Chapter Nine

Arguing for Apperception

1. Introduction

Kant's theory of the transcendental unity of apperception emerges in the course of his argument for the objective validity of the categories. In this central chapter, I try to show how the argument for apperception works and what the resulting theory looks like. My analysis begins with a more precise characterization of the type of cognition from which the argument regresses. I then consider exactly what the principle of apperception says. I examine the argument for apperception where it first enters the (first) text, in the discussion of the so-called 'Third Synthesis' in the A deduction (though with some clarifying interpolations from the second edition). I then turn to the version of the argument presented in the B deduction, which includes some important additional considerations.

Although my focus is on the argument for rather than from the principle of apperception, I show how that argument fits into the overall defense of the categories for the parade case of causation—a clear test for any interpretation. I conclude with a discussion of the relation that between empirical and transcendental apperception. On this point I am somewhat critical, arguing that Kant's claim that empirical apperception presupposes transcendental apperception is true, but only half the truth. As his theory stands, transcendental apperception also presupposes the empirical variety.

The preparatory materials in Chapters 7 and 8 were intended to permit a relatively uninterrupted presentation of the deduction argument here. For similar reasons, I briefly consider two alternatives before offering a continuous line of interpretation. In showing why these *prima facie* attractive approaches should be rejected, I defend and motivate my competing analysis.

2. 'I-think' as the Cogito; the One-Step Deduction from Judgment

As I read it, the deduction examines the requirements of (rational) empirical cognition and argues for the necessary togetherness of different representations for such cognition; from that intermediate result, it moves to a defense of the categories. Two textual clues invite different construals of the deduction argument. First, Kant describes the principle of apperception as the 'first and supreme' principle of his philosophy and he casts it in terms of the possibility that the representation 'I-think' can accompany any representation. Second, in an oft-cited note to the *Metaphysical Foundations of Natural Science*, he claims that he can solve the problem of how the categories relate to experience

> almost through a single inference from the precisely determined definition of a **judgment** in general (an action through which given representations first become cognitions of an object) (4.475-476, C1781190).

The first clue makes it tempting to think that the argument is very like that of Descartes' *Meditations* and so begins with 'apperception' as the bedrock on which all of cognition rests. The second may suggest that judgment—as opposed to apperception—provides the key to linking intellectual concepts and activities to the manner in which objects are represented by the senses. These

clues can mislead. Kant could not have begun with apperception and his musings about judgment do not imply that he ever thought that the deduction could be completed without a full exploration of apperception, without his '*Egologie*.' It's not that judging isn't of central importance to the deduction; it's that it forges an indissoluble link between judging and apperception.

We've already encountered the contextual and textual reasons for rejecting the suggestion that the principle of apperception is like the *cogito* in being first known and so a suitable first premise for an argument about cognition.[1] Wolff and his followers actively debated Descartes' claim that the *cogito* ranked first in the order of knowing (Chapter 5, section 3). Against the background of this debate, and of the general (unresolved) question of how self-consciousness is possible, Kant should not have assumed that '*Ich denke*'—I am conscious of myself as thinking—was acceptable as a first premise. The debate also suggests a further hypothesis for the deduction's complex structure. His weaving together of considerations about the requirements of object cognition and about the role of the subject in cognition may reflect his appreciation of the solid arguments on each side that the other side was one-sided.

We have also seen some of the textual reasons for denying that Kant cast his argument in a Cartesian mold. As did many of Wolff's successors, he formulates the *cogito* as an inference—and then rejects it (e.g. A355). 'I exist' could not be established inferentially. In the Anthropology lectures, he claimed instead that it was an 'intuition' (25.10, 244, 473-74). Had he continued to believe that cognizers knew of their existence (as simple and identical beings)

through intuition, then he might have offered the principle of apperception as a first premise. In fact, he explicitly rejects intuition as a possible basis for 'I exist' in the *Critique* (e.g. B135). He discusses three types of intuitions, intellectual intuitions, outer intuition of objects in space, and inner intuitions of the mind. Intellectual intuitions are impossible for humans. In the B edition, he is also explicit that

> If I were to think of an understanding that itself intuited ... then in regard to such cognition the categories would have no signification whatever (B145).

Not only are intellectual intuitions impossible for humans, if they were possible, then the assumption that humans have an intellectual intuition of themselves would be a uniquely bad premise for a deduction of the categories. Put positively his claim is that the absence of an intellectual intuition of the 'I' is an essential presupposition in the argument for the categories. That leaves outer and inner sensory intuitions. Outer intuitions can present only bodies, not minds. So, only inner intuition could present a thinking self. But he is clear that inner sense can provide no intuition of a self (A107, B134). Given the historical context and his views about intuition, it is hard to see how he could take the unity and identity of a thinking self as a first premise.

Further, the deduction texts are evidence against the '*cogito* as premise' hypothesis. Especially in the A deduction, the discussion does not begin with apperception, but works its way to apperception as a necessary condition for cognition. But even the B deduction does not begin with apperception; it begins with the need for combination and unity in cognition. Apperception enters as a

solution to the problem of the unity required for combination. Although Kant does not argue from the *cogito*, he tries to situate it in relation to his views in the B edition. Chapter 11 explains how he saw the relation between his 'I-think' and that of his illustrious predecessor.

Kant's speculation about the possibility of a one-step deduction from a precise definition of 'judgment' makes two different approaches seem attractive. We considered one of them in Chapter 8. Kant's gloss of judging as 'an action through which given representations first become cognitions of an object' strongly suggests that objective reference is the key to the deduction. Chapter 8 argued that objective reference cannot be the whole story, because it cannot explain crucial aspects of the theory. One could also take the focus on judgment to indicate that the 'forms of judgments' are as central to the transcendental deduction as they are to the so-called 'metaphysical deduction.'[2] This seems to be Longuenesse's (1998) view, because she devotes most of her interpretive efforts to explaining how the forms of judgments relate to the syntheses of the imagination, thereby resolving the Critical problem. Her analyses are extremely useful, but they tend to displace the unity of apperception and the categories from their central role in the deduction.

The *Metaphysical Foundations* reference to 'judgment' is misleading, because it mentions only one aspect of Kant's use of this term—its relation to object cognition. It is a firm part of his general theory that cognition always has two sides, a subjective and an objective side. Under the heading of 'Cognition in General,' his (i.e. Jäsche's) *Logic* explains that

> All our cognition has a **twofold** relation**,** first a relation to the **object**, **second**, a relation to the subject. In the former respect it is related to **representation**, in the latter to **consciousness**, the universal condition of all cognition in general (9.33, CLog 544).

Given the generality of this claim, a complete or adequate definition of 'judgment' should have a second, subjective side. He presents that side—indeed only that side—in his definition of 'judgment' in the *Logic*. Under the heading '*Definition of a judgment in general,'* he explains:

> A judgment is the representation of the unity of the consciousness of various representations, or the representation of their relation insofar as they constitute a concept (9.101, CLog 597).

Checking Jäsche's version against Kant's handwritten notes reveals that he offers an abbreviated, but faithful, representation of the Reflections:

> The representation of the way in which different concepts (as such) belong to a consciousness is judgment. They belong to a consciousness according to laws of imagination, thus subjectively, or [according to laws] of understanding, that is [are] objectively valid for every being that has understanding. The subjective connection depends on the special situation of the subject in experience (16.633).

In light of these fuller discussions of 'judgment,' the *Metaphysical Foundations* proposal of a 'one-step' deduction from the definition of judgment does not shift the enterprise away from a study of consciousness and apperception.

The book on science appeared a year before the B edition of the *Critique*. So we may go to that text to see how the 'one-step' deduction was carried out. There we encounter the familiar discussion of the apperceptive principle. The focal point of Section 19 of the B deduction—the paragraph about judgment—is

the subjective side of the complete definition. It repeats the points made in the notes:

> But suppose that I inquire more precisely into the reference of given cognition in every judgment, and that I distinguish it, as belonging to the understanding, from the relation in terms of the laws of the reproductive imagination (a relation that has only subjective validity); I then find that a judgment is nothing but a way of bringing given cognitions to the objective unity of apperception (B141).

Hoppe and Longuenesse are well-aware of these texts.[3] Hoppe makes an interpretive decision to avoid the '*Egologie*.' Longuenesse uses the close relation between apperception and judgment to interpret apperception—and so the deduction—through judgment rather than judgment through apperception (1998, 57-58, 69). Below I argue that we can understand crucial moves in the deduction only by recognizing the independent status of the *a priori* representation 'I-think' and its principle. My interpretive point here is that the somewhat autobiographical note in the *Metaphysical Foundations* should not be understood as moving judgment—rather than apperception—to the center of the deduction. Given the subjective aspect of Kant's 'judgment' definition, the note does not undermine, but reinforces the centrality of the unity of apperception, a centrality that is obvious in the texts.

3. What Kind of Cognition is at issue in the Transcendental Deduction?

There is fairly wide agreement that the distinctive feature of the transcendental deduction is that it argues for the legitimate use of the categories by arguing (in the Principles chapter) that the categorial principles that invest

them with meaning are necessary conditions for the possibility of experience.[4] Chapter 7 presents reasons for glossing 'experience' as 'empirical cognition'. But exactly what sort of 'empirical cognition' does Kant propose to investigate? Or better, how does he understand 'empirical cognition'?

One of the most well-known theses of the *Critique* is the claim that cognition requires both intuitions and concepts (e.g., A51/B75). In Section 1 of Chapter 1 of the Analytic of Concepts, Kant maps out the relation he sees between intuitions and concepts. A concept is or involves a function of unity among representations. It collects or combines different representations, intuitions or (other) concepts, under a common representation (A68/B93). So, for example, the concept 'dog' collects many intuitive representations of dogs under it. It also can contain other conceptual representations, for example, 'animal,' 'possessing a tail' (A68/B93ff.). He explains further that concepts (and the judgments they make possible) cannot stand in direct relations to objects:

> The only kind of representation that deals with its object directly is intuition. Instead the concept is referred directly to some other representation of the object (whether that representation be an intuition or is already a concept) … In every judgment there is a concept that is valid [*gilt*] for many representations, and, among them it comprehends [*begreift*] also a given representation that is referred directly to the object (A68/B93, amended translation).

That is, judgments about objects are possible only when the concepts they contain can be understood as representations that express what is common to many intuitive representations, including an intuition that stands in a direct

relation to the object (through that object causing certain changes in the mind [A19/B33]).

Although the *Critique* stresses that the cognition at issue is conceptual or discursive, it does not spell out how such cognition should be understood. Fortunately Kant's *Logic* is more explicit about the relations among 'discursive' cognition, concepts, and the 'marks' of concepts:

> From the side of the understanding, human cognition is discursive, i.e., it takes place through representations which take as the ground of cognition that which is common to many things, hence through marks as such. Thus we cognize things **through** marks and that is called **cognizing** [*Erkennen*], which comes from being acquainted [*Kennen*].
>
> A **mark is that in a thing which constitutes a part of the cognition of it,** or—what is the same—**a partial representation, insofar as it is considered as a ground of cognition of the whole representation**. All our **concepts** are marks, accordingly, and all **thought** is nothing other than a representing through marks (9.58, CLog 564, my underscoring).

