of mental processing. In his idealist view, the objects of experience result from this mental processing, and it is due to the role that the categories have in this processing that they correctly apply to these objects. Thus in the transcendental deduction he intends to secure a normative claim, that the categories correctly apply to the objects of our experience, by establishing a psychological theory (Kitcher 1990: 2-29). One should note that Kant's transcendental deduction presents *general* considerations supporting the applicability of all the categories to the objects of our experience; it does not concentrate on the applicability of specific categories (Bird 1962/1973: 112-5). This more focused task is taken up in the Analytic of Principles (A130-235/B169-287).

In the *metaphysical deduction* (A66-83, B92-116) Kant sets out to derive the categories from what he calls the *logical forms of judgment*. The metaphysical deduction has a key role at a specific point in the transcendental deduction, and we will discuss its claims when we reach that point.

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 processing or ordering of mental states, but he does not believe that such an account requires a priori concepts. According to his theory,

*associationism*, our mental repertoire consists solely of perceptions, all of which are sensory items – the more vivid impressions, and their less vivid copies, the ideas, which function in imagination, memory, reasoning, and conceptualization (1748, §§2, 3). Association proper is the process by which these perceptions are ordered. The hallmark of this theory is that mental processing requires no resources beyond what perceptions provide; how perceptions are ordered is solely a function of the perceptions themselves. A subject that is distinct from these perceptions cannot have a role Hume's picture, since for him the subject is merely a collection of perceptions (1739: Bk. I, IV, vi).

In Kant's theory, the ordering of mental states most prominently features the process of *synthesis*. Synthesis is "the act of putting different representations together, and grasping what is manifold in them in one cognition" (A77/B103); it is that which "gathers the elements for cognition, and unites them to form a certain content" (A78/B103). Synthesis is a process by which multiple representations – in Kant's term, a 'manifold' – are connected with one another to form a single further representation with cognitive content. This process can employ concepts as modes of ordering representations. Crucial to the transcendental deduction is the claim that it is the categories by means of which our representations are synthesized. Since the understanding of the subject is the source of the categories, and also a faculty that produces synthesis, the subject has a central role in mental processing. For Kant this subject is distinct from its representations, as I shall argue.

This discussion will focus on the transcendental deduction in the second edition (1787) of the *Critique of Pure Reason* – the *B-Deduction*. I shall argue that in §§16-20 of the B-Deduction Kant employs a two-pronged strategy for defeating associationism and establishing synthesis.

The first argument, contained in §16, is designed to show that association lacks adequate resources for explaining an aspect of self-consciousness, and that synthesis is required to provide this explanation. This type of argument is appropriately called an *argument from above* (A119). Correlatively, in §§17-20 we find an *argument from below*, by which Kant aims to demonstrate that synthesis by means of the categories is needed to explain certain features of how we represent objects.

### Apperception

The argument from above in §16 divides into two stages. The first aims to establish the various features of *the principle of the necessary unity of apperception*. The second advances to a priori synthesis by explaining how we might grasp an aspect of self-consciousness that this principle highlights. *Apperception* is the apprehension of a mental state, a representation, as one's own. In his view, my apperception has necessary unity since all of my representations 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:

*(The principle of the necessary unity of apperception)* 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 can be conscious of its representations (A116, B131-2, B134-5).

Consider three observations about the meaning of this principle.

(i) Kant maintains, against Hume, that the apperceiving subject is not a collection of representations. In §16 he 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). If intuition is distinct from the 'I' in this sense, then, plausibly, intuitions are not components of this subject. Furthermore, Kant would be implausibly interpreted as holding that the 'I' consists merely of a collection of concepts. Supposing that all representations are either intuitions or concepts, it follows that the 'I' does not consist merely of 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, he affirms that I have no inner intuition of the subject (e.g. B157), and this claim would be at odds with the subject's being a collection of representations, since he maintains that I can intuit my representations by inner sense (e.g. A33/B49).

