TRANSCENDENTAL ARGUMENTS

The Oxford Handbook of Philosophical Methodology, John Hawthorne, Herman Cappelen and Tamar Szabó Gendler, eds., Oxford: Oxford University Press, 2016, pp. 444-62.

Derk Pereboom, Cornell University

Penultimate Draft

Among Immanuel Kant's (1724-1804) most influential contributions to philosophy is his development of the transcendental argument. In Kant's conception, an argument of this kind begins with a compelling first premise about our thought, experience, knowledge, or practice, and then reasons to a conclusion that is a substantive and unobvious presupposition and necessary condition of the truth of this premise, or as he sometimes puts it, of the possibility of this premise's being true. Transcendental arguments are typically directed against skepticism of some kind. For example, Kant's Transcendental Deduction targets Humean skepticism about the applicability of a priori metaphysical concepts, and his Refutation of Idealism takes aim at skepticism about an external world. A focus on anti-skeptical objectives suggests that this method addresses only a fairly narrow range of philosophical topics. However, many issues in philosophy can be represented as confrontations between skeptical and anti-skeptical points of view. For example, the utilitarian can be represented as a skeptic about various non-consequentialist moral considerations, and the incompatibilist about free will and

determinism as a skeptic about free will given determinism. Yet at the same time it is not essential to transcendental arguments that they have anti-skeptical intent.

The Nature of Transcendental Arguments

An important issue for transcendental arguments concerns the epistemic qualifications of the initial premise. It is sometimes specified that the first premise of a successful transcendental argument must be one of which we – and this includes a targeted skeptic – are or can be certain. But whether this standard must be met depends on the skepticism the argument targets. Janet Broughton (2002) interprets Descartes's *cogito ergo sum* as a transcendental argument against the dream and evil demon skepticism introduces in the First Meditation. Given that the standard Descartes sets for the acceptability of a claim is indubitability, the initial premise must also be indubitable. But not every transcendental argument is advanced in such an epistemically rarified context. For instance, Justin Coates (forthcoming) provides a compelling interpretation of P. F. Strawson's "Freedom and Resentment" (1962) as a transcendental argument against skepticism about moral responsibility. This skeptic is concerned that moral responsibility is incompatible with determinism and that determinism might well be true. Skepticism of this sort does not appeal to the claim that only indubitable propositions are acceptable, and consequently the first premise need not meet this standard.

An alternative and plausible standard is contextual – the first premise must be one the skeptic at issue will accept. It would be valuable in addition if this

premise had a particular sort of resilience for the skeptic, so that once she understands that a necessary condition of the truth of that premise is a claim she doubts or denies, she won't be readily disposed to respond by denying the premise. This point might be made more precise when stated in terms of credences, i.e., degrees of belief in a proposition or beliefs in probabilities that a proposition is true. If the transcendental argument is to succeed against the skeptic, the skeptic's credence in the first premise conditional on its necessitating the falsity of the skeptical claim must be not significntly lower than the skeptic's initial credence in the skeptical claim. True, no actual skeptic is likely to be convinced to reject his skepticism by a transcendental argument, human nature being what it is, and he will surely find fault with either the first premise or the reasoning. For this reason, 'the skeptic' in this characterization must be relevantly idealized.

Some transcendental arguments gain strength by appealing to a first premise that is a supposition the skeptic must accept either because it is a premise of the skeptic's argument for her skeptical conclusion, or since it is transparently entailed or presupposed by such a premise. If such a transcendental argument is successful, the ground of the falsity of the skeptic's claim would turn out to be a premise the skeptic cannot reject while retaining her skepticism. The resilience of such a premise would be especially strong. Any reduction of the skeptic's credence in the premise upon understanding that it necessitates the falsity of the skeptical claim would result in a corresponding weakening of the skeptic's argument, and in particular a reduction in her rational credence in the skeptical conclusion.

The crucial steps in the reasoning featured in transcendental arguments are claims to the effect that a subconclusion or conclusion is a presupposition and a necessary condition of a premise; i.e., that the premise presupposes and necessitates the subconclusion or conclusion. On one proposal, these steps must display logically necessary conditions in particular, and weaker connections are insufficient. But here too it's plausible that the requisite strength of the necessary connection varies with the type of skeptic targeted by the transcendental argument. The necessity might be logical, metaphysical, nomological, or explanatory. If the skeptic doubts metaphysical necessitation but not logical necessitation, the necessary conditions appealed to in the argument must be logical. If the skeptic does not doubt either logical or metaphysical necessitation, then both logical and metaphysical necessary conditions are fair game. But the moral responsibility skeptic, for instance, takes no issue with nomological necessitation, and thus a transcendental argument that takes aim at this position is free to employ such a weaker condition, and arguably Strawson's (1962) version does (Coates, forthcoming). Furthermore, in many philosophical contexts, the relevant sort of skeptic takes no issue with the notion of only possible explanation or best explanation. In such a context, the steps of a transcendental argument need show only that the subconclusion or conclusion is a necessary condition for the premise in the sense that it is the only possible explanation for it, or in the still weaker sense that it is the best explanation for it. As we shall see, the key steps of Kant's Transcendental Deduction invoke such explanatory conditions. This is not a defect in the argument, for the reason that

Humean skepticism about the applicability of metaphysical concepts is not also skeptical of explanatory necessary conditions.

Perhaps the best known contemporary transcendental arguments are *world-directed* in the sense that they aim to secure an anti-skeptical claim about mind-independent reality (Peacocke 1989; Cassam 1999; Stern 2012). But transcendental arguments need not be world-directed in this sense. They might, for example, be ethical in import without aiming to establish a moral realist conclusion, by contrast with one that is, say, constructivist instead. For example, Strawson's transcendental argument against skepticism about moral responsibility leaves a constructivist account of moral responsibility open. Christine Korsgaard's (1998) transcendental argument for the conclusion that we must value ourselves as rational agents from the premise that we make rational choices also does not commit to moral realism. As we shall see, the world-directed transcendental arguments face an important objection, advanced by Barry Stroud (1968), according to which the *existence* of the external feature will not be a necessary condition of the aspect of experience or knowledge invoked by the first premise, for a belief about or representation of the external feature would then also suffice. A world-directed transcendental argument vulnerable to this objection would fall short of its anti-skeptical ambitions.