(Here Jäsche faithfully reproduces materials from R 2277, 2279, and 2281, 16. 297-98).

The text that Kant used for the course, G.F. Meier's *Excerpts from the Doctrine of Reason*, describes one use of marks (16.296). Marks are a ground of cognition, because they enable cognizers to differentiate the object of cognition from other things (24.113, 16. 299). For example, it is orange and other things are not. By contrast, Kant maintains that marks not only have this 'external' use but also an 'internal use.' In the latter case marks or partial representations are the ground on which the whole concept is applied to the object, not via identity

and differences with other objects, but via identity of the marks (R2282, 16.298). The internal use of marks is not a matter of differentiation but of derivation (16.299). So for example, the concept 'body' might include the marks 'impenetrable,' 'extension' and so forth (2.58, C1755 1027, A106) and so be applied through the tacit judgments 'this thing is impenetrable', etc.. Under these circumstances, one could derive 'x is impenetrable' from 'x is a body.'

This material is neither new nor controversial,[5] but I wish to highlight two important aspects of Kant's 'mark' theory of concepts. First, marks are concepts. The part-whole relation is not between intuitive representations and conceptual ones, but between concepts that are part of the content of other concepts and the latter 'whole' concepts. This is not a foundationalist account of conceptual cognition—not an explanation of how sensations or intuitions lead to concepts. The second noteworthy aspect of Kant's mark theory is that for a representation to be a mark, it must be considered as such. It is not that cognizers must have the concept 'mark' or the concept 'concept' or the very abstract concept 'representation.'[6] Rather, they must recognize that a mark—say 'impenetrable'—is a partial ground or basis for their application of the concept 'body.' It is part of why they call something a 'body' or part of what they presuppose in calling something a 'body.' Even when not considered in relation to a complex concept, but just on its own, a mark is still a basis of cognition, because it is the ground of cognition of the objects in its extension[7]:

> As one says of a **ground** in general that it contains the **consequences** under itself, so can one also say of the concept that as **ground of cognition** it contains all those things

> under itself from which it has been abstracted, e.g. the concept of metal contains under itself gold, silver, copper, etc. (9.96, CLog 594).

'Metal,' for example, is a ground of cognition of copper things, because it classes them together with other metals. In describing a copper kettle as 'metallic' a cognizer implies that it is similar to some things and different from others. To recognize a mark as such is to use the term with that understanding.

Humans not only have sensory systems that detect similarities in the properties of objects, they are also discursive cognizers. As such, they can recognize their representations of, e.g., the color orange as presenting a humanly detectable property common to many things and (so) as marks—as things that can be offered in answer to the question of why they group those various together (and exclude others) and of why they label something as a particular kind of fruit, an 'orange.' Since the cognition at issue in the *Critique* is discursive, conceptual, it should be understood as mediated by marks that are considered as such.

Kant also stresses the importance of marks to human cognition in his objection to Meier's suggestion that animals use concepts, a criticism discussed in Chapter 2. I repeat a portion of the relevant citation for easy reference:

> I would go still further and say: it is one thing to **differentiate** things from each other, and quite another thing **to recognize** [*erkennen*] the difference between them (2.59-60, C1755 103-104).

Having conceded to Meier that an ox can differentiate its stall by seeing that it has a door and so, in a sense, through one of its marks, Kant points to the crucial

difference with the human case. It is the same point just highlighted in the *Logic*. Something can function as a mark only if it is recognized as the basis of differentiation.[8]

The cognition that is the topic of the *Critique*'s investigation is not merely empirical, that is, linked to sensory representations. It is also conceptual and for Kant that means that it is a matter of 'marks' or grounds of cognition that are recognized as such. This point will be crucial to Kant's argument linking cognition to apperception. Since a standard definition of 'rational cognition' is cognition where the cognizer can give the reason—or ground—of the cognition, I use the slightly odd description 'rational empirical cognition' (or, more briefly, 'RE cognition'[9]) to characterize the *Critique*'s project. He does not use the expression 'rational cognition' for two reasons. First, he complains in his Logic lectures about the expression. It is too vague and such cognitions should be called 'cognitions of reason,' because that gives their source (24.151, CLog 119). Second, in the *Critique*, he has reasons for distinguishing the 'higher' faculties of 'reason' and 'understanding.' Given the latter usage, the cognition at issue might better be called 'cognition of the understanding' or less formally, 'conceptual cognition,' because one of its sources is the understanding, the faculty of concepts. I use 'RE cognition,' because it captures more fully the sort of cognition whose source is the 'higher faculties,' *viz*., cognition where the subject knows the reason.

Kant's interest in RE cognition may suggest that he would be on the 'internalist' side of recent debates.[10] As noted in Chapter 2, however, he lists

other varieties of cognition besides the rational. The deduction does not argue that all cognition requires internal justification, but that the RE cognition of which humans are capable is possible only when certain conditions are met. It is not an argument for internalist justification, but an argument from the assumption that, unlike animals, humans can know the reasons for their cognitions.

4. What is the Principle of Apperception?

In earlier chapters and earlier in this chapter I refer to the 'principle of apperception.' I have also indicated the general shape of the deduction by reference to the principle. It regresses from the possibility of rational empirical cognition to the truth of the apperceptive principle; it will then argue that the principle of apperception requires that the categorial principles hold across all possible objects of the senses. Since Kant describes it as the supreme principle of cognition and of his theory, one might expect a fairly clear statement of it. The interpretive issue is complicated because of the two *prima facie* inconsistent treatments in the two editions.

Before turning to these texts, it may be useful to recall the longer and shorter historical context. Leibniz took the soul to be a special sort of monad. As a monad it was a simple substance whose states changed through time *via* the operation of a timeless internal principle. Souls possess not only substantial unity, but also memory and some awareness of the transitions from one state to another (though many perceptions are unconscious). The internal principle is also the ground of the conscious transitions and memories that provide subjects access to their own identities (Chapter 4, section 2). Although it is more

nuanced, Leibniz's theory fits the description that Hume offered for the Rationalist accounts that were his target:

> Some philosophers …imagine we are every moment intimately conscious of what we call our Self; that we … are certain, beyond the evidence of a demonstration, both if its perfect identity and simplicity (1739/1978, 251).

Lecture notes from his courses in Metaphysics and Anthropology suggest that Kant held something like this position in his pre-*Critique* days. Apparently he maintained that in thinking subjects are conscious of themselves as simple, substantial, and as continuing subjects (Chapter 4, p. nn). In this tradition a cognizer who is aware of his representations is aware of something that, as a matter of ontological necessity, belongs with other representations to a simple, permanent, and continuing self. A Reflection from the *Duisburg Nachlaß* presents a more subtle view. In thinking cognizers are conscious of representations as set in the mind according to three exponents: relation to subject, relation of following and relation to the whole (Chapter 6, citation **F**, p. nn). This claim is not ontological. It is not that the representations of which the subject is conscious must belong with others to a subject-substance. It is that, in thinking, cognizers represent their representations as belonging to a subject, as successive, and as coherent.

Given this background and Kant's general program of 'reforming' metaphysics by recasting some of its key doctrines as presuppositions of cognition (most famously, 'all events have causes'), we should expect him to argue that the assumption that different representations all or necessarily belong

to a common subject is required for RE cognition. That appears to be just what he is doing at the beginning of the A deduction's 'systematic' presentation of the argument for apperception:

> We are conscious a priori of the thoroughgoing identity of ourselves in regard to all representations that can ever belong to our cognition, and are conscious of it as a necessary condition for the possibility of all representations. (For any such representations represent something in me only inasmuch as together with all others they belong to one consciousness; and hence they must at least be capable of being connected in it.) This principle holds a priori, and may be called the transcendental principle of the unity of whatever is manifold in our representations (A116, my underscoring).

A cognizer's consciousness of his identity is no longer a matter of intuiting something that is, as a matter of ontological necessity, a continuing self. It is an awareness of self-identity as a necessary condition for the possibility of representations that can represent something. In a note to the text, he elaborates the point:

> All representations have a necessary reference to a possible empirical consciousness…But all empirical consciousness has a necessary reference to a transcendental consciousness (a consciousness that precedes all particular experience), *viz*., the consciousness of myself as original apperception. It is therefore absolutely necessary that in my cognition all consciousness belongs to one consciousness (that of myself). Here, then, is a synthetic unity of the manifold (in consciousness) which is cognized a priori…The synthetic proposition that all the varied **empirical consciousness** must be combined in one single self-consciousness is the absolutely first and synthetic principle of all our thought as such (A117n., my underscoring).

Any representation (that can be a part of cognition) must be able to be conscious. Although Kant does not explain why this is so, the rationale is clear from his general view of RE cognition. Were such consciousness impossible, a cognizer could never know the grounds of his cognition and so could not be a rational cognizer.

The 'epistemological turn' comes in the next step. The metaphysical proposition that any representation must belong with all others to a continuing self is replaced by the claim that in human cognition, empirical consciousness has a necessary reference to original apperception, that is, to one (continuing) consciousness. Matters become slightly confusing because the principle that Kant highlights has a complex structure: 'All the varied empirical consciousness *must* be combined in one single self-consciousness (for cognition to be possible).' He describes the proposition as 'synthetic.' Since it is also cognized *a priori*, it appears to be a synthetic *a priori* proposition, which is the same status as that of categorial principles. The principle of determinism is that 'all changes occur according to the law of cause and effect'; the claim that would be parallel to the apperceptive principle for the case of determinism would be: 'All changes must occur according to the law of cause and effect (for cognition to be possible).' The latter is a description of the thesis to be proved in relation to the causal principle. It is not the causal principle itself.

We can sort out the complications here by recognizing that the role of the deterministic principle in transcendental idealism involves two necessities and two apriorities. It is necessary for cognition that events are understood as

standing in relations of necessary succession. The concept 'cause' cannot be extracted from experience, but is '*a priori*'; the proof that the causal principle is required for cognition is carried out by reason and so is also *a priori*. Kant is not explicit about the double use of these concepts, because he takes it to be obvious that the only possible proof of a universal and necessary proposition ('all events have causes,' 'all representations belong with others to a single self) would show the necessity of the proposition for cognition. He also takes it to be evident that the only possible defense of the use of an *a priori* representation or principle would have to be *a priori* from reason.

With this background, we can understand the structure of the apperceptive principle and also see that the roles played by the principle of determinism and by the metaphysical principle that 'for anything that is a representation, it must belong with others to some continuing self' are largely parallel. For ease in reference, I call the latter principle the 'I-rule.' I should note, however, that my nomenclature reflects Kant's position that the difficult philosophical problem about the ownership of representations concerns their 'togetherness' and not their 'mineness.' What needs to be shown is how different mental states can be unified in a single self and not how an individual can attribute a particular mental state to himself (which he thinks is *via* inner sense). So he assumes that it is sufficient to show that different mental states belong to a common *self*-consciousness or to the same 'I' to show that they necessarily belong together. At various points below, I take note of the fact that being necessarily connected to each other is only a necessary condition for being referred to a common '*I*.'