(ii) The ability to attribute my representations to a subject is *pure*, rather than *empirical* apperception. This means, in part, that I cannot attribute my representations to a single subject just in virtue of Humean inner perception, or Kantian empirical inner intuition. However, 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). Pure apperception is original since I am not conscious of the self-consciousness, the apperceiving I that is the subject of apperceptive thoughts, in any manner independent of what is contained in these thoughts. I cannot have an intuition or any other type of representation of this subject other than by 'I think...'- type thoughts, and thus, these thoughts

are the original representations of this subject (e.g., A350). Nevertheless, in virtue my capacity for apperception, I can have a type of propositional grasp of the apperceiving subject; in apperception, I am conscious *that* I exist as subject(B157).

Kant begins the first stage of the argument in §16 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 I could not attach the 'I think' to it is just that I could not become conscious of it (Guyer 1987: 139-44) 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. One might even understand why Kant would think a claim like this to be an analytic truth. 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. Some of my representations are thoroughly subconscious, while they should still be classified as mine in virtue of the types of causal relations they bear to my perceptions, my behavior, and representations 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, and my being conscious of this identity.

Several commentators maintain that Kant's argument requires the unity of apperception to be a claim about simultaneous consciousness of representations, and that this undermines its soundness. For instance, one of Robert Howell's primary difficulties with the argument of §16 is that Kant does not establish what he considers to be a crucial premise

(S) All of the elements of the manifold of *i* (where *i* is some arbitrary intuition) are such that *H* is or can become conscious, in thought, that all of those elements, taken together, are accompanied by the *I think*

that is, all the individual elements of the intuition are such that the subject can become conscious of them simultaneously, by contrast with a weaker claim

(W) Each element of the manifold of *i* is such that *H* is or can become conscious, in thought, that the *I think* accompanies that element. (Howell 1992: 161)

which is to say that all the individual element of the intuition are such that the subject can become conscious of each (in turn). Howell supposes that if (S) cannot be established, the argument of the B-Deduction collapses, for only if Kant can show that the different elements of a manifold together and at the same time are accompanied by the same *I think* can he establish that *H*'s mind must synthesize these elements (Howell 1992: 162). He contends that, by contrast, the unity expressed by (W) is insufficient to generate this need for synthesis. Howell goes on to argue that Kant cannot in fact demonstrate (S) – and it is indeed implausible that such co-consciousness for any arbitrary intuition is really possible for us -- and that hence the argument of the B-Deduction falters.

But perhaps Kant does not require a premise as strong as (S) — as opposed to (W) -- for the argument of §16. First of all, although the co-consciousness claim is *suggested* in §16 by the following sentence:

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 (B133)

it is not clearly stated here. By “am conscious of the synthesis of them,” Kant might mean that I am conscious that these representations stand in a certain relation to one another, which need not involve actually my being conscious of them at the same time. I might be conscious that my representations are inferentially integrated with each other in a way distinct from how mine are integrated with yours, without being conscious of all of the representations that are so integrated at the same time. Moreover, perhaps Kant intends not that I am *actually* co-conscious of these representations, but that they *could* become co-conscious for me. This interpretation is suggested by Kant’s assertion:

For without such combination nothing can be thought or known, since the given representations would not have in common the act of apperception ‘I think’, and so *could* not be apprehended together in one self-consciousness. (B137, emphasis mine).

James van Cleve argues, however, that if the representations can only possibly become co-conscious, Kant can only conclude is that they are possibly subject to the categories (van Cleve 1999: 84). Is this a correct diagnosis?