Let us now inspect a number of specific transcendental arguments, two from Kant and several contemporary examples. We will begin with a substantial discussion of Kant's Transcendental Deduction. It's still the paradigmatic transcendental argument, and due to its ambitiousness and promise, it has been the main inspiration for the ensuing tradition. We then turn to Kant's Refutation of

Idealism, because it inspires the widespread strategy of using transcendental arguments to undermine external-world skepticism. Subsequently we will discuss a number of contemporary arguments, focusing on their problems and prospects.

Kant's Transcendental Arguments

Kant's most famous transcendental arguments are found in the *Critique of Pure Reason* (1781, 1787/1987): the Transcendental Deduction of the Categories, the Second Analogy, and the Refutation of Idealism. There are many others, in, for example, the *Critique of Practical Reason,* the *Critique of the Power of Judgment*, and in the *Opus Posthumum* (Forster 1989). Here I single out the two that are most celebrated: the Transcendental Deduction the Refutation of Idealism. Discussion of the Transcendental Deduction, the most influential of all, and the part of Kant's theoretical philosophy that he believed to be his greatest achievement, illustrates the structure of a transcendental argument, and in particular the epistemic requirements for the first premise and for the necessary conditions such an argument involves. Consideration of the Refutation of Idealism highlights in addition the type of objection Stroud raises against world-directed transcendental arguments.

1. The Transcendental Deduction

In the Transcendental Deduction (1781/1787/1987: A84-130, B116-169) Kant aims to demonstrate against an empiricist that certain priori concepts legitimately apply to objects featured in our experience. A deduction in this context

is an argument intended to justify the use of a concept, one that shows that the concept legitimately applies to real things. For Kant a concept is a priori just in case its source is in the mind of the subject and in not sensory experience (A80/B106; Strawson 1966: 86). The particular a priori concepts whose applicability to objects of experience Kant aims to vindicate are given in his Table of Categories (A80/B106); they are *unity*, *plurality*, and *totality* (the Categories of Quantity); *reality*, *negation*, and *limitation* (the Categories of Quality); *inherence and subsistence*, *causality and dependence*, and *community* (the Categories of Relation), and *possibility-impossibility*, *existence-non-existence*, *necessity-contingency* (the Categories of Modality).

David Hume denies that a deduction can be provided for a number of metaphysical concepts – *ideas*, in his terminology – including those of personal identity, of identity over time more generally, of the self as a subject distinct from its perceptions, and of causal power or force (1739, 1748). In his view, a concept can be validated only by finding a sensory experience, that is, an impression, in particular one that is the 'original' of that idea, which must resemble the idea. However, because any attempt to locate, for example, an impression corresponding to the idea of causal power turns outs to be unsuccessful, he concludes that this idea does not apply in our experience (1748: §7). In Kant's terminology, Hume is testing to see whether there is an *empirical* deduction of the concept of causal power (A85/B117), and from the failure of the attempt to produce one, he concludes that this concept lacks *objective validity*, that is, it does not apply to the objects of our experience.

Hume's view about the impossibility of a deduction of a priori metaphysical concepts is Kant's target in the Transcendental Deduction. Kant agrees with Hume, however, that no empirical deduction is forthcoming for such concepts. Instead, he aims to produce a different sort of justification for their, one that is transcendental rather than empirical. A transcendental deduction begins with a premise about any possible human experience, a premise to which the participants in the debate will at least initially agree, and then contends that a presupposition of and necessary condition for the truth of that premise (or for the possible truth of that premise) is the applicability of the a priori concepts at issue, the categories. Kant's Transcendental Deduction features a number of subsidiary transcendental arguments. Each begins with a premise either about the self-attributability of mental items, *apperception*, or else a premise that affirms the necessity and universality of a feature of our experience of objects. Kant's strategy is to establish a specific theory of mental processing, *synthesis*, by arguing that its truth is a presupposition of and a necessary condition for the truth of such a premise, and then to show that the categories have an essential role in this mental processing. On a metaphysical idealist interpretation of his position, the objects of experience are produced by this mental processing, and it is due to the role that the categories have in this production process that they legitimately apply to these objects.

For Kant the most significant rival theory of mental processing is that of his target, Hume. Hume agrees that a theory of experience will feature an account of the processing of mental items, but he denies that such an account should involve a priori concepts, and *a fortiori* that it issues in their applicability to experience. In 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). Association itself is the process by which these perceptions are related and ordered (1748, §3). An important characteristic of association is that it allows no resources other than the perceptions themselves. How perceptions are ordered is accounted for solely by facts about the perceptions themselves. Significantly, a subject not constituted solely of perceptions has no role in Hume's theory; the Humean subject is just a collection of perceptions (1739: I, IV, vi). These last two features make Hume's associationism a particularly economical theory, which results in a prima facie advantage over Kant's more complex view. Kant contends, however, that associationism cannot accommodate the compelling premises of the Transcendental Deduction, and this makes the case for synthesis by a priori concepts.

For Kant synthesis is "the act of putting different representations together, and grasping what is manifold in them in one cognition" (A77/B103); it is a process that "gathers the elements for cognition, and unites them to form a certain content" (A78/B103). Synthesis takes multiple representations – in Kant's term, a 'manifold' – and connects them with one another to produce a single further representation with cognitive content (Kitcher 1990, 2011). This process employs concepts as ways of ordering representations. A claim crucial to the Transcendental Deduction is that it is the categories by means of which manifolds of representations are synthesized. Because the understanding of the subject is the source of the categories, and since it

is also the faculty that generates synthesis, the subject plays an essential role in mental processing. It is important for Kant's theory that this subject is distinct from its states, and this is a further respect in which it differs from Hume's.