In the A deduction, Kant rules out the possibility of an *a posteriori* basis for connecting a representation to others in a continuing self (A107). Since he takes representations to be either *a posteriori* or *a priori*, the representation of the unity of consciousness must be *a priori*. The B deduction presents the representation 'I-think' as *a priori* (B132). As in the case of the causal principle, the I-rule is *a priori*. If it is also synthetic, then the only way it can be established by the standards of scientific metaphysics is through an *a priori* argument showing that it too is a necessary condition for cognition. And that appears to be the thesis expressed in the apperceptive principle: 'All representations must (if cognition is possible) belong to a single self-consciousness.' Alternatively: it is necessary for cognition that representations are necessarily connected to a single 'I.' Part of the burden of A deduction is to establish this highest principle, the thesis that the holding of the I-rule is necessary for cognition. In this way, it makes the same move with respect to the I-rule that the Second Analogy makes with respect to the causal principle. Both are removed from metaphysics and relocated among the assumptions that must hold for cognition to be possible. What is somewhat out of place is the focus on the synthetic status of the claim that cognition requires that different representations belong to a common self.[11] If it is to be established by *a priori* reasoning, then the claim should be analytic.

The A and B deductions appear to conflict on this point. In B, the apperceptive principle is characterized as 'analytic' (B135). The apperceptive principle of the B deduction is, however, a different principle. It is not the famous 'I-think' claim:

> The **I think** must be **capable** of accompanying all my representations; for otherwise something would be represented in [*in*] me which could not be thought … (B131-32, amended translation and punctuation.)

In this text, as in the 'absolutely first principle' in A, the 'must' is an indication of a requirement of cognition. The principle that is characterized as analytic is presented later:

> All **my** representations in some given intuition must be subject to the condition under which I can ascribe them as **my** representations to an identical self, and hence under which alone I can collect them together [*Zusammenfassen*] as synthetically combined in one apperception through the common expression 'I think' (B138).

In this principle, the 'must' does not indicate a general requirement of cognition: i.e., these representations must be combined together in one identical self for cognition to be possible. It refers instead to the fact that, if different representations are linked to a common 'I,' then they must meet whatever conditions are required for belonging to a common 'I.'

This claim is not as banal as the claim that 'anything classified as green must meet the conditions for greenness,' because the condition of belonging to an 'I' is a matter of belonging with other representations to the same 'I.' Kant says this explicitly (though without the contrast to other sorts of concepts):

> For the manifold representations given in a certain intuition would not one and all be **my** representations, if they did not one and all belong to one self-consciousness. I.e., as my representations (even if I am not conscious of them as being mine), they surely must conform necessarily to the condition under which they **can** alone stand together in one universal self-consciousness (B132).

Still, the claim that 'all my representations must conform to whatever condition or conditions are necessary for them to be collected together as my representations' or, more generically, the claim that 'any representation that belongs with others to one consciousness must meet the conditions of belonging with others to one consciousness' is a tautology. Since a tautology can have no implications beyond logical truths, it is natural to ask about the relevance of this tautology to the *Critique*'s project. Kant showcases its relevance in the dramatic B132 claim about the 'I-think.' For a representation to be of any cognitive use to its bearer it must be possible to think something through it,[12] and to think something through it, it must belong with others to a single 'I-think': It must be possible to attach a communal 'I-think' to it. Hence the opening sentence of §16 that is rightly thought to be the essential point of the B deduction (the I-think must be able to accompany all my representations) is the equivalent of A's principle of apperception. Although the principle of the necessary unity of apperception in B prepares the way for the argument for the categories, it is a tautology that would be of no interest—because it might have no instances—except for the 'I-think' doctrine of B132. Given that doctrine, if cognition is possible, then there must be instances of different representations belonging to a common 'I.'

Two questions remain. Is the A deduction apperceptive principle/B deduction 'I-think' doctrine synthetic? Is the I-rule synthetic or analytic? Kant does not say clearly in either case. He describes A's apperceptive principle as 'synthetic,' but is silent about the status of its equivalent in B. The formulation of the 'I-think' passage and the apperceptive principle in A both suggest that he

thinks of the I-rule as having a synthetic *a priori* status and thus as needing to be established as necessary to cognition. But he doesn't say. Instead, he stresses that although the B deduction's principle of apperception is analytic, it presupposes a synthesis (B134, B135). By this, he means that synthesis is required for there to be any instances that meet the conditions of the I-rule, that is, any representations that belong with others to a single self-consciousness. Since being able to attach representations to an 'I-think' is necessary for cognition, such syntheses would also be necessary for cognition. His silence on the status of some key claims may reflect his belief that these last links among cognition, the I-rule and synthesis are all that he needs for his argument for the categories. And perhaps his earlier claim that the A edition apperceptive principle was 'synthetic' was only an awkward attempt to express the necessity of synthesis to the possibility of applying the I-rule and so to the possibility of cognition.[13]

Since the proof of the A edition principle of apperception/B edition 'I-think' doctrine is a matter of bringing out the conditions required for RE cognition, the principle must 'analytic,' even if it takes an enormous amount of work to reveal its presuppositions. The same is true for the I-rule. When 'representation' is understood as something that can take part in RE cognition, it turns out that, on careful analysis, any representation must belong with all others to a single consciousness. In this respect, the I-rule is not parallel to the causal principle. But the difference is not particularly significant for Kant's reform program in metaphysics. He redeploys both allegedly ontological principles as requirements

of cognition. The difference between them is that the I-rule is so closely connected to the requirements of RE cognition, that the rule itself and not just the principle that cognition requires the rule turns out to be analytic.

In sum, A and B deductions present four interrelated principles connected to apperception.

1. The principle of apperception in A: 'All the variety of empirical consciousness must be combined in one single self-consciousness' (A117n).
2. The I-rule that is embedded in the A edition principle of apperception and in the 'I-think' doctrine of the B deduction: 'Representations … [belong] with all others to one consciousness' (A116); 'The **I think** …[accompanies] all my representations' (B132).
3. The 'I-think' doctrine of B: 'The **I think** must be **capable** of accompanying all my representations' (B132).
4. The tautological B edition principle of apperception: 'All **my** representations in some given intuition must be subject to the condition under which I can ascribe them as **my** representations to an identical self' (B138).

The last is the least important. What the transcendental deduction is to establish is the I-think doctrine/A edition principle of apperception. In subsequent sections, I use 'principle of apperception' and 'I-think doctrine' interchangeably to refer to that thesis. Since the I-rule is embedded in the principle, it will also be a central

topic, particularly in relation to the claim that it can be satisfied only when representations are synthesized.

5. The Apperceptive Synthesis of Recognition in a Concept

With a clearer understanding of the sort of 'experience' or 'empirical cognition' that constitutes the beginning of the regressive argument and of the conclusion to be established, we can turn to the text. 'Apperception' enters the A edition in the 'Third Synthesis,; which is variously described as the 'synthesis of apperception' and as the 'synthesis of recognition in a concept' (A94, A103, A115). My clunky section title—which Kant avoids, presumably on stylistic grounds—reflects the intimate connection between the unity of apperception and conceptual cognition (or judgment) in his theory. I do not begin the discussion with the Third Synthesis just because that is the point of entry for apperception. There are philosophical problems with his discussion of the Second Synthesis. It is supposed to explain the operation of the law of association, but cannot do so (Chapter 8, pp. nn). The placement of the synthesis of apprehension first in the A edition is also problematic. Kant's considered view is that apprehension is subject to the rules of the intellect, not vice versa (B161-62). Presenting the synthesis of apprehension as the 'First' in A gives the impression that it operates 'before' the others. The B deduction's treatment, which turns to the *synthesis speciosa* after the argument for the 'I-think' doctrine, is better in this respect. Although I start with the presentation in the A edition, I follow the better order of B. Moreover, I draw on some better formulations of the B deduction to clarify A's line of argument—and some of the more expansive discussions in A to fill in

more cryptic moves of the B argument. I also refer to both editions in this section and the next to establish the commonality of some claims.

Kant's discussion of the Third Synthesis is unusual, because it employs a detailed example. The example is counting. Three reasons help to explain the selection. Mathematics and logic are the two obvious counter-examples to his thesis that cognition requires both concepts and intuitions. The example illustrates the role of sensory materials in arithmetic. A second reason is that he believes that construction in mathematical proofs can serve as a model for the constructive activities of the understanding. In this case, the point isn't to illustrate the production of continuous lines, but the fact that mathematical concepts enable cognizers to construct sensory (written or imagined) objects fitting the concepts. The third reason is connected to the second, though it may have a different motivation. In mathematics, cognizers do not have to wait for their senses to supply them with information in order to apply a concept. They can construct the 'data' for themselves. This feature enables him to present the example without having to be concerned about transitions between sensory states caused by perception, which are something that his theory will need to explain.

Although Kant has reasons for choosing a mathematical example, it is meant to illustrate judging or recognizing in a concept in general and this is somewhat problematic. He believes that all concepts are associated with rules[14] (e.g., A106). Since mathematical concepts are usually understood as having definitions, the example may suggest that the associated rules are necessary

and sufficient conditions for the applicability of the concept. In fact, he thinks that neither empirical concepts nor categories can be defined (9.141-41, CLog 632, A727/B755-56).[15] The associated rules are not definitions, but expositions (incomplete analyses) of the concept (A728-29/ B756-57). He thinks of these rules as universal, but for the case of empirical concepts at least, it wouldn't matter to his theory if the rules were probabilistic.

Once the suggestion of necessary and sufficient conditions is rejected and the rules are allowed to be probabilistic, Kant's assumption that concepts are associated with rules can be seen as a version of the standard contemporary view that concepts stand in inferential relations to other concepts and can be used only by individuals who explicitly or implicitly recognize those relations. The rules indicate some of those relations. In the case of concepts that are either not complex or not clear (where the subject doesn't know the inferential relations), the rule would be that of the external use of marks (above, p.n)—the rule that the concept indicates a property that can be detected by humans and is common to this object and others.

On Kant's telling, counting is more complicated than one might think. In the context of the A discussion of the Third Synthesis, he explains that it is insufficient for cognition just to apprehend something in intuition (the First Synthesis), and just to be able to recall what has been apprehended (the Second Synthesis).

> Without the consciousness that what we are thinking is the same as what we thought an instant before, all reproduction in the series of representations would be futile. For

> what we are thinking would in the current state be a new representation, which would not belong at all to the act by which it was to be produced little by little. Hence the manifold of representations would never make up a whole, because it would lack that unity that only consciousness can impart to it. If, in counting, I forget that the units that now float before my mind or senses [*Sinnen*] were added together by me one after another, I should never cognize the amount [*Menge*] being produced through this successive addition of unit to unit; nor, therefore, would I cognize the number. For this number's concept consists solely in the consciousness of this unity of synthesis (A 103, amended translation).

His first point is that the mere ability to reproduce the series of stroke symbols, say four of them in a row, is insufficient for cognition. He does not refer to animals here, but the contrast is useful. An ox could have an image of four stroke symbols, but he could not recognize its contents under the concept '4.'