At this point the argument of §16 features a subtlety that is often overlooked. In fact, the central feature of this argument is Kant’s attempt to demonstrate that only a priori synthesis can -

explain *how I might represent the identity of my apperceptive consciousness* (B133) or *how I might represent the identity of the apperceiving subject* (B135) for different elements of the manifold of intuition to which I can attach the *I think*. The difficulty Kant 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 (*und ohne Beziehung auf die Identität des Subjects*)" (B133). One implication of this key passage is that Hume's theory 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 one would not want to initially deny, that the conscious subject of different apperceptive self-attributions is the same. 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. Hume might propose to explain our sense that the conscious subject of different self-attributions is the same by the perceptions of perceptions being elements of a single causally coherent bundle. Still, the bundle is not conscious of perceptions; consciousness of perceptions would be a feature of individual perceptions. In Kant's view, explaining how the conscious subject of different self-attributions can be the same requires that this subject be distinct from its representations.

The second stage of the argument of §16 involves a further implication of the claim that "the empirical consciousness, which accompanies different representations, is dispersed and without relation to the identity of the subject" (B133): 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). One might envision several types of explanation for my representation of this identity. One possibility is that inner sense enables me to represent this identity for each of my various self-attributions, and the way I represent the identity of the subject is similar to the way I typically represent the identity over time of ordinary objects -- by noting similarities among the intrinsic properties represented. But Kant and Hume would agree that this is not the way I represent the identity of the apperceiving subject, since both would agree that by inner sense I do not represent any intrinsic properties of such a subject. The second type of explanation, which Kant endorses, is that I have a less direct way of representing this identity. As Henry Allison points out, this representation must instead depend on my apprehending a feature of my representations (Allison 1983: 142-4; Guyer 1987: 133-9). The relevant feature is some type of unity or ordering of these states. Kant's idea is that if the representations I can attribute to myself possess a unity of the appropriate type, and if I apprehend this unity, then I can represent the apperceiving subject of any one of them as identical with that of any other.

Thus 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). What sort of unity must I recognize among my representations that would account for my representation of this identity? Need it be actual co-consciousness? Note that I represent the subject as identical for self-attributed representations that are not co-conscious, so actual co-consciousness could not explain generally how I represent this sort of identity. A plausible alternative is that the unity consists in certain intimate ways in which representations in a single subject are typically related. Perhaps the key aspect of this unity is that a single subject's

representations are inferentially integrated to a high degree, by contrast with representations across discrete subjects. This integration might in turn be analyzed, at least in part, by the *possibility* of my representations becoming co-conscious. Memory and the capacity to become conscious of non-occurrent states are means by which representations can become co-conscious.

Association and synthesis are possible explanations for the how this unity comes about 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 adduced 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). Since the understanding provides concepts for synthesis, and since for synthesis to be a priori in this context is, in part, for it to employ a priori concepts, Kant is contending here that synthesis by means of a priori concepts is required to account for my production of or recognition of the unity at issue.

As Paul Guyer contends, however, demonstrating 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 the unity among my representations (Guyer 1987:146-7). It might be, for all Kant has shown, that this recognition requires awareness of information derived from inner experience. Kant does not at this point attempt to dislodge such rival empiricist hypotheses, and it seems he would need to do so to confirm the need for a priori synthesis. To advance his claims, he might try to point out features of this unity that would resist such an

empiricist account. As we shall see, Kant employs this tactic in his account of our representations of objects.

### Representations of objects

One 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. Kant's proposal is that an object is "that in the concept of which a manifold of a given intuition is united" (B137). I think that here we should read 'object' in the broad sense of *objective feature of reality* – a feature whose existence and nature is independent of how it is perceived (Bird 1962/1973 130-1; Guyer 1987: 11-24). Such objective features might be physical; they could also be mental. According to Allison's interpretation, §17 does not simply contain this challenge to Hume, but also a demonstration of our representation of objects on the basis of the claims about self-consciousness developed in §16. This reading is part of Allison's broader picture, in which Kant establishes that the unity of apperception entails the representation of objects, and, conversely, that the representation of objects entails the unity of apperception (Allison 1983: 144ff.) The crucial claim for Allison's interpretation is that the unity of apperception is not only a necessary but also sufficient condition for the representation of objects. Other commentators, including Aquila (1989: 159) and Howell (1992: 227-8), agree. In my interpretation, by contrast, the unity of apperception, 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 proposes in §17, and only aims

to demonstrate in §§18-20.