Here we will focus on the core part (§§16-20) of the Transcendental Deduction in the second edition of the *Critique of Pure Reason* (1787/1987), the B-Deduction. On my reading, in §§16-20 of the B-Deduction Kant employs a two-pronged strategy for defeating associationism and establishing synthesis, each of which is a transcendental argument. The first, contained in §16, is designed to demonstrate that association cannot account for an aspect of consciousness of the self that Kant refers to as the consciousness of its unity, and that such an account requires synthesis instead or in addition. This kind of transcendental argument he calls an *argument from above*, signifying that it begins with a premise about self-consciousness. Correlatively, §§17-20 features an *argument from below*, by which Kant intends to establish that synthesis by the categories is needed as a necessary condition for certain features of how we represent objects (the above/below terminology derives from A119).

The argument from above in §16 can be divided into two stages. The aim of the first is to establish the various components of *the principle of the necessary unity of apperception*. The second stage aims to show that synthesis is a necessary condition for the aforementioned aspect of self-consciousness, which this principle highlights. *Apperception* is the apprehension of a mental state, a *representation* (*Vorstellung*) in Kant's terminology, as one's own; one might characterize it as the self-ascription or self-attribution of a mental state (Strawson 1966: 93-4). In Kant's

conception, apperception of my representations has *necessary* unity in the sense that 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, emphasis mine). 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).

Kant initiates the first stage of the argument in §16 by claiming:

> 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 interpretation, the sense in which a representation would be impossible or nothing to me if it could not accompanied by the 'I think' is simply that I could not then become conscious of it (Guyer 1987: 139-44). It is credible that for any representation of which I am conscious, I can attribute it to myself as subject, assuming my mental faculties are in working order, and if no controversial account of the nature of the subject is presupposed. But the claim that I can become conscious of *each* of my representations, and that it is therefore possible for me attribute each of them to myself as their subject, is likely to be false. Plausibly, some of my represen-

tations are so thoroughly subconscious that I cannot attribute them to myself, while they are nevertheless mine due to the causal relations they bear to other representations and to actions that are paradigmatically mine. Fortunately, however, the premise that each of my representations is such that I can attribute it to myself is not crucial for the argument from above. Rather, the one Kant ultimately singles out:

> I am conscious of the identity of myself as the subject of different self-attributions of mental states

is significantly less committed, and also highly credible and resilient.

The argument from above crucially turns on the proposal 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 inadequacy Kant claims for "empirical consciousness," that is, for consciousness according to Humean psychological theory, is that "it is in itself dispersed and without relation to the identity of the subject" (B133). One idea expressed here is that Hume's theory does not have the resources needed to account for one's ability to attribute representations to one's self conceived as a subject that is both conscious of them and the same subject for each act of self-attribution. Humean theory can accommodate the view that apperceptive consciousness consists in perceptions that are intrinsically self-conscious, or else perceptions of perceptions. But intrinsically self-conscious perceptions would be distinct from one another, as would perceptions of perceptions; and thus they too would be "dispersed" (B133),

and have no common subject. Hume might propose to explain one's sense of the identity of the conscious subject of different self-attributions by the intrinsically self-conscious perceptions or the perceptions of perceptions being components of a single causally coherent bundle. Still, this bundle would not itself be conscious of perceptions. Consciousness of perceptions would instead be an intrinsic feature of individual self-conscious perceptions or a feature of individual perceptions of a perception. On Kant's proposal, by contrast, accounting for one's sense of the identity of the conscious subject of different self-attributions requires that this subject be distinct from its representations.

The second stage of the argument from above of §16 involves a further implication of the claim that "empirical consciousness, which accompanies different representations, is dispersed and without relation to the identity of the subject" (B133), i.e., that Hume's theory lacks the resources to account for my *representation-relation* to the identity of the subject. His view cannot explain how I can "represent to myself the *identity of the consciousness in [i.e. throughout] these representations*" (B133). We might imagine several kinds of explanation for my representation of this identity. One candidate is that inner sense accounts for it. On this suggestion, the way I represent the sameness of the subject would be akin to how I commonly represent the identity over time of ordinary objects – by sensory apprehension of intrinsic properties, and noting that these intrinsic properties remain the same, or similar enough, over time. However, Kant and Hume concur that this is not the way I could represent the identity of the apperceiving subject, since they agree that by inner sense I cannot represent any intrinsic properties of

such a subject. A second kind of explanation, which Kant endorses, is that I have an indirect way of representing this identity. This representation must instead depend on my apprehending a feature of my representations (Allison 1983: 142-4; Guyer 1977: 267, 1987: 133-9). The appropriate feature is a type of unity or ordering. Kant's idea is that if the representations I can attribute to myself possess a unity of the right kind, and if I am conscious of this unity, then I will be able to represent the apperceiving subject of any one of them as identical with that of any other. 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).

This consciousness is plausibly interpreted as conscious awareness not of the act or process of synthesis itself, but rather of the unity that is its outcome (Strawson 1966: 94-6; Dicker 2004: 133-4)). What sort of unity must I consciously recognize among my representations that would account for my representation of this identity? A credible proposal is that the unity consists in certain intimate ways in which representations in a single subject are typically related. Arguably, the essential feature of this unity is that a subject's representations be inferentially and causally integrated to a high degree, and in this respect they are unified in a way in which representations possessed by distinct subjects are not. When mental states fail to exhibit inferential and causal integration, as in the case of multiple-personality disorder, we have a tendency to posit distinct subjects, and we do not when such integration is present.