Using the counting rule (or any rule associated with a concept) involves a number of skills. A counter must be aware of his performance in such a way as to catch possible errors and, in this case, to know where he is in the process. Kant presupposes all this in making his second and positive claim: The counter needs to be conscious that he designates the first stroke symbol as '1' etc. in order to cognize the amount. In the terms used in the *Logic*, the counter needs to be conscious of his representations of '1,' '2,' etc., as 'marks' or 'partial representations' that are the ground or basis of his conceptual representation '4.' (The representation '2' is a partial representation of the whole representation '4,' because any because any group of 4 objects must contain a group of at least 2 objects.[16]) The last sentence of the passage notes that, in this case, being conscious of applying the counting rule to the individuals and sub-groups of a

group is not only necessary for applying the concept to the group, it is also sufficient.

Kant elaborates and refines the account in the further discussion of this example:

> The very word concept could on its own lead us to this observation. For this one consciousness is what unites in one representation what is manifold, intuited little by little, and then also reproduced. Often this consciousness may be only faint, so that we do not [notice it] in the act itself, i.e. do not connect it directly with the representation's production, but [notice it] only in the act's effect. Yet, despite these differences, a consciousness must always be encountered, even if it lacks striking clarity; without this consciousness, concepts, and along with them cognition of objects, are quite impossible (A 103-104).

He allows that thinkers do not have to pay much attention to individual steps, adding up the stroke symbols little by little. Still, they must be conscious of the judgment '4' as the effect of applying to the represented units the rule that a group has the size of '4' if and only if it consists in four units.[17] The discussion concludes with the very strong claim that without this consciousness, cognition of objects would be impossible.

Given Kant's view of how concepts are employed, we can understand why he thinks RE cognition requires consciousness in acts of judging. Conceptual cognition requires that partial representations, '1', etc., are not merely representations that float before the mind. They must be understood as partial representations, as the basis of the whole representation '4.' If cognizers did not consciously apply the concept '4' on the basis of the representations of the units,

then they would not know the basis of their judgments. They would not know that they judge '4,' because they have already noted '2' units with more to follow. With arithmetical concepts the rules offer necessary and sufficient conditions for the applicability of the concept. So a cognizer could infer that the number of units in the group is '4.' Where the rules are merely partial explications, e.g., 'bodies are extended,' the judgment that x is a body has 'x is extended' as its basis or partial ground, but the judgment 'x is a body' is not a valid inference from that ground.

Since the hypothesis of an awareness of mental acts can seem strange to contemporary readers, I offered some evidence for the phenomenon in Chapter 2. It would, however, have been familiar to Kant's contemporaries from the Port Royal Logic and from other standard texts. So he needn't elaborate on mental act awareness *per se*. His efforts at clarification take the form of a contrast between 'apperception' and 'inner sense.' Inner sense is introduced in the Transcendental Aesthetic. In arguing that a further faculty of apperception is needed, he rejects his earlier view that Lockean inner sense is the key faculty in RE cognition. We can get a better sense of how he understands both cognition and the active self-conscious faculty of apperception by considering why he no longer believes in the adequacy of inner sense.

Chapter 2 describes Tetens's efforts to explain why 'inner sense' should be understood as a sense (p.nn). On his account, inner sense is a sense, because it records acts of thinking. A mere record of having judged would, however, be insufficient for RE cognition as Kant understands it. The proper use

of concepts requires the counter to be conscious, not that he has judged '4' or even that he is judging '4' right now. It is not even enough that he be aware through inner sense of his partial representations, '1', etc. *He must regard his partial representations as such, as the grounds of his cognition; he must be aware of judging '4' on the basis of his partial representations. To be capable of rational cognition, the subject must be aware—as she makes the judgment—of her act as having the appropriate basis. And that requires act-consciousness.* The creation of impressions of mental actions in a Tetensian inner sense is too little and too late to contribute to the rationality of the judgment.[18]

Locke had introduced 'inner sense' in terms of the awareness of mental states and of mental actions. Chapter 2 presented Kant's canonical account of the difference between inner sense and apperception in the *Anthropology* (p. nn). He invites his readers to contemplate being aware of their mental states through inner sense and being aware of their mental acts through apperception so that they may see how very different these awarenesses are. In that chapter, I raised a question: Why can't the difference lie in the objects of awareness rather than in the faculty that is so aware? The answer is now clear. The awareness in judging that is essential to cognition is not a perception of an act that could be separate from it. It is an indissoluble component of the self-conscious act itself.[19] Inner sense will not do, because when properly understood as a sense, it involves perception of having acted, not self-conscious action.

At this point, the A deduction has argued for an active self-conscious faculty of apperception. Chapter 10 examines the faculty of apperception in

more detail. This chapter follows the argument for the unity of apperception. After the counting example, Kant broadens the argument to any object of cognition. He also introduces the idea of the concept of an object as providing a rule or standard for what makes something an object (A104-105). At one level, the object-rule is just another example of a rule associated with a concept, because all such rules provide standards for using concepts. For example, something falls under the concept 'dog' only if it falls under the concept 'animal.' On the other hand, the object-rule is extremely broad, because it covers all objects of cognition *per se*. He then makes a *prima facie* peculiar equation:

> We are, however, dealing only with the manifold of our representations. And since that x (the object) which corresponds to them is to be something distinct from all our representations, this object is nothing for us. Clearly, therefore, the unity that the object makes necessary can be nothing other than the formal unity of consciousness in the synthesis of the manifold of representations. When we have brought about synthetic unity in the manifold of intuitions—this is when we say that we cognize the object (A105).

Since he was describing the unity of an object (the rule is that the representations of an object of cognition must not be haphazard or arbitrary but determined in a certain way), it seems strange suddenly to introduce the unity of consciousness.

Kant's move is more intelligible against the background of the disagreement over the priority of object and self consciousness. How can one claim that object consciousness (reference to an object) comes first? On the representational theory of perception common to the disputing parties, all that

cognizers have available to them are particular representations of sensible properties. How is the cognizer to refer these to a common object? Kant's assumption is that she can do so only if she has available some *a posteriori* or *a priori* object rule that enables her to determine which representations could be representations of the same object. As the discussion of the counting example makes clear, however, applying such a rule requires her to be aware of her partial representations (e.g., the representations of a certain shape and of self-propelled motion) and of her mental act as the basis on which she judges the object to be a 'dog.' This awareness not only permits cognition of the object, however, it essentially involves recognizing a relation across the partial cognitions or mental states and the judgmental state, the relation of necessarily belonging together. Since the partial representations are the basis of the judgment, it could not be a judgment—an example of RE cognition—in their absence. Unless the partial representations could be used in producing such a judgment, they could not participate in RE cognition. It follows that since an object rule can lead to cognition if and only if it can be applied to (partial) representations in such a way that it involves recognizing a relation of necessary connection across states, the functioning of an object rule implies the unity of consciousness. And since cognizers have no way to refer representations to objects by grabbing onto an object independently of their representations, the only way that they can refer them to objects is through an object rule. Hence, oddly enough, the key to understanding how representations can refer to an object are the conditions required for a unity of consciousness.

Beginning with objective reference (alone) makes no sense, because the conditions that permit it also bring about a unity of self-consciousness. On the other hand, understanding this process provides some insight into why cognition could not begin with self-consciousness (alone) either. How is the relation of necessarily belonging together to be established in the absence of the functioning of an object rule? Most importantly, understanding this process and the mutual dependence of object and subject consciousness it involves tells us why examining the possibility of cognition can lead to the discovery of *a priori* rules or templates for forming '*a posteriori*' rules, rules that can be applied to the data of sense in such a way as to create relations of necessary connection among the representations of a cognitive subject. Neither object cognition nor subject consciousness is possible in the absence of rules, rules that supply standards for objects, and rules whose use creates relation of rational dependence and hence necessary connection across mental states. If *a posteriori* object rules are possible only because they are created on the basis of templates—e.g., a template that requires that properties be attached to objects (and ultimately to substances[20])—then both reference to objects and consciousness of a continuing subject require that cognition be mediated by the use of such (categorial) templates.

Before launching a full-scale exploration of the unity of consciousness, Kant underlines the importance of object rules. Concepts serve as rules. So, for example, the concept of a body necessarily involves representations of impenetrability, shape, etc. (A106). He then explains that a concept can be

> a rule for intuitions only by representing, when appearances are given to us, the necessary reproduction of the manifold and hence the synthetic unity in our consciousness of these appearances. Thus when we perceive something external to us, the concept of body makes necessary the representation of extension, and with it the representations of impenetrability, shape, etc (A106).

The reference to 'reproduction' is somewhat awkward (see below, p. nn) and is dropped in the B deduction. His claim that the materials in the partial representations, 'extended' etc., must be 'reproduced' in the resulting representation 'body' repeats the point that the latter representation must be understood as being built up out of the partial representations, which are 'repeated' in it. That is, he maintains that the representations of 'body' must be understood as necessarily connected in consciousness to the representation of 'extended', etc. Instead of pursuing 'necessary reproduction', he turns somewhat abruptly to the claim that any necessity must be grounded in a transcendental condition, so we need to find a transcendental basis for the 'unity of consciousness in the synthesis of the manifold' (A106.) By a 'transcendental basis,' he means an *a priori* representation that is required for cognition.[21]

Kant then makes a dramatic pronouncement:

> This original and transcendental condition is none other than transcendental apperception (A106-107).

We then come to another *prima facie* curious twist in the argument. Having just introduced a crucial theoretical term, he considers and criticizes the possibility that self-consciousness could arise empirically, through the operation of inner sense.

> There is, in inner perception, a consciousness of oneself in terms of the determinations of one's state. This consciousness of oneself is merely empirical and always mutable; it can give us no constant [*stehendes*] or enduring [*bleibendes*] self in this flow of inner appearances. But what is to be represented **necessarily** as numerically identical cannot be thought as such through empirical data. A condition that is to validate such a transcendental presupposition must be one that precedes all experience and that makes experience itself possible (A107, also cited in Chapter 3).

Introducing the inadequacies of an empirical derivation of the I-representation here serves a number of purposes. Kant's general approach is to clarify classifications by contrasting them with neighboring classifications. Empirical consciousness of states offers an obvious contrast for consciousness of a numerically identical subject; similarly inner sense offers a useful contrast for the faculty of transcendental apperception.

Still the thrust of the passage is less comparative than negative. Hume's and Tetens's failures to provide an empirical derivation of the 'I-representation' provide an opening for Kant's unusual theory that self-consciousness is a requirement for cognition and known through that guise. He appears to exploit that opening in rolling out his theory: Empirical derivations do not establish an enduring or constant self; they could not establish necessary identity. Eberhard had claimed that cognizers were aware of their numerical identity through being aware of constancy [*Stätigheit*] or unbroken continuity in the transition from one state to another. Eberhard could be an object of Kant's criticism. He is a somewhat unlikely candidate, because his general approach is Leibnizean[22] and the target seems to be more straightforward Empiricist accounts. The obvious

foils are Hume, the discoverer of the insufficiency of Empiricist accounts of self-identity, and Tetens who failed to grasp the strength of Hume's demonstration, though the passage may be directed at Eberhard as well.