By Allison's reading, §§15-20 comprise a single argument whose only assumptions premises about self-consciousness that Kant defends in §16. This interpretation of the B-Deduction is widespread. Demonstrating that we represent objects has 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. However, there are reasons to be concerned about this interpretation. For instance, 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 he never attempts to establish by argument that our representations of objects have such characteristics, either in §17 or elsewhere in the *Critique of Pure Reason*. 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.

In Allison's view, the argument from the unity of apperception (or, equivalently, of consciousness) for the existence of representations of objects is found in the following passage:

(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. He argues, first of all, that 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...

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 given this picture, Allison points out, Kant's reasoning seems to involve a gross *non sequitur*, since (2) would support only the claim that the unity of consciousness is a necessary condition for the representation of an object, and not that it is also a sufficient condition. Howell voices a similar worry: "In §17 Kant simply does not make this inference clear, and an air of blatant fallacy hovers over this part of his reasoning" (Howell 1992: 228)

Allison and Howell both venture that Kant's sentence in paragraph (A)

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

should be read as a statement of the sufficiency claim. In (A) Kant contends that cognitions of objects consist in some determinate relation of representations to objects, and as (1) indicates, this relation is constituted or produced by a synthesis that crucially involves the unity of consciousness. But (1) does not entail that the synthesis that involves unity of consciousness cannot take place without its resulting in a relation of a representations to an object. By analogy,

the smelting and molding of steel are processes that constitute or produce steel girders, but it does not follow that the processes of smelting and molding steel cannot take place without the production of steel girders. Just as producing steel girders requires in addition molds of particular shapes, so producing representations of objects could require particular concepts of objects in addition to the synthesis that involves the unity of consciousness.

How might §17 function in the argument absent the sufficiency claim? In my view, the main role of this section is to provide a characterization of an object that incorporates a challenge to Humean associationism and thereby initiates an argument from below. Kant proposes that an object is "that in the concept of which a manifold of a given intuition is united" (B137). In his view, this unification of a manifold requires the unity of apperception together with synthesis, 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). This depiction is designed to present his anti-Humean theory about the processing required for a representation of an object. This characterization does not set out a position that Kant expects his readers to accept without argument, but rather, one he intends to establish in §§18-20.

#### Universality and necessity

In §18 Kant continues the argument against association and for synthesis by drawing our attention to certain features of our representations of objects:

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)

Here Kant characterizes the empirical unity of consciousness as non-universal, contingent, and subjectively valid, distinguishing it from the transcendental unity of apperception, which he describes as universal, necessary, and objectively valid. The empirical unity of consciousness is an ordering of representations produced by association. Hence, Kant is maintaining that association can achieve only an ordering that is non-universal, contingent, and lacks objective validity. The transcendental unity of apperception, by contrast, results in an ordering of representations produced by synthesis, and accordingly, Kant is claiming that synthesis can generate an organization that is universal, necessary, and objectively valid.

In this argument, a key feature of certain representations of objects is their objective validity. For a representation to be objectively valid, it must be a representation of an objective feature of reality, that is, a feature whose existence and nature is independent of how it is perceived (Guyer 1987:11-24). In the argument of §§17-18, Kant assumes that the representations that make up experience are objectively valid. He then argues that association is inadequate

because it can yield only representations that are not objectively valid.

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). Kant here invokes considerations about the ordering of phenomena in time that foreshadow the discussion of the Second Analogy. There Kant argues that our representations, considered independently of their content, are always successive. What is represented by these successive representations, however, is not always itself successive. When viewing a boat floating downstream, its various positions are represented as successive, but in scanning the parts of a house from the roof to ground, these parts are represented not as successive but as simultaneous. The parts of the house are represented as objectively simultaneous, while the positions of the boat are represented as objectively successive. How do we account for this difference?