In Kant's view, the candidates for accounting for this kind of unity – or, less ambitiously, for my ability to recognize this sort of unity – are association and synthesis. At this point in the argument he seems to suppose that because Hume's psychological theory has already been ruled out, synthesis is the only remaining option. So for me to represent the identity of the subject of different self-attributions, I must generate or at least recognize the right sort of unity among these representations, and synthesis must be recruited to account for this unity. Thus Kant contends 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 because for synthesis to be a priori is, at least 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 the unity in question.

Here is an austere representation of the structure of the argument so far:

(1) I am conscious of the identity of myself as the subject of different self-attributions of mental states. (premise)

(2) I am not directly conscious of the identity of this subject of different self-attributions (premise).

(3) If (1) and (2) are true, then this consciousness of identity is accounted for indirectly by my consciousness of a particular kind of unity of my mental states. (premise)

(4) This consciousness of identity is accounted for indirectly by my consciousness of the particular kind of unity of my mental states (1, 2, 3)

(5) If (4) is true, then my mental states indeed have the particular kind of unity (premise).

(6) This particular kind of unity of my mental states cannot be accounted for by association (5, premise).

(7) If (6) is true, then this unity of my mental states is accounted for by synthesis by a priori concepts. (premise)

(8) This unity of my mental states is accounted for by synthesis by a priori concepts (6 and 7)

The structure of this part of the deduction as a transcendental argument is clear. Premise 1 is intended as a claim the metaphysical-concept skeptic will accept. The crucial necessary conditions, expressed by (3) and (7), are at root necessary conditions of only possible explanation. It's not especially plausible, however, that Kant has ruled out all of the competing explanations. However, the argument might well still have force against the skeptic's position if necessary conditions these premises express are those of best explanation. The skepticism targeted by the Transcendental Deduction does not question conditions of this sort.

Paul Guyer forcefully argues that establishing synthesis by a priori concepts would require ruling out the alternative view that empirical information and concepts derived from experience are sufficient to account for the recognition of the unity at issue (Guyer 1987:146-7). In particular, it remains open, given what Kant has shown, that this recognition requires only awareness of information derived from inner sense or introspective experience. At this juncture in the argument from

above Kant does not take on the task of ruling out such a rival empiricist proposal, but he would need to do so to establish the need for synthesis by a priori concepts.

In the next phase of the Transcendental Deduction (§§17-20), an argument from below, this is exactly the task Kant takes on. In §18 he draws our attention to certain features of our representations of objects that, in his view, will serve to defeat associationism, the empiricist's rival proposal, and establish a priori synthesis (Ameriks 1978; Pereboom 1995, 2009; Kitcher 2011). For Kant, a key characteristic of our 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). Kant contends that our objectively valid representations must in a sense be necessary and universal. However, the empirical unity of consciousness, which involves an ordering of representations produced by association, can only be non-universal, contingent, and hence merely subjectively valid, by contrast with the transcendental unity of apperception, which is or involves an ordering that is universal and necessary, and is therefore objectively valid. In Kant's conception, it is the fact that the transcendental unity of apperception is generated by synthesis by a priori concepts that allows it to yield an ordering that is universal, necessary, and objectively valid.

To illustrate and support these claims, Kant here invokes the example of the ordering of phenomena in time that has the key role in the discussion of the Second Analogy (cf. Guyer 1987: 87-90; Dicker 2004: 137-44). There he argues that in our representations, considered independently of their content, are always successive.

For example, when I view the front, sides, and back of a house when walking around it, and when I watch a boat float downstream, my representations of the individual parts and states occur successively. The objective phenomena represented by these successive representations, however, can be represented as either successive or as simultaneous – I represent the positions of the boat as successive, but the parts of the house as simultaneous. So despite the representations in each of these sequences being s*ubjectively successive*, I represent the parts of the house as *objectively simultaneous*, and the positions of the boat as *objectively successive*. How might we account for this difference in objectivity despite the similarity in subjectivity (Melnick 1973: 89)?

The important clue for answering this question is that these representations of objective simultaneity and succession are universal and necessary. On Kant's proposal, it is the universality and necessity of our representing the parts of the house as simultaneous that accounts for our representing them as objectively simultaneous, and the universality and necessity of our representing the positions of the boat as successive that accounts for our representing them as objectively successive. Association is inadequate for accounting for this objectivity because it is incapable of yielding such universality and necessity, a defect not shared by synthesis.

A first approximation of the import of 'universal' in the house example is:

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

The addition of necessity has the following result:

(U-N, first pass) Necessarily, any human experience of the parts of the house is an experience of these parts as objectively simultaneous.

Hume would resist this claim if the necessity were specified as ranging over all possible circumstances, since his theory would allow for the possibility of a deviant ordering in unusual empirical conditions. But (U-N, first pass) can be reformulated more precisely as

(U-N) 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 proposal is that given only the resources of association, the truth of (U-N) cannot be explained. His reason is "whether I can become empirically conscious of the manifold as simultaneous or as successive depends on circumstances or empirical conditions," and so "the empirical unity of consciousness, through association of representations, itself concerns an appearance, and is wholly contingent" (B139-40). Association does not yield an explanation the truth of (U-N), for given only the resources of association, the parts of the house will not necessarily or universally be represented as objectively simultaneous even supposing only normal empirical conditions. Kant has us 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). In conversations people make different associ-

ations in similar circumstances. Kant's point is that if the paradigms for association fail to exhibit the sort of necessity and universality at issue, then the proposal that association can yield such an ordering of representations – wherever we find it – is excluded.