Kant repeats the criticism that, if one tries an empirical derivation, then one will end up with nothing in more florid language in the B deduction:

> Only because I can comprise the manifold of the representations in one consciousness, do I call them one and all **my** representations. For otherwise I would have as self as many-colored and varied as I have representations that I am conscious of (B134, also cited in part in Chapter 8).

Later in the A deduction, in the 'argument from below,' he considers representations apart from the unity of apperception in terms reminiscent of those in which Tetens presented Hume's view: For Hume there is a crowd [*Menge*] of impressions that follow each other singly [*einzelner*] and are separated [*getheilter*] and scattered [*zerstreut*] and are brought together in a whole subject representation through the faculty of imagination. Kant explains that without relation to apperception

> different perceptions are in themselves encountered in the mind scattered [*zerstreut*] and individually [*einzeln*] (A120).

If the unity of apperception were impossible then

> much [*eine Menge*] empirical consciousness could be found in my mind—yet found as separate [*getrennt*] and without belonging to a consciousness of myself (A122).

He also describes representations as on their own amounting to nothing more than a random 'heap' [*regellose haufen*] (A121), Hume's famously disparaging

term for the impressions of an allegedly single mind, a term that Tetens often uses as well (e.g. 1. 385, 386, 387).

Kant breaks the flow of his positive argument to criticize other approaches for good reason. It gives him the opportunity not just to contrast his view to theirs and, perhaps, to criticize particular predecessors. It also allows him to showcase the superiority of scientific metaphysics in general, by arguing that Empiricists cannot explain the obvious fact that there are continuing cognizers. Rationalist approaches will be shown to be inadequate in the Paralogisms Chapter, thereby completing his case that, in this crucial area, only transcendental philosophy can provide a viable solution. Passages like those cited above led me to suggest in earlier work (1990, Chapter 4 *passim*) that a central goal of the argument for apperception was to defeat Hume's skepticism about personal identity. Given the importance of the unity of apperception to the proof of the categories, which Kant partly saw in the *Duisburg Nachlaß*, prior to reading Tetens, that claim is too strong. Still, the shape of the A deduction demonstrates that he does not introduce the theory of apperception only to establish the categories. It is not just a lemma for that proof, but an important result in its own right, because it offers a way of dealing with the unity of self-consciousness that avoids both skepticism and dogmatism. His theory of apperception has the twin advantages of resolving the contemporary puzzle about the unity of self-consciousness and of providing a bridge to the long-sought argument for the categories.

Having explained how he is about to solve the problem of the human representation of unity of consciousness (by revealing it to be a presupposition of experience or empirical cognition) Kant presents his solution:

> Now there can take place in us no cognition, and no connection and unity of cognitions among one another, without that unity of consciousness which precedes all data of intuitions, and by reference to which all representations of objects is alone possible. Now this pure, original, and immutable consciousness I shall call **transcendental apperception** …[23] Now this transcendental unity of apperception brings about, from all possible appearances whatever that can be together in one experience, a coherence of all these representation according to laws. For this unity of consciousness would be impossible if the mind, in cognizing the manifold, could not be conscious of the identity of function whereby it synthetically combines the manifold in one cognition (A107-108).

What does he mean by the claim that the unity of consciousness precedes all data of intuitions? One possibility is that he is agreeing with Crusius and Merian against Wolff that self-consciousness must precede any awareness of objects that would be provided through the data of intuitions. As Thiel notes, the use of the term 'original' [*ursprüngliche*] may be a reference to Merian's position. Still, the last sentence makes this hypothesis problematic. If the unity of consciousness literally 'precedes' all data of intuition, then how could it be impossible except under certain conditions? As we saw in the last chapter, Kant says in a number of passages that the unity of apperception must be achieved, that representations must be brought to or under the unity of apperception.

We can resolve the apparent inconsistency between Kant's claims that the unity of consciousness both precedes cognition and is created in the course of

the mind's cognizing activities by recalling that the representation of unity is *a priori* and by turning to his clarification of *a priori* representations in the Reply to Eberhard. When criticized for appealing to innate representations, he explained more precisely how his claims about *a priori* representations should be understood.

> The *Critique* admits absolutely no implanted or innate **representations.** One and all, whether they belong to intuition or to concepts of the understanding, it considers them as **acquired**. But there is also an original [*ursprüngliche*] acquisition (as the teachers of natural right call it), and thus of that which previously did not yet exist at all, and so did not belong to anything prior to this act. According to the *Critique*, these are, **in the first place**, the form of things in space and time, **second**, the synthetic unity of the manifold in concepts; for neither of these does our cognitive faculty get from objects … rather it brings them about **a priori**, out of itself. There must indeed be a ground for it in the subject, however, which makes it possible that these representations can arise in this and no other manner, and be related to objects which are not yet given, and this ground at least is **innate**. …The ground of the possibility of sensory intuition is neither of the two, neither **limit** of the cognitive faculty nor **image** [as Eberhard had suggested]; it is the mere **receptivity** peculiar to the mind, when it is affected by something (in sensation) to receive a representation in accordance with its subjective constitution … [T]ranscendental concepts of the understanding … are acquired, and not innate, though their acquisition, like that of space is no less original and presupposes nothing innate save the subjective conditions of the spontaneity of thought (in conformity to the unity of apperception) (8.221-223, C1781, 312-313, my underscoring.)

On his view there are no innate representations of spaces, causes, or anything else, presumably including 'I's.' Rather, there are ways of arranging sensory data and of forming concepts that depend on innate tendencies of the mind's

receptive and active faculties. In the latter case, the tendencies order conceptual representations in such a way that representations of both categorial concepts and the unity of apperception arise in consciousness.

Presumably the innate ground of the original acquisition of causal representations is the tendency to scrutinize the data of sense for patterns that could be instances of causal relations.[24] According to the passage, this tendency is governed by what appears to be a further consideration. The representations so formed are subject to the unity of apperception. This suggests an extra layer of scrutiny. Cognizers not only come with innate tendencies to scrutinize the contents of representations for indications of causal relations; they also have innate tendencies to scrutinize not the contents of representations, but representations themselves to find representations that meet some other principle. Given that the 'I-think' is *a priori* and that *a priori* representations and principles arise through the [scrutinizing] activities of the mind (B1), this must be his view. If so, it resolves the 'prior to' and 'produced by' problem. The tendency to scrutinize by a principle associated with the 'I-think' precedes cognition; the representation of the unity of consciousness would be produced by that scrutiny.

Still, this resolution raises two obvious questions: Where to cognizers get the idea of 'representation' *per se*? By what principle do they scrutinize representations? Kant's official answer to the first question is straightforward and disappointing. Humans are aware of the world around them through outer sense and aware of the condition of their 'inner world' through inner sense (A22/B37). This answer is not only disappointing, but also in tension with the picture of

judging presented through the counting example. Mathematics in general and counting in particular are taken to require some sort of sensory representations (real or imagined stroke symbols), but the faculty that judges '4' on the basis of adding up the stroke symbols one after another is not inner sense but apperception. By the time of the *Critique*, inner sense is no longer the faculty that enables cognizers to know the basis of their judgments. Apperception performs that crucial function. But Kant does not make any effort to reconcile his new view with his continuing claim that cognizers know about their mental states through inner sense. As I suggest below, however, some type of reconciliation might be possible.

The answer to second question is more helpful. For reasons we have seen, the principle according to which representations are scrutinized cannot be the apperceptive principle, because that principle states that the unity of apperception is necessary for cognition of objects. The principle in question must be the one embedded in the principle, the I-rule that 'all representations must belong to a single consciousness.' This rule is also in tension with the inner sense theory, since, for all we know, inner sense might divulge a condition of the inner world that could not be understood as belonging with others to a single consciousness—and hence could not be a representation. On the other hand, the rule indicates a possible avenue for representing representations as such apart from inner sense. In rational cognition, some states that are representations (judgments) come to be understood as depending on other states that are representations (partial representations that are combined in the

whole). It is not possible to represent that dependency or that necessary togetherness without representing the states as such. So this capacity must come along with the capacity for rational cognition. Unlike the case of causation, the unity of apperception is produced through the creation of relations of necessary connection across representational states—and the states thereby come to be recognized as such. More briefly, the activity of thinking produces relations of necessary connection across representations, which thanks to scrutiny *via* the I-rule, enables the cognizer to represent her representations as such and as states of an 'I.'[25] Above I characterized the inner sense theory as Kant's 'official' account of humans' cognition of their representations, but the view just outlined is at least a semi-official second view.

The discussion of A107 concludes:

> Hence the original and necessary consciousness of one's own identity is at the same time a consciousness of an equally necessary unity of the synthesis of all appearances according to concepts, i.e., according to rules (A108).

How can humans be conscious of representations as necessarily belonging to a single self-consciousness? Besides having faculties capable of synthesizing or combining materials from different representations, they must employ concepts or rules in the manner exemplified by counting or by the use of the concept 'body' and thereby bring about a relation between the partial cognitions and the whole cognition or concept application (or judgment). Through the consciousness of the synthesis they are conscious of the relation between, for example, the representation of the first stroke symbol as '1' and the judgment '4.' They can

see these representations as exemplifying the relation of necessarily belonging together, since the judgment '4' would be impossible without the representation of the '1,' and the representation of '1' would be impossible as a ground of cognition without the judgment.[26] In recognizing the relation of necessarily belonging together between these representations, they recognize them as exemplifying the I-rule.

We can now see the contours of the central argument of the A deduction: It is necessary for RE cognition that representations can be understood as necessarily belonging together. We can also see why Kant believed that the unity of apperception and object cognition (under concepts) are necessary and sufficient conditions for each other and why his position is plausible—and may be correct. RE object cognition requires partial cognitions that are recognized as such and so an implicit awareness of the necessary togetherness relation across different states. Further, given the absence of any intuition of an 'I' and the impossibility of establishing relations of necessary connection empirically, he sees no way for cognizers to be aware of their representations as belonging to the unity of consciousness—as necessarily belonging together—except by being aware of a relation between partial cognitions and a whole cognition that suffices for object cognition.

Whether or not this was his intent, Kant's theory offers an elegant solution to the self-consciousness versus object-consciousness priority argument between Wolff and his critics. His theory splits the difference between them while criticizing both for not seeing the intimate connection between object

cognition and the unity of consciousness. The critics were correct that there is no cognition worthy of the name prior to self-consciousness and that it is impossible to derive a representation of a unified subject from object representations. It is a *sui generis* representation. On the other hand, Wolff was right that Descartes' scenario of an 'I-think' without any object of thought is impossible.[27]

Once the relation between RE cognition and the possibility of a unity of self-consciousness is clear, we can see how an argument that starts by exploring the necessary conditions for the former can end with a claim about the application of the categories to any object that can affect the human senses (B160). An important part of the argument concerns the necessity of ordering human mental states in time, so that they can have the intuitive coherence required by the sorts of (spatiotemporal) that cognizers humans are. But the general argument that is meant to cover any discursive cognizer, regardless of form of intuition, already provides the solution. The deduction is able to perform the apparently logic-defying feat of arguing from a narrower class (representations that belong to cognition) to a conclusion about *prima facie* wider class (any representation that can belong to the unity of consciousness) because of the peculiarities of the I-representation.