In §18, Kant implies that an important clue for answering this question is that these representations of objective simultaneity and succession are 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.

Now (5) can be denied by Hume, for example, if we take the necessity to range over all possible circumstances, because his theory can countenance the possibility of a deviant ordering in unusual empirical conditions. But (5) can be recast as

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

Kant's view would then be that given only the resources of association, the truth of (5') could not 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).

Guyer charges that at various places in the transcendental deduction Kant illegitimately assumes knowledge of necessity, and perhaps this argument falls to such a concern (Guyer1987: 146-7). However, Hume might well not deny the necessity at issue. For example, he 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. (1748: §5)

Thus Hume himself contends that given certain specific empirical circumstances, a particular type of ordering of perceptions necessarily or unavoidably comes about.

To forestall a further objection, asserting (5') does not amount to presupposing that one of the categories, viz., necessity, correctly applies to the objects of experience. Although (5') does make a claim to necessity or unavoidability, it does not presuppose that the 'necessity' in the premise is an a priori concept. Moreover, the objection that no premise for the transcendental deduction may employ concepts whose applicability to experience Kant aims to establish is unreasonably 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 a further concern of Guyer's, that Kant merely assumes that all knowledge of necessity is grounded in a priori concepts. One need not interpret him as arguing directly from (5') to the claim that the categories correctly apply to objects in our experience. Rather, one should see him as advancing his claim for the applicability of the categories by ruling out association as an explanation for (5'). One might divide the next point he sets out to reach into three steps:

- (6) To explain the truth of (5'), there must exist a mental faculty for ordering representations.
- (7) This faculty does not consist solely of sensory items.
- (8) This faculty must employ a priori concepts, the categories in particular.

Kant's challenge is to explain why, under normal conditions, the ordering in question is universal and necessary. Part of the best explanation, he believes, is (6), that there must exist a faculty for

ordering the representations. Hume might agree with this conclusion, as long as a sufficiently thin conception of 'faculty' is permitted. But he would deny (7), that this faculty does not consist solely of sensory items. Kant argues that the Humean proposal for a faculty that consists solely of sensory items, the faculty of association, cannot account for the truth of propositions such as (5'). He asks us to consider an activity, word association, which functions as a paradigm for association. 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). Hume's own paradigm for association in is the relations among parts of a conversation (1748: §3). Kant's point is that in such conversations, people make different associations in the same circumstances. If the very paradigms for association fail to exhibit necessity and universality, then the hypothesis that association is powerful enough to yield such an ordering of representations – wherever we find it – on insecure ground.

The associationist might argue that sensory experience is sufficiently uniform for association to produce the universalities and necessities at issue. Perhaps Kant was too quick to conclude that the argument from universality and necessity was decisive, for in addition, associationist objections of this sort must be answered – as contemporary discussions of proposals for innate concepts indicate (Pereboom 1995: 31-3). But this does not detract from the considerable anti-associationist force provided by the sorts of universalities and necessities Kant has in mind, and this fact is also recognized by the contemporary discussion.

Logical forms of judgment and categories

In §§19-20, Kant argues that judgment is the medium that brings about synthesis, and that

this medium employs certain forms of judgment, which are in turn appropriately connected to the twelve categories. By linking synthesis to judgment in this way, Kant aims to show that we use the categories in the synthesis of experience.

In §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 are brought to the objective unity of apperception" (B141). Judgment, Kant claims, is objectively rather than subjectively valid, and hence exhibits the kind of universality and necessity that characterizes objective validity (B142). He then claims that without synthesis and judgment as its vehicle, an ordering of representations might reflect what appears to be the case, but would not allow us to make distinctions between objects and the subjective states they induce.

In §20, Kant links this notion of judgment to the twelve forms of judgment presented in the metaphysical deduction (A70/B95), and he in turn ties these forms of judgment to the twelve categories (A76-83/B102-9). 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 under-argued. Guyer, for example, 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 (Guyer 1987: 94-102). It is fair to say that these concerns have merit. Kant's claims about these connections continue to be more obscure than the preceding part of the transcendental deduction, and it continues to be a serious challenge for interpreters to clarify and vindicate them.