Here we should see Kant as advancing his claim for the applicability of the categories by ruling out association as an explanation for (U-N). The structure of the resulting transcendental argument can be represented as follows:

9. We have representations of objects, i.e., of objectively valid phenomena. (premise)

10. All of our representations of objects are of universal and necessary features of experience. (premise)

11. Necessary and universal features of experience cannot be explained by association. (premise, from reflection on the nature of association)

12. If (10) and (11) are true, all of our representations of objects require a faculty for ordering mental states distinct from association. (premise)

13. All of our representations of objects require a faculty for ordering mental states distinct from association. (11, 12)

14. If (13) is true, all of our representations of objects require a faculty for synthesis by a priori concepts. (premise)

15. All of our representations of objects require a faculty for synthesis by a priori concepts – that is, the same faculty that accounts for my consciousness of the identity of myself as the subject of different self-attributions of mental states. (8, 13, 14)

To this we can add the final moves, which are explained in the subsequent sections of the B-Deduction:

16. Insofar as our representations of objects require a faculty for synthesis by a priori concepts, certain a priori concepts -- the categories -- legitimately apply to these objects. (premise)

C. We have representations of objects, and they are all such that the categories legitimately apply to these objects. (9, 15, 16)

The key necessary conditions expressed by (12) and (14), like those of the argument of the first stage, are conditions of only possible explanation. Here again, if these conditions turned out to involve best explanation instead, the argument would retain its force against the targeted skeptic.

In summary, the challenge Kant issues in this second stage of the Transcendental Deduction is to explain why, under normal conditions, ordering of representations in experience is universal and necessary. Part of the only explanation, he believes, is that we must have a faculty for ordering the representations. Hume might agree with this conclusion, supposing a sufficiently thin conception of 'faculty' on which it might consist solely of sensory items and associative tendencies among them. Kant argues that the Humean proposal cannot account for the truth of propositions such as (U-N), for the very paradigms of association, such as word association, and the association of topics in a conversation, do not exhibit the requisite universality and necessity. The alternative that can account for the truth of propositions such as (U-N) involves affirming the conclusion (C), that we have a faculty for synthesis by a priori concepts, which is the

same faculty that was shown earlier to be required for my consciousness of the identity of myself as the subject of different self-attributions of mental states.

Briefly, here is the rest of the story. In §19, Kant argues 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 proposes, is objectively rather than subjectively valid, and hence exhibits the type 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 it would not explain how we make distinctions between objective valid phenomena (i.e., objects) and the subjective states they induce. In §20 Kant ties this notion of judgment to the twelve forms of judgment presented in the Metaphysical Deduction (A70/B95), and then connects these forms of judgment to the twelve categories (A76-83/B102-9). The challenge has often been raised that the links Kant specifies between synthesis and judgment, judgment and the forms of judgment, the forms of judgment and the categories are not sufficiently supported (Guyer 1987: 94-102). Béatrice Longuenesse (1998), in her state-of-the-art interpretation of the Metaphysical Deduction, takes up this challenge with impressive results.

How resilient will the first premises for the two component transcendental arguments be? They, in effect, are:

(1) I am conscious of the identity of myself as subject of different self-attributions of mental states.

(9-10) We have representations of objects, all of which are of universal and necessary features of experience.

The Humean skeptic might try to reject (1), but denying this consciousness of subject-identity is a radical and unattractive move, even for Hume. Regarding (9-10), Hume would not disavow the necessity and universality this premise invokes, on a proper understanding of the kind of necessity at issue. He maintains that it is in some sense impossible, given an experience of constant conjunction, that the mind not be carried from an impression of the first conjunct to an idea 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)

Hume himself contends that given certain specific empirical circumstances, a particular type of ordering of perceptions in a sense necessarily (and universally) comes about, and this is just the type of claim Kant is making in (9).

The Transcendental Deduction has been highly influential as the paradigm of the method of transcendental argument. More specifically, it pioneers the alluring idea of using this method to draw significant anti-skeptical conclusions from premises about self-consciousness alone, and the now-standard tactic of arguing for

concepts whose source is in the mind from universal and necessary features of experience.

2. The Refutation of Idealism

Kant's quarry in the Refutation of Idealism is Cartesian skepticism about the external world (B274-279; Bxxxix-Bxli). His intent is to refute what he calls *problematic idealism*, according to which the existence of objects outside of me in space in space is "doubtful and indemonstrable" (B274). Kant's strategy is to show that the existence of such objects is a necessary condition of my awareness that my representations have a specific temporal order. At the present time I am aware of the specific temporal order of many of my past experiences. This awareness is produced by memory, but what is it about what I remember that allows me to determine the temporal order of these experiences? There must be something by reference to which I can correlate the remembered experiences that allows me to do this. However, first, I have no conscious states that can play this role. In addition, this reference cannot be time itself, for "time by itself is not perceived;" As Guyer observes, it is not as if the content of memories of individual events are evidently indexed to particular times, the way in which sportscasts and videotapes often are (Guyer 1987). Kant argues that the only other candidate for this role is something outside of me in space, something that is permanent (cf. First Analogy, B224-5).

Kant's proposal is perhaps made plausible by how we often actually determine the times at which our experiences occur. We use the observations of sun's positions, or of the changing shadow on a sundial, or of a clock that indicates

time by means of the period of a pendulum. Kant's argument can be viewed as exploiting this fact, together with the observation that there is no similar periodic process in our conscious experience considered independently of any spatial objects it might represent, and that we lack any awareness of time by itself, to show we must perceive objects in space. For then it would be only by reference to such objects that we can determine the objective temporal order of our experiences.