Different representations do not manifest a constant 'I' representation in inner sense; there is no I-impression or I-intuition. The I-representation is *a priori*. As such, it is tied to a principle, in this case, the I-rule that all representations must belong to a common 'I.' But how can the necessary

connections across representations be forged (and recognized)? In Kant's view, the only possibility is through the rule-governed activity of thinking that produces object cognition. Through that activity, the unity of consciousness is created and recognized: Some representations are, and are understood as, rationally dependent on others; the latter representations (intuitive or conceptual) are participants in cognition (and so representations) and are understood as such only through their relation as grounds for the former. Absent this relation none of the representations, conceptual or intuitive, could be referred to a single subject. In particular, intuitions that could not participate in cognition could not be attributed to a common subject. If the object rules needed for rational cognition are built on *a priori* templates, then it follows that any representation that could be understood as such—as the representation of a common subject of many representations—must be suitable for the application of categorial concepts. Kant summarizes these connections in a letter to Herz:

> If we can demonstrate that our knowledge of things, even experience itself, is only possible under those conditions [the forms of intuition and the categories of the understanding], it follows that all other concepts of things … are for us empty and utterly useless for knowledge. But not only that; all sense data for a possible cognition would never, without those conditions, represent objects. <u>They would not even reach that unity of consciousness that is necessary for knowledge of myself (as object of inner sense). I would not even be able to know that I have sense data; consequently for me, as a knowing being they would be absolutely nothing</u> (11.51-52, CLetters 314, my underscoring).

Unless a subject engages in object cognition she can have no grasp of a sensory state as belonging to the unity of consciousness and so no understanding of her sensory states as sensory states or as belonging to a common subject.[28]

After arguing that the consciousness of a subject's identity is at the same time a consciousness of the synthesis of appearances according to concepts or rules at A107, Kant seems to shift to a different reason for connecting representations to a common 'I,' a reason that concerns a common act:

> … For the mind could not possibly think its own identity in the manifoldness of its representations, and moreover think this identity a priori, if it did not have the identity of its act before its eyes [*vor Augen hätte*]—the identity that brings all synthesis of apprehension under a transcendental unity, and thereby first makes possible the coherence of those representations according to a priori rules (A108).

What does Kant mean by the 'identity of the act'? One possibility is suggested by the counting: There would be one extended act of adding the units, '1' etc. The idea would be that the singleness of the act of generation connects the representations so generated to a single self.

Carl stresses the importance of the generation of numbers in a particular order in the counting example. In his view, the purpose of the example is to illustrate how a concept can determine the order of states in intuition (1992, 163). He also sees the generation by a single (extended) act as the source of the belongingness to a common subject.[29] Although the demonstration of ordering states in counting is very important to Kant's project of explaining how the categories can determine the order of states in inner sense, that cannot be all

that is intended. His central point is that without being aware of his acts of counting, the would-be counter could not cognize the total amount. The counter needs not just to apply the rule and so to produce representations in a particular order; he needs to grasp the relation between his partial representations and his judgment '4.' Exactly the same point applies to the idea of a single act as producing the varied actions of counting. Unless that act includes the final act of recognizing that the representation '4' depends on the results of earlier phases of the (extended) act, the would-be counter will not have cognized the amount. This point harkens back to the original definition of synthesis, where a cognizer not only adds various representations together, but also comprehends or understands their manifoldness in one [resulting] cognition. The crucial aspect of the synthetic process is not so much a common act or process of synthesizing, but the relation between the synthesized representation and the representations from which it is synthesized in accord with a rule. Bringing some object x under the concept 'body' or judging that 'x is a body' requires that the cognizer grasp that her judgment depends on a further, partial representation 'x is extended.' This 'recognition in a concept' according to a rule permits the object to be rationally cognized. It also demonstrates that the different representations stand in the relation required for being states of a single 'I.'

We are now better able to appreciate the differences between Gareth Evans's discussion of the identity of a person in tracking an object and the Kantian analysis of my keeping track of our puppy Teddy sketched in Chapter 1. For Evans, the identity of the person through the tracking is a function of two

features: The tracking activity is the expression of a belief and the belief can be criterionlessly attributed to the subject. On the Kantian account of keeping an eye on Teddy, I judge that Teddy is still in the room, by noting that he is at a particular place in the room. My judgment is rational, because I know its grounds, and because I know its grounds I take the two states, that of seeing Teddy under the window, and that of judging him to be here, to be necessarily connected together and so to exemplify the rule for being states of a common 'I.'[30] As in the case of an extended act, a temporally extended exercise of a disposition would not, by itself, produce either rational object cognition or necessary connection across the temporally diverse states. On Kant's view, for either to occur, the cognizer must possess some rule that enables her to grasp the logical relations across the contents of her representations. So the approaches are very different.[31]

The importance of rules to object cognition and to thinking one's own identity is why Kant begins and ends the long and difficult A108 paragraph by talking about the necessity of laws or *a priori* rules. But, if the central point of A108 is that the concepts and rules are needed to permit the unity of consciousness, why does Kant emphasize the 'identity' of the act? He claims that a subject would be unable to understand his self-identity if he were not aware of the identity of the act that brings all synthesis of apprehension under a transcendental unity. The first question to ask is whether the identity in question is numerical or qualitative. As the basis of a claim of the numerical identity of the self, it seems that it must be a numerically identical act. On the other hand, it is

implausible to think that different representations are brought to the unity of apperception by a single act that extends for a lifetime—or at least implausible to think that that is Kant's view.[32]

It may be easier to understand A108 by considering a related observation from the B deduction:

> We readily become aware that this act of synthesis must be in origin single [*ursprünglich einig*] and equally valid (or of equal weight [*gleichgeltend*]) for all combination (B130, amended translation).[33]

This passage tilts towards a reading of qualitative identity. Although '*einig*' indicates singularity or perhaps uniqueness, what are we to make of the modifier '*ursprünglich*' which means 'in origin' or 'originally'? The possibility that the act was originally single and later something else seems to make no sense. Further, Kant uses this term to indicate a special sort of origin in the activities of the faculties. Although the question is vexed[34], I think the most plausible reading is that the act has a singular origin in the distinctive activity of the thinking faculty.[35] The additional and unusual characterization, 'equally valid,' supports that interpretation. Because the act has its origin in the characteristic activity of the thinking faculty, it always has equal weight or is equally valid.

On Kant's theory, representations are attached to a common 'I' through the operation of synthesis:

> This synthesis is called the original synthetic unity of apperception. All representations given to me are subject to this unity; but they must also be brought under [*gebracht unter*] it through a synthesis (B135-36).

Yet, as we have seen, synthesis is an operation carried out on some representations that yields other representations. It is carried out in accord with rules that are attached to different concepts—the rule for 'body,' for '4,' for 'dog' etc. So how can an operation that is carried out according to different rules produce a univocal relation of belonging to a common subject across all representations? Kant's preemptive reply to this objection is to stress the 'identity' of the act or 'singular character' or its 'equal weight': Even if different representations are combined in accord with a variety of rules, there is no difference in the act of combining itself and hence no difference in the relation of necessary connection created across representations. Representations of self-propelled motions and of certain shapes might be combined in the representation 'dog', and representations of 'extension' and 'impenetrability' may be combined in the representation 'body,' but the act of combining and the relation of 'being combined in' are the same. That 'synthesis' is in origin singular and equally valid for all acts of combination is not sufficient to connect the subject who judges 'dog' to she who judges 'body,' but it is a necessary condition for doing so. As we see below, the equal validity of all acts of synthesis is also crucial in explaining how different sorts of mental states, perceptions as well as conceptions, can be understood as standing in the same relation to each other and so to a common 'I-think.'

Although the single extended act of counting does not explain the unity of the representations in a single consciousness, it does explain something very important. Intending to apply the counting rule, the cognizer starts with real or

imagined stroke symbols and assigns them number ‘1’ etc. The singe act explains the transition from one representation to another. No Leibnizean inner principle unfolds the career of a substance; a cognizer intends to carry out an act of cognition. Besides setting up relations of ground and whole cognition across all the representations—and so the relation of necessarily belonging together—the example illustrates how the order of representations in inner sense can be determined by rules (as Carl helpfully points out) thereby also illustrating the transition of a mind from one state to another.

The counting example also shows how the ‘chicken-and-egg’ problem that divided Kant’s predecessors can be solved. It is true that there is no RE cognition without self-consciousness and true that there is no self-consciousness without representations of objects. The solution can be understood in terms of the simultaneous application of two rules, the counting rule and the I-rule. It is a mistake to model the situation as one in which the counter gets to ‘4’ and through his consciousness of his act of synthesis then realizes (through a further act of synthesis) that the states representing ‘1’ etc. and ‘4’ stand in a relation of necessarily belonging together and so are instances of the I-rule.[36] He applies both the counting rule and the I-rule prospectively. He starts counting and building up partial cognitions in the expectation that the partial representations will enable him to determine the number and that the judgment will be based on the partial representations produced by counting. In being aware of the mental act of judging ‘4’ on the basis of the partial cognitions, he knows both the amount

and relation of necessarily belonging together across his partial representations and the judgmental state.

Towards the end of the A deduction, Kant returns to the representation of a constant and enduring 'I' that eluded the Empiricists:

> For in this constant and enduring I (of pure apperception) consists the correlate of all our representations insofar as becoming conscious of them is so much as possible. And all consciousness belongs to an all-encompassing pure apperception, just as all sensible intuitions belongs, as representation … to time (A124).

Since 'I-think' is an *a priori* representation, the I-rule should work in the same way as the categorial principles, prodding the understanding to scrutinize the perceptions available to inner sense in search for representations that are instances of it. What is different is that since the only representations that the understanding takes note of are those that fit categorial rules, these representations will automatically also fit the I-rule. (In this way, the 'extra scrutiny' by the I-rule considered above, would not eliminate any representations, but only allow for the application of 'common subject' across them.) Hence the 'I' will be a ubiquitous feature of representations of which the subject is conscious.

The A deduction somewhat muddies its message of necessary unity by considering the possibility of 'representations' that do not belong to the unity of apperception. One frequently discussed passage dismisses such 'representations' as 'less than a dream' (A112). A second passage raises the possibility of a multitude of perceptions—empirical consciousness—arising in a human mind, but without belonging to the unity of consciousness (A122). Here

such ‘perceptions’ are said to be impossible. Insofar as ‘representations’ or ‘perceptions’ are meant to indicate those mental states that enable humans to have their distinctive form or cognition, however, these occurrences of the terms are misuses. Alternatively, Hoppe is right that these passages are not intended to show that such situations are possible—but rather impossible (Chapter 8, p. nn). The B deduction refrains from such speculation, making it clearer from the beginning that the representations at issue are representations properly so-called, representations that are something for a cognitive subject.