Béatrice Longuenesse (1998), in her interpretation of the metaphysical deduction, has taken up this challenge. In her view, the faculty at issue in the production and use of concepts, the understanding, is the power to judge (*Vermögen zu Urteilen*), which is ultimately a disposition or a *conatus* to make judgments and to shape the what affects us so that we can make them (Longuenesse 1998: 208, 394). The logical forms of judgment are, most fundamentally, the forms of combination of concepts in judgments. One such form is the categorical, which is the form of subject-predicate judgments; another is the hypothetical, the form of if/then judgments. Kant maintains that the logical form of a judgment is what makes it capable of truth or falsity; it is that by which a judgment expresses the relation of a representations to an objective feature of reality (Longuenesse 2000, 93-4). He thinks, for instance, that by virtue of its categorical form, the judgment ‘the boat is moving’ (whether or not it is true) expresses the relation of my representations of a boat and of motion to an objectively existing boat in motion.

One paradigmatic role of the logical forms of judgment is in the process of analysis, by which, for example, the objects we intuit are subsumed under concepts. What results from this process is a judgment that expresses an *analytic unity* – in this case the subsumption of objects under a single concept. But a logical form of judgment can also be called into service to perform a different function: the synthesis of a manifold of intuition, which now results in a *synthetic unity*. The understanding, as the power to judge, has both roles; “the same function that gives unity to concepts in judgment, also gives unity to the mere synthesis of representations in intuition” (A79/B104-5). When the logical form has the synthetic role, it is transformed or expressed as a category:

The same understanding, and indeed by means of the very same actions through which it

brings the logical form of judgment into concepts by means of the analytical unity, also brings a transcendental content into its representations by means of the synthetic unity of the manifold in intuition in general. (A79/B105).

The content added to the logical forms of judgment is some feature of intuition – more precisely, of the forms of intuition – that emerges when the power to judge sets out to unify a manifold of intuition (B128-9). The categories are thus generated from the forms of judgment in the process of synthesizing intuitions. For example, what is generated from the categorical form of judgment is the category of substance, and from hypothetical form of judgment is the category of cause.

The second step of the B-Deduction

In §20 Kant concludes: "Consequently, the manifold in a given intuition is necessarily subject to the categories" (B143). It is tempting to think that this is precisely what Kant intended to show in the transcendental deduction, and hence that the argument is completed in §20.

However, in §21 Kant says: "Thus in the above proposition a beginning is made of a deduction of the pure concepts of understanding" (B144). He goes on to explain:

In what follows (cf. §26) it will be shown, from the mode in which the empirical intuition is given in sensibility, that its unity is no other than that which the category (according to §20) prescribes to the manifold of a given intuition in general. Only thus, by demonstration of the a priori validity of the categories in respect of all objects of our senses, will the purpose of the deduction be fully attained. (B144-5)

Here an old interpretive question arises: how should we conceive of the argument we find in §26, together with material from §24, which has become known as second step of the B-Deduction?

Erich Adickes (1889: 139-4) and H. G. Paton (1936: v. 1, 501) argue that while what precedes §21 is an *objective* deduction, what we find in §24 and §26 is a *subjective* deduction. This distinction has its source in the Preface to *A*, where Kant says:

This enquiry, which is somewhat deeply grounded, has two sides. The one refers to the objects of pure understanding, and is intended to expound and render intelligible the objective validity of its a priori concepts. It is therefore essential to my purposes. The other seeks to investigate the pure understanding itself, its possibility and the cognitive faculties upon which it rests: and so deals with it in its subjective aspect. Although this latter exposition is of great importance for my chief purpose, it does not form an essential part of it. (Axvi-xvii)

Henrich (1968-9) rejects the Adickes/Paton proposal for the reason that in §21 Kant says that the demonstration of the validity of the categories is completed only in §26, and the passage from the *A*-Preface indicates that this is a task for the objective deduction. In defense of Adickes and Paton, in §20 Kant claims to have shown that the categories apply to the manifold in any given intuition, and says that he will now show that categories apply to any object that presents itself to the senses. In view of the aim of the objective deduction, this move would seem to require only a straightforward application of the result of §20 to any empirical intuition we might have.