George Dicker sets out a compelling representation of Kant's argument (Dicker 2004, 2008):

> (1) I am conscious of my own existence in time; that is, I am aware, and can be aware, that I have experiences that occur in a specific temporal order. (premise)
>
> (2) I can be aware of having experiences that occur a specific temporal order only if I perceive something permanent by reference to which I can determine their temporal order. (premise)
>
> (3) No conscious state of my own can serve as the permanent entity by reference to which I can determine the temporal order of my experiences. (premise)
>
> (4) Time itself cannot serve as this permanent entity by reference to which I can determine the temporal order of my experiences. (premise)
>
> (5) If (2), (3), and (4), are true, then I can be aware of having experiences that occur in a specific temporal order only if I perceive persisting objects in space outside me by reference to which I can determine the temporal order of my experiences. (premise)

(C) I perceive persisting objects in space outside me by reference to which I can determine the temporal order of my experiences. (1-5)

Two of the most pressing objections that have been raised against the Refutation are that the skeptic would resist the first premise, and that the argument is vulnerable to an instance of Stroud's objection. So first, a skeptic could reject the initial premise on the ground of a general skepticism about memory (Allison 1983: 306-7). Bertrand Russell, for example, proposes that for all I know I was born five minutes ago (Russell 1912). On this skeptical hypothesis, I would be mistaken in my belief that I had experiences A, B, and C which occurred more than five minutes ago, first A, then B, and lastly C. It's credible that a skeptic who claimed that we lack adequate justification for a belief that external objects exist would also be disposed to contend that I lack justification for my belief that I had experiences that occurred in the past in that particular temporal order. Accordingly, Kant is not clearly justified in supposing that Premise (1) provides leverage against an external-world skeptic (cf. Dicker 2008, Chignell 2009).

Second, consider the proposal that states of the self are as well-suited as objects in space to function as a reference whereby I can accurately discern the temporal order of my past experiences. Imagine I had available as such a reference solely the mere appearance of a digital clock in one corner of my field of consciousness. This would not clearly be less effective than an actual clock in space (cf., van Cleve, reported in Dicker 2004: 207; Dicker 2004: 207). This objection is an instance of the type of concern Stroud raises against world-directed transcendental arguments, viz., that mere representation of some feature, by contrast with the

existence of the external feature the skeptic targets, is all that can be established as a necessary condition of the first premise. To this one might reply, with Dicker, that there are in fact no states of the self that can serve as such a reference. However, and this is the deeper worry, according to Berkeley's idealist view in which the *esse* (to be) of objects in space is their *percipi* (to be perceived), any spatial object would amount to no more than mental states of the subject. But Berkeleyan spatial perceptions would seem to be as effective a reference by which to ascertain the temporal order of my past experiences as perceptions of objects distinct from my mental states (cf. Allison 1983: 300-301; Chignell 2009).

The Refutation of Idealism is an especially ambitious transcendental argument, and it has inspired many others for a similar conclusion. But critics largely agree that the Refutation itself falls to instances of certain standard forms of objection to transcendental arguments: that the skeptic need not commit to the first premise, and that the argument can establish at most a conclusion about our representations or beliefs and not about mind-independent reality.

Contemporary Kantian Transcendental Arguments

1. Practical transcendental arguments.

Transcendental arguments against various sorts of skepticism were developed with vigor in the mid-twentieth century, and it was P. F. Strawson who led this effort. One of Strawson's most influential works is his essay on moral responsibility, "Freedom and Resentment" (Strawson 1962). The reasoning in this article has not traditionally been interpreted as a transcendental argument, but

recently Justin Coates (forthcoming) has made a strong case for such a reading. In Coates's account, the argument begins with the premise to which the moral responsibility skeptic would agree, that meaningful adult interpersonal relationships are possible for us. It continues by pointing out that relationships of this sort require that the participants show each other good will and respect, and that they be justified in expecting this of one another. Expectations for good will and respect in turn require susceptibility to the reactive attitudes, such as moral resentment, indignation, and gratitude, and in particular, justified expectations for good will and respect presuppose that the participants are apt recipients of these reactive attitudes. But to be an apt target of the reactive attitudes is just what it is to be a morally responsible agent. Consequently, that we are morally responsible agents is a necessary condition of the possibility for us of meaningful adult interpersonal relationships.

Note that not all the connections among the steps of the argument are plausibly instances of appeals to logical or even metaphysical necessary conditions. True, some are: if being an apt target of the reactive attitudes is what it is to be a morally responsible agent, the necessary connection invoked would be conceptual or metaphysical. But if expectations for good will and respect do require susceptibility to the reactive attitudes, this would be plausibly a case of nomological necessitation, where the relevant laws are psychological. But given the sort of skepticism targeted, nomological necessitation is not too weak a connection; it is not called into question by the arguments of the moral responsibility skeptic.

Critics have in effect taken issue with a number of steps of this argument, for example that expectations for good will and respect require susceptibility to the reactive attitudes, and that justified expectations for good will and respect presuppose that the participants are apt recipients of the reactive attitudes. Perhaps human relationships do not require susceptibility to moral resentment and indignation, but only to the nonreactive attitudes of moral concern, disappointment and sorrow (Pereboom 2001). Another avenue of criticism involves separating moral responsibility from being an apt target of the reactive attitudes. It may be that a forward-looking, that is, what Strawson calls an 'optimistic' notion of responsibility, is all that's required for good relationships, and it is not characterized by being an apt target of the reactive attitudes. But these criticisms are controversial, and Strawson's argument is widely accepted and acclaimed.

Another prominent transcendental argument in the practical sphere is the sort Korsgaard's (1996) develops for claim that we must value ourselves as rational agents. Here is Robert Stern's (2012) representation of one such argument. It begins with a premise about rational choice, and crucially features the notion of one's practical identity, the distinctive nature of oneself as an agent, which may include, for example, being a Harvard philosophy professor and an American citizen:

1. To rationally choose to do X, you must take it that doing X is the rational thing to do.

2. Since there is no reason in itself to do X, you can take it that X is the rational thing to do only if you regard your practical identity as making X the rational thing to do.

3. You cannot regard your practical identity as making doing X the rational thing to do unless you can see some value in that practical identity.

4. You cannot see any value in any particular practical identity as such, but can regard it as valuable only because of the contribution it makes to giving you reasons and values by which to live.

5. You cannot see having a practical identity as valuable in this way unless you think your having a life containing reasons and values is important.

6. You cannot regard it as important that your life contain reasons and values unless you regard your leading a rationally structured life as valuable.