Before turning to the somewhat different way in which the argument is cast in the B deduction, it may be useful to summarize the main points in the A edition argument for apperception. As I understand it, there are 6 essential steps:

1. Cognition requires intuitions and concepts (e.g., ‘body’).
2. Intuitions can participate in cognition only insofar as they are understood as providing the basis for concept use (they are the states from which the concept was abstracted) and concept application (they are the basis of concept application), hence, only insofar as they are understood as standing in relations of necessary connection to representational states of applying concepts.
3. The use of concepts requires implicit use of rules (e.g., a rule governing ‘body’, ‘bodies are extended’) or the general rule of similarity for simple concepts or concepts used in the manner of simple concepts.

4. In making implicit use of the rules associated with concepts, the cognizer must recognize some of his representational states (e.g., that representing 'body') as depending on others (e.g., that representing 'extended") and so must recognize the states as standing in relations of necessary connection.

5. When representational states are recognized as necessarily connected together, they are recognized as instances of the I-rule.

6. Thus any representational state that can be understood as participating in cognition must also be understood as meeting a necessary condition for belonging to a common 'I.'

6. Combination and Self-Consciousness in the B Deduction

Given the hint from the *Metaphysical Foundations of Natural Science*, we might expect the B deduction to begin with an exact definition of 'judgment.' It does not. Kant simply assumes that cognition must involve combination. He needn't spell out why judgment requires combination, because judgments were standardly taken to involve a relation between two concepts (B140-41). Further, cognition requires both intuitions and concepts and concepts, like judgments, require combination.

The scope of the opening discussion of the B deduction is broader than concept application and judgment. Kant does not consider combination just as a prerequisite for (all kinds of) judging. His topic is all combination or synthesis of representations, whether the combination is conscious or unconscious and

whether or not it is carried out on concepts or on sensible or non-sensible intuitions. Again, he does not spell out why intuitions must be combined. He can, however, presume familiarity with his initial discussion of concepts in the Analytic where he presented them as 'functions of unity' among intuitions (and other concepts). A concept is something that unites different intuitive (or conceptual) representations under it (A68/B93). Hence insofar as intuitions are understood as being brought under or united under concepts—which they must be to participate in cognition—they too need to be thought of as combined under a conceptual representation that encompasses them all.

Because combination is essential to judgment and concept use, and so too to cognition, Kant proposes to investigate the necessary conditions for combination as such. He begins by implicitly comparing the sort of combination he has in mind with that offered by Empiricism in its laws of association:

> We cannot represent anything as combined in the object without ourselves' having combined it beforehand; and that, among all representations **combination** is the only one that cannot be given through objects, but—being an act of the subject's self-activity—can be performed only by the subject himself (B130).

According to Empiricism, representations simply associate themselves and so form combined representations. His claim is that the sort of combination relevant to cognition requires that combination not simply be effected in some manner, but that it is represented as such, as combination. Against the background of his views of rational cognition, we can understand this terse observation as presupposing that in proper cognition, the cognizer knows the ground or basis of,

for example, his judgments. In that case, however, he must understand his judgment, 'x is a body' as combined out of partial representations, e.g. 'x is extended.' Further, in proper cognition, a cognizer understands a concept as such—as something that unites different intuitions (or concepts) under it. In which case, he must also understand his intuitions as combined in a certain sense under concepts.

Having specified the sort of combination relevant to cognition, Kant offers a necessary condition for it:

> But the concept of combination carries with it, besides the concept of a manifold and of its synthesis, also the concept of the manifold's unity. Combination is a representation of the **synthetic** unity of the manifold[n] (B130-31, my underscoring).

He goes on to deny that the unity involved is the category of unity (B131), whose principle is that all intuitions are extensive magnitudes. Rather the unity is qualitative. He explains further that the use of categorial concepts involves combination. Since the categories thus presuppose the unity in question, they cannot be its source:

> We must therefore search for this unity (which is qualitative unity; see section §12) still higher up; *viz*., in what itself contains the basis for the unity of different concepts in judgments, and hence contains the basis for the possibility of understanding, even as used logically (B131).

When read on its own, the claim that combination is the representation of the manifold's unity is mysterious. What is this unity? Is it the unity of the materials (suggested by the phrase 'the unity of the manifold') or is it a unity of the faculty of understanding that combines them? Fortunately Kant appends a

note that dispels much of the mystery, particularly when read in light of his views about rational cognition. He is specifically addressing the case of analytic judgments.

> Whether the representations themselves are identical, and thus one could be thought analytically through the other; that does not come into consideration. Insofar as we are speaking about the manifold [of representations, i.e., the different representations] the **consciousness** of the one is nevertheless always different from the consciousness of the other, and we are here only concerned with the synthesis of this (possible) consciousness (B131n).

Even in the case of an analytic judgment, e.g. 'F=F', where the contents are identical, combination is required. For a cognizer to be conscious that F is F, that the representations are identical, she must be conscious of one F and conscious of the other, and the states of representing the common content 'F' are different. To make the judgment, the cognizer must combine or synthesize the materials in the diverse states. Alternatively, to understand her analytic judgment as a judgment, she must understand the judgmental state as being based on the states of representing, say, the left-hand-side F and the right-hand-side F. As we saw in the discussion of the A deduction, however, when a cognizer grasps some his states as dependent on others and so as necessarily connected, he simultaneously grasps them as instances of the I-rule. Thus the unity that is presupposed when combination can be understood as such is the unity of consciousness.

In arguing against taking the *cogito* as the premise of the deduction, I noted that the discussion of combination as such serves as a prologue to the

introduction of the 'I-think' doctrine. Upon closer inspection, we can see that it lays out a rich background and even the beginnings of an argument (in the note) for the claim,[37] which I repeat for easy reference:

> The **I think** must be **capable** of accompanying all my representations; for otherwise something would be represented in [*in*] me that could not be thought at all, which is equivalent to saying: the representation either would be impossible, or at least would be nothing to me. Representation that can be given prior to all thought is called **intuition**. Hence everything manifold in intuition has a necessary reference to the **I think** in the same subject in whom this manifold is found (B131-32, amended translation and punctuation.).

The second portion of the claim (after the semi-colon) modifies the first. As in A, the scope of the 'I-think' doctrine in the B edition is the set of representations that can participate in cognition. For an intuition to so participate, it must be thought. That is, it must be able to be brought under concepts in a thought or judgment (about an object, as we learn slightly later). But to bring an intuition under a concept is to combine it with others under that concept. Hence in so far as any intuition is understood as something that can be part of cognition, it must be understood as something that is combined with other representations in a resultant representation. For example, the concept 'orange' can be used in rational cognition only when it is understood as abstracted from various intuitions which are 'combined' in it or collected under it. And to understand intuitions as combined in concepts is to see the intuitive and conceptual representations as necessarily connected to each other and so as meeting the necessary condition

for belonging to a single I-think. Given the necessity involved, Kant observes that

> the representation [I-think] cannot be regarded as belonging to sensibility (B132).

As noted in section 4, the argument is clearer if we are cognizant of the double dose of necessity. Why an Empiricist account of the 'I-representation' would be unavailing (even if it could be given, which it can't) is that it is necessary for cognition that different mental states are understood as combined and so as standing in relations of necessary connection. And no Empirical account could establish necessary connection.

Although Kant does not present it in this guise, the preceding argument undercuts Wolff's view that object cognition precedes self-consciousness. The claim asserted and defended is that any representation (that could serve in object cognition) must also be understood as belonging to a single self-consciousness. After noting that the representation 'I-think' must be spontaneous and *a priori*, he goes on to discuss the possible ramifications of the argument. If any representation that can serve in cognition must be understood as belonging with others to a common I-think then it must meet the requirements for doing so and those requirements could turn out to be significant.

Kant then turns to the problem of how cognizers are conscious of the identity of apperception. One target of his implicit criticism seems to be Locke's Empiricist account. Locke's 'reflection' theory could not explain how a cognizer is conscious of her identity across different representational states:

> The same thoroughgoing identity of the apperception of a manifold given in intuition contains a synthesis of representations, and is possible only through the consciousness of this synthesis. For the empirical consciousness that accompanies different representation is intrinsically sporadic and without any reference to the subject's identity. Hence this reference comes about not through my merely accompanying each representation with consciousness, but through my adding one representation to another and being conscious of their synthesis. Hence only because I can combine a manifold of given representations **in one** consciousness, is it possible for me to represent the **identity of the consciousness itself in these representations** (B133, my underscoring).

The objection is that the Lockean consciousness—the consciousness that is inseparable from thinking (Locke, 1690, 2.27.9, 335)—is momentary or episodic. As such, it cannot provide a basis for representing a common subject. In this passage, he makes the same objection that Merian raised against Wolff: A 'reflecting' consciousness can provide no evidence of a *continuing* self[38] .

Although Kant's criticism agrees with Merian's objection to Wolff, it applies with equal force to his and to Crusius's competing proposal that self-consciousness is prior to the differentiation of objects. The opening sentence of the passage claims that the identity of the apperception of different states contains a synthesis and is possible only through a consciousness of this synthesis. Later in the passage he claims that to be conscious of her identity, a cognizer must add representations to each other—must combine them—and be conscious of the combining or synthesizing.

Chapter 8 dealt with some problems in Kant's view that synthesis can be both a conscious and an unconscious operation. The set-up of the B deduction raises an additional puzzle. He stresses that he is going to consider combination in general, whatever the nature of the materials or the conscious or unconscious character of the operation. But at B 133, he says that consciousness of the act of combining is essential for linking different representations to a common 'I-think.' His point is that, even though some, indeed many, syntheses are unconscious, there is still something special about conscious syntheses. If a cognizer were never conscious of synthesizing, then she could not be self-conscious: She could not be conscious of herself as the possessor of different states. And that is because she could never recognize any of her states as standing in relations of necessary connection. The counting example in the A deduction makes Kant's claim more vivid and so easier to grasp. Through the conscious act of combining partial representations according to a rule, a cognizer can both cognize the amount and the relation of necessary connection across his states. In B, the argument is completely abstract: A cognizer can recognize that different representations belong to a single consciousness only through a conscious act of combining materials from some representations in a resultant representation. Through the conscious act of combining, he can both grasp the contents of the resultant state and the relation of necessary connection across the states which contribute to and result from the combining and which is essential to the state being the 'proper' cognitive state that it is. Despite the context of a broad notion of combination (of which more below), conscious

combination thus plays the same role in B that it did in A. Consciousness of synthesis informs cognizers of the special relation that holds across their representations in thinking and thus of the necessary connection of those representations with one another.

One aspect of Kant's objection to Locke can be applied to his theory as well.[39] He complains that Locke cannot explain the reference of different representations to a common 'I,' but he has given no account of how a representation can be referred to an 'I' at all. Presumably he thinks (wrongly, as we have seen) that the 'mineness' question can be resolved through inner sense: A representation is mine' just in case I am aware of it *via* inner sense. By contrast, he notes in the B edition as well as in A that inner sense can provide no representation of a common self or constant 'I' across varied representations (B134), thereby completing his argument that conscious combining is necessary for recognizing a relation of necessary togetherness of different representations in a single 'I.' Nevertheless, he has not provided an adequate account of how cognizers refer to themselves as particular 'I's either.