Henrich points out that although in the *B*-Deduction Kant sought to avoid the problems of a subjective deduction, this “does not mean that he neglected the demand for an explanation of the possibility of relating the categories to intuitions.” But by Kant’s account, a subjective deduction contains not only an examination of cognitive faculties, but also an investigation of “the possibility of the pure understanding,” which would include an investigation of precisely

how the categories might be related to intuitions and to the objects of experience they represent. This is just the type of inquiry that is to be found in §24 and §26 – and Henrich agrees. So in the last analysis, the views of Henrich, Paton, and Adickes can be reconciled: they can all concur that the second step is an attempt to show how the categories are related to objects of experience – in such a way as to show how the categories might correctly apply to them. Moreover, this common ground provides a reasonable interpretation of the text. For in §26 Kant argues that our representations of space and time, because they themselves contain a manifold, must also be synthesized by the categories. Since all of the objects we experience are given to us in space and time, these objects too will be synthesized by the categories. Thus Kant provides an explanation of how the categories apply to the objects of our experience by way of the manner in which these objects are given to us (Longuenesse 1998: 211-33).

#### A final word

The legacy of the transcendental deduction includes not only the successes of Kant's actual argument, but also a number of influential philosophical strategies: the now-standard tactic of arguing for concepts whose source is in the mind from universal and necessary features of experience; the idea of drawing significant philosophical conclusions from premises about self-consciousness alone; and the notion of a transcendental argument, which from an uncontroversial premise about our thought, knowledge or experience, reasons to a substantive and unobvious necessary condition of this claim.

## References and Further Reading

Adickes, E. (1889). *Immanuel Kants Kritik der reinen Vernunft*. Berlin: Mayer & Müller.

Allison, H. (1983). *Kant's Transcendental Idealism*. New Haven: Yale University Press.

Aquila, R. (1989). *Matter and Mind*. Bloomington: Indiana University Press.

Bennett, J. (1966). *Kant's Analytic*. Cambridge: Cambridge University Press.

Bird, G. (1962/1973). *Kant's Theory of Knowledge*. London: Routledge and Kegan Paul; reprinted, New York: Humanities Press, 1973.

Guyer, P. (1987). *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.

Henrich, D. (1968-9). "The proof-structure of Kant's Transcendental Deduction," *Review of Metaphysics*, 22, 640-59.

Howell, R. (1992). *Kant's Transcendental Deduction*. Dordrecht: Kluwer.

Hume, D. (1739). *A Treatise of Human Nature*. Oxford: Oxford University Press, 1978.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press, 2005.

Kitcher, P. (1990). *Kant's Transcendental Psychology*. Oxford: Oxford University Press.

Longuenesse, B. (1998). *Kant and the Capacity to Judge*. Princeton: Princeton University Press.

Longuenesse, B. (2000). "Kant's categories and the capacity to judge" *Inquiry*, 43, 91-110.

Paton, H. J. (1936). *Kant's Metaphysic of Experience*. London: Allen and Unwin, 2 vols.

Pereboom, D. (1995). "Self-understanding in Kant's Transcendental Deduction". *Synthese* 103: 1-42.

Pereboom, D. (2000). "Assessing Kant's Master Argument." *Kantian Review*, 5, 90-102.

Strawson, P. F. (1966). *The Bounds of Sense*. London: Methuen.

van Cleve, J. (1999). *Problems From Kant*. Oxford: Oxford University Press.

Wolff, R. P. (1963). *Kant's Theory of Mental Activity*. Cambridge: Harvard University Press.