7. You cannot regard your leading a rationally structured life as valuable unless you value yourself qua rational agent.

C. Therefore, you must value yourself qua rational agent, if you are to make any rational choice.

Stern (2012) explains this argument as follows. To act is to do or choose something for a reason. But one has reasons to act only because of one's practical identity; one does not have reasons to act independently of that identity. However, a practical identity can yield such a reason only if one regards that that identity as valuable. Merely being a father gives one no reason to care for one's children; rather, valuing one's fatherhood has this force. But one cannot regard a particular practical identity as valuable in itself – Korsgaard argues that this sort of realism about value is implausible. The only remaining explanation is that one regards it as valuable because of the contribution it makes to providing one with reasons and values by which to live. But then one must believe that it matters that one's life has the sort of

rational structure that having such identities provides. However, to see that as mattering, one must regard leading a rationally structured life as valuable. Then, in conclusion, to regard leading such a life as valuable, one must see one's rational nature as valuable.

Various steps in this reasoning are again controversial, but this argument and others in the same family have attracted much attention, and it is a fine illustration of the potential the methodology of transcendental argument.

2. Transcendental arguments against external world skepticism.

The second half of the twentieth century featured a revival of transcendental arguments against external world skepticism inspired by the example of Kant's Refutation of Idealism. Perhaps the most prominent example is Strawson's main argument in *The Bounds of Sense* (1966), although Hilary Putnam's (1981) much-discussed argument from the causal theory of reference against this sort of skepticism has also been interpreted as a transcendental argument (Stern 2012). In *The Bounds of Sense* Strawson sets out a number of transcendental arguments inspired by Kant's Transcendental Deduction and his Refutation of Idealism. The one that is best known and most influential (1966: 97-104) is modeled on the Transcendental Deduction, but intentionally without invoking the controversial and arguably dated features of Kant's transcendental psychology. His target is a skeptic who claims that our experience consists just of sense-data, and thus does not feature objects "conceived of us distinct from any particular states of awareness of them."

One might think of the skeptical target as a Berkeleyan account according to which the *esse* of spatial objects of experience is to be perceived (1966: 98).

The essential structure of Strawson's transcendental argument is as follows. It begins with the premise that every (human) experience is such that it is possible for its subject to become aware of it and ascribe it to herself. It is a necessary condition of the truth of this premise that in every experience the subject is capable of distinguishing a recognitional component not wholly absorbed by, and thus distinct from, the item recognized (1966: 100). To be capable of distinguishing these components it is necessary that the subject conceptualize her experiences in such a way so as to contain the basis for a subjective component – how the experienced item seems to the subject – distinct from an objective component – how the item actually is. Strawson argues that "collectively," this comes to "the distinction between the subjective order and arrangement of a series of such experiences on the one hand, and the objective order and arrangement of the items of which they are the experiences on the other" (1966: 101). Conceptualizing experience as involving an objective order and arrangement of items amounts to making objectively valid judgments about it, which, in turn, requires the conclusion that experience must consist of a rule-governed connectedness of representations (1966: 98). Summarizing, from a premise about self-consciousness we can conclude, as a necessary condition, that the subject conceptualizes her experience so as to feature a distinction between "the subjective route of his experience and the objective world through which it is a route," where the experience of the objective world consists in a rule-governed order of representations (1966: 105).

As noted earlier, Barry Stroud, in his 1968 article "Transcendental Arguments," issued a telling objection the enterprise of defeating the external-world skeptic by transcendental arguments of this sort. These arguments reason from some aspect of experience or knowledge to the claim that the contested feature of the external world in fact exists. In each case the existence of the external feature will not be a necessary condition of the aspect of experience or knowledge featured in the initial premise, because a belief about the external feature would always suffice. Although the claim about existence of the aspect of the external world could be secured if certain kinds of verificationism or idealism were presupposed, these views are highly controversial. Moreover, one could make as much of an inroad against the skeptic armed with the verificationism or idealism alone, without adducing the transcendental argument at all (Brueckner 1983, 1984).

Although Strawson's transcendental argument in *The Bounds of Sense* is not a specific target of Stroud's (1968), Anthony Brueckner (1983: 557-8) points out that it is susceptible to the line of criticism that Stroud develops. For Strawson's argument, despite its objective, can only conclude that experience must be conceptualized in a certain way, that is, such as to allow the subject to make the distinction between an objective world and her subjective path through it. This is not a conclusion about how a mind-independent world must be, but only about how it must be thought.

More recent development of world-directed transcendental arguments reflects chastened expectations about what they might establish. One more modest sort of transcendental argument begins with a premise about experience or

knowledge that is acceptable to the skeptic in question, and then proceeds not to the existence of some aspect of the external world, but in accord with Stroud's criticism, to a belief in the existence of some aspect of the external world. Stroud himself advocates a strategy of this sort (Stroud 1994, 1998), as does Stern (1998b). The kind Stern proposes begins with the premise that we think of the world as being independent of us, and it concludes, as a necessary condition of this premise, that we must think of it as containing enduring particulars. Such an argument does not claim that it is a necessary condition of this premise that there must exist such particulars. It contends only for "a connection solely within our thought: if we think in certain ways, we must think in certain other ways" (Stern 1998b: 165). A belief or thought to which one reasons in this way would, in Stroud's assessment, have a certain *indispensibility*, "because no belief that must be present in any conception or any set of beliefs about an independent world could be abandoned consistently with our conception of the world at all," and it would be *invulnerabl*e "in the special sense that it could not be found to be false consistently with its being found to be held by people" (Stern 1998b: 166; Brueckner 1996).

Stern advances a conception of this modest sort of transcendental argument on which it targets a skeptic who questions whether certain beliefs cohere with others in one's set, by contrast with a skeptic who questions whether certain beliefs are true (Stern 1998b). A modest transcendental argument of this sort would aim to show that a belief whose coherence with the other beliefs is challenged so coheres after all. The requisite coherence might be demonstrated by showing that the belief in question is actually a necessary condition of a belief that is indispensable (in

some coherentist sense) to one's set. Mark Sacks (1998) objects that if at the same time one admits that the belief might not be true, one's sense that one is justified in holding the belief will be undermined. This worry seems serious. Sacks contends that it arises because of a tension between the coherentist theory of justification and the realist correspondence theory of truth that the external world skeptic presupposes. He points out that one might respond by accepting a coherence theory of truth as well, but this would be to adopt a version of idealism. Moreover, even if one accepted a coherence theory of truth, one would still have to admit that for specific instances of a belief one might be mistaken, even if one did think that one was justified in holding that belief on grounds of coherence.

Final Words

The legacy Kant's Transcendental Deduction and Refutation of Idealism is the notion of a transcendental argument, which from an uncontroversial premise about our thought, knowledge, or experience reasons to a substantive and unobvious presupposition and necessary condition of this premise, often an anti-skeptical conclusion. Much of the effort spent devising transcendental arguments in the second half of the twentieth century focused on refuting skepticism about the external world, and the prospects for this project do not seem especially bright. But transcendental arguments can be recruited for other purposes, as indicated by Strawson's argument concerning moral responsibility and Korsgaard's argument about valuing oneself as a rational agent. It's credible that the reasons for pessimism about their significance for refuting external world skepticism will not transfer to

such other uses, and that therefore transcendental argument remains a promising philosophical methodology.

Bibliography

Allison, H. (1983). *Kant's Transcendental Idealism*, New Haven: Yale University Press; (2004), second edition.

Ameriks, K. (1978) "Kant's Transcendental Deduction as a Regressive Argument," Kant-Studien 69, pp. 273-87.

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

Berkeley, G. (1713). *Three Dialogues between Hylas and Philonous*, Robert Adams, ed., Indianapolis, Hackett, 1979.

Broughton, J. (2002) *Descartes's Method of Doubt*. Princeton: Princeton University Press.

Brueckner, A. (1983). "Transcendental Arguments I," *Noûs* 17, pp. 551-575.

Brueckner, A. (1984). "Transcendental Arguments II," *Noûs* 18, pp. 197-225.

Brueckner, A. (1996). "Modest Transcendental Arguments," *Philosophical Perspectives* 10, pp. 265-80.

Cassam, Q. (1999). *Self and World*. Oxford: Oxford University Press.

Chignell, A. (2009). "Causal Refutations of Idealism," *Philosophical Quarterly*, forthcoming.

Coates, J. (forthcoming). "Responsibility without (Panicky) Metaphysics."

Dicker, G. (2004). *Kant's Theory of Knowledge*. New York: Oxford University Press.

Dicker, G. (2008). "Kant's Refutation of Idealism," *Noûs* 42, pp. 80-108.

Forster, E., ed. (1989). *Kant's Transcendental Deductions*, Stanford: Stanford University Press.

Guyer, P. (1977). "Review of W. H. Walsh, *Kant and the Criticism of Metaphysics*," *Philosophical Review* 86.

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

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.

Kant, I. *Gesammelte Schriften,* ed. *Koniglichen Preussischen Academie der Wissenschaften*, 29 Vols. Berlin: Walter de Gruyter et al., 1902-

Kant, I. (1781/1787/1987) *Critique of Pure Reason* (trans. P. Guyer and A. Wood). Cambridge and New York: Cambridge University Press, 1997. (References are in the standard pagination of the 1st (A) and 2nd (B) editions. A reference to only one edition indicates that the passage appears only in that edition.)

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

Kitcher, P. (2011). *Kant's Thinker*. New York: Oxford University Press.

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.

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

Melnick, A. (1973). *Kant's Analogies of Experience*, Chicago: University of Chicago Press.

Peacocke, C. (1989). *Transcendental arguments in the theory of content*. Oxford: Oxford University Press.

Pereboom, D. (1990). "Kant on Justification in Transcendental Philosophy," *Synthèse* 85, 1990, pp. 25-54.

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

Pereboom, D. (2001). *Living without Free Will*. Cambridge: Cambridge University Press.

Pereboom, D. (2009). "Kant's Transcendental Arguments", *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/kant-transcendental/>.

Putnam, H. (1981). *Reason, Truth, and History*, Cambridge: Cambridge University Press,

Russell, B. (1912). *The Problems of Philosophy*. London, Williams and Norgate; New York, Henry Holt and Company.

Sacks, M. (1998). “Transcendental Arguments and the Inference to Reality,” in Stern (1998a).

Smith, J. and P. Sullivan (2011), eds. *Transcendental Philosophy and Naturalism*, Oxford: Oxford University Press.

Stern, R. (1998a), ed. *Transcendental Arguments*, Oxford: Oxford University Press, 1998.

Stern, R. (1998b). “On Kant’s Response to Hume: The Second Analogy as Transcendental Argument,” in Stern (1998a).

Stern, R. (2012) "Transcendental Arguments", *The Stanford Encyclopedia of Philosophy (Fall 2012 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2012/entries/transcendental-arguments/>.

Strawson, Peter F. (1962). "Freedom and Resentment," *Proceedings of the British Academy* 48, pp. 1-25.

Strawson, P. F. (1966). *The Bounds of Sense: An Essay on Kant’s Critique of Pure Reason.* London: Methuen.

Stroud, B. (1968). "Transcendental Arguments," *Journal of Philosophy* 65, pp. 241-56.

Stroud, B. (1994). "Kantian Argument, Conceptual Capacities, and Invulnerability," in Paolo Parrini, ed., Kant and Contemporary Epistemology, Dordrecht, Kluwer, pp. 231-51.

Stroud, B. (1998). "The Goal of Transcendental Arguments," in Stern (1998a).