Kant goes on to observe that asserting that different representations belong to a common subject presupposes a synthesis. Whoever makes the claim is connecting a variety of representations in a representation of an 'I' that has them all: the 'I' with state R_1 and R_2, R_3, etc.

> The thought that these representations given in intuition belong one and all to me is, accordingly, tantamount to the thought that I unite them, or at least can unite them, in one self-consciousness. And although that thought itself is not [*noch nicht*] the

> consciousness of the **synthesis** of representations, it still presupposes the possibility of that synthesis. I.e., only because I can comprehend the manifold of the representations in one consciousness, do I call them one and all **my** representations. For otherwise I would have a self as many-colored and varied as I have representations of which I am conscious (B133-34, my underscoring, amended translation.)

We have to be careful about what is being combined with what. My gloss above can mislead. Kant's point is precisely not that someone who asserts that I am the possessor of R_1, R_2, and R_3 manages to make that claim by combining R_1 with a representation of an 'I,' R_2 with a representation of 'I' etc. This is not possible, because inner sense provides no 'I' representation to combine with others, as he notes in the next paragraph:

> Through the 'I,' as a simple representation, nothing manifold is given; only in intuition which is distinct from this representation, can a manifold be given (B135).

Different representations are not combined with some permanent 'I' intuition. Instead, they are combined with each other and that can come about only through an operation of synthesis that establishes their necessary connection to each other—and so their conformity to the condition for belonging to one consciousness or an 'I-think.' Combination of representations is thus presupposed in any assertion of coconsciousness. It is presupposed because, absent an intuition of an 'I,' representations can be connected to a common 'I-think' only by being understood as necessarily connected to each other and Kant sees no alternative for meeting that condition other than by grasping some representations as combined in others.

The ‘I-think’ is a very odd representation, because its use depends on the possibility of combining other representations. Its peculiar status is also highlighted in a note appended to the passage about the necessity of being conscious of combining representations to self-consciousness. For easy reference, I repeat the passage to which the note is appended.

> Hence only because I can combine a manifold of given representations **in one** consciousness, is it possible for me to represent the **identity of the consciousness itself in these representations**. I.e., the **analytic** unity of consciousness is possible only under the presupposition of some **synthetic** unity of apperception[n] (B133).

By the ‘analytic unity of consciousness’, Kant means that different representations, R_1, R_2, etc., are attached to the same representation (though not intuition) ‘I’: ‘I have R_1,’ ‘I have R_2,’ etc. This claim is related to the analytic truth that any representation which belongs to the same ‘I’ as others must meet whatever conditions are necessary for doing so, but in the note, he takes the argument in a different direction:

> The analytic unity of consciousness attaches to all concepts that are, and inasmuch as they are, common [to several representation]. E.g. in thinking **red** as such, I represent a property that can be found (as a mark [*merkmale*]) in something or other, or can be combined with other representations; hence only by virtue of a possible synthetic unity that I think beforehand can I represent the analytic unity. A representation that is to be thought as common to different representations is regarded as belonging to representations that, besides having it, also have something **different**, about them. Consequently it must beforehand be thought in synthetic unity with other representations (even if only possible ones). Only then can I think in it the analytic unity of consciousness that makes the representation a **conceptus communis**. And thus

> the synthetic unity of apperception is the highest point, to which we must attach all use of the understanding, even the whole of logic, and in accordance with it transcendental philosophy; indeed, this power is the understanding itself (B133-34n, amended translation.)

The analytic unity of consciousness is unique in that it is not required just to use a common representation 'I.' It is also required by the use of any concepts where the concept is thought of as common to different (complex) representations in which it is contained. Consider a case where a cognizer recognizes that a representation, say 'red,' is common to many representations, 'cardinal,' 'fire engine,' 'tomato', etc. She can do so only if she understands her 'abstraction' of the common element as depending on her consciousness of the (different) complex representations. To do that she must grasp that her representation of the abstracted concept, 'red,' and the states representing the complexes as necessarily connected and so as belonging to a common 'I.' Further, to represent the complexes as such, she must see them as having been produced through the combination of partial representation (and hence as themselves presupposing a common 'I'). Since the 'I' must be common, the use of any concept thought of as common to many representations also requires the use of 'I' and so the analytic unity of consciousness. Since Kant believes that all concepts must be understood as representations that are (potentially) common to many representations, he draws the conclusion that all concept use requires the synthetic unity of consciousness.

An important passage in the Paralogisms Chapter (A341/B399) describes the 'I-think' as the 'vehicle' of all concepts and hence of the categories. It is

necessary for their use, because that use involves combination. In this note, Kant argues that the 'I-think' is the vehicle for all concept use (and so cognition). Were cognizers unable to recognize their states as necessarily connected they could not recognize their concepts as common to many different representations, because they could not recognize complex representations as combinations. In that case, however, they would be incapable of the proper use of concepts. So the 'I-think' is a peculiar sort of representation in that its use both makes possible the use of concepts that are combined with others and also depends on the possibility of combining representations with others.

The crucial 'I-think' section (§16) of the B deduction concludes by reiterating the A edition's insistence that the unity of apperception is not only required for the combination presupposed by rational cognition; the unity of apperception is also made possible through the synthesis of representations. The final paragraph of the section makes the claim twice. It does so first in summing up the argument considered above about conscious combination, combination, and referring different representations to a common 'I':

> Hence the synthetic unity of the manifold of intuitions, as given a priori, is the ground [*Grund*] of the identity of apperception which precedes[40] a priori all my determinate thought (B134-35, amended translation).

The final thought of the section repeats its crucial lesson:

> All representations given to me are subject to this unity [the original synthetic unity of apperception]; but they also must be brought under it through a synthesis (B136-37)

Given the prominence of this doctrine, I disagree with Carl's suggestion that Kant doesn't mean to claim that representations are brought to the unity of apperception, but only that they must belong to a single self-consciousness.

In Chapter 7, I suggested that because cognition is complicated, an argument that examines the possibility of RE cognition will have many strands. We have just seen a number in the crucial opening sections of the B deduction (§15 and §16), where considerations about the requirements of judgments, of concept use, of intuitions that can participate in cognition, and even of making claims about representations as belonging to oneself are explored and inter-related. But the main line of argument in the B edition is not very different from that of A. As I understand it, the A deduction argument is that both recognition in a concept (and so object cognition) and self-consciousness require the simultaneous operation of two rules, an object-rule and the I-rule. The focus of the B deduction is on combination. Both cognition and self-consciousness require combination that is understood as such. And for combination to be understood as such, both combination of the materials of representations in further representations and combination of representations in a subject are required. Despite the differences in presentation, the arguments for the unity of apperception in the two editions are thus surprisingly similar, especially when the next section of the B edition (§17) explains that cognition requires a particular sort of combination—that of different representations in an object in accord with the concept of an object. The deductions are similar because they develop the

same twin themes of the lack of an intuition of an 'I' and of the mutual dependence of RE cognition and self-consciousness.

In the next three sections of the B deduction, Kant draws some conclusions and turns explicitly the issue of judgment and the unity of apperception. Section 17 repeats the A edition claim that the principle of the original synthetic unity of apperception is the highest principle of the understanding (B136-37). It also explains that the ultimate combination of representations required for cognition is in the representation of an object—i.e., according to the concept of an object—thus repeating the doctrine of A105 already discussed. Section 18 concerns the subjective unity of consciousness and will be discussed below in conjunction with the relation between transcendental and empirical apperception.

We have already briefly considered the section about judgment and apperception (§19) in looking at the proposal of a 'one-step' deduction from the exact definition of judgment. This discussion is interesting in part, because it resolves a puzzle about the differences between the arguments of the A and B editions. The introduction of 'transcendental apperception' in A is intertwined with a discussion of how cognition of an object is possible and of rules for cognizing objects. At least in advertisement, the B deduction is supposed to take a different approach, one which revolves around the definition of judgment. This approach would produce a tighter link to the metaphysical deduction where the categories are linked to forms of judgment. When we look at §19, however, the project of elucidating what is meant by 'judgment' and what is necessary for

judgments to be possible blends seamlessly into considerations about the requirements of object cognition.

> the [relation or] reference of different cognitions in a judgment … is nothing but the way of bringing given cognitions to the unity of apperception… E.g. [in the judgment] 'bodies are heavy' … the representations belong to one another **in virtue of the necessary unity** of apperception in the synthesis of intuitions, i.e., they belong to each other according to the principles of the objective determination of all representations insofar as these representations can become cognition (B142, my underscoring).

Since 'objective determination' is determination in accord with the rules of concepts of objects, the relation between concepts in a judgment is possible only through the unity of consciousness—which is possible only through the representations belonging to a single self [also] falling under rules for representing objects.

One notable difference between the editions reinforces the importance of object concepts. When stressing the key role of concepts and rules in cognition in A, Kant offered an explicative judgment, 'bodies are extended' as his parade case (A106). This was the exemplary case of an analytic judgment in the introduction of the analytic/synthetic distinction (A7/ B11). By contrast the example in B is the Introduction's paradigm of a synthetic judgment, 'bodies are heavy.' As he notes in the segment I elide from the citation, although this is an empirical or contingent judgment, the presence of the copula 'is' indicates that it is a judgment about objects (and not about the state of the subject) and so can be made only according to the rules governing objects. The appropriate rule is something like 'objects have properties,' which provides an *a priori* template for

any particular claim such as ‘bodies are heavy’ [have the property of being heavy].[41] The A deduction might have given the misleading impression that only analytic judgments are governed by rules; that possible interpretation is ruled out in B. In brief, the argument is that certain forms of judgments are possible only because certain kinds of object concepts are possible, object concepts that are necessary for the unity of self-consciousness. I discuss the relation between the forms of judgments and the kinds of object concepts (i.e. the categories) further in Chapter 13 and its Appendix. What seems clear in §19, however, is that the different strategy that Kant may have anticipated using in the B edition *via* the exact definition of judgment turns out not to be very different from the central line of argument of A that object cognition and self-consciousness mutually imply each other.

Although the multi-faceted B edition argument for the unity of apperception in sections 16 through 19 is not fundamentally different from that offered in the first edition, later parts of the chapter represent an enormous improvement on the treatment of perception in the first synthesis of the A deduction. I follow the argument of B through some of its later stages with an eye to addressing two worries about my interpretation of the argument for apperception. As I have presented Kant’s view, he takes judgmental states and, e.g., states representing sensory properties to stand in relations of rational dependence: The judgmental state is formed through the appreciation of the rational relations among a rule, the materials represented in sensory states (or in other conceptual states) and the judgment. But if that is how the relation of necessary connection across

mental states is to be understood—in terms of relations of rational dependency—how can that relation hold for the case of intuitions? Intuitions don't stand in rational relations to each other. This worry is particular pressing in the B edition where one thesis to be established is that everything that is manifold in *intuition* has a necessary reference to the *same* I-think (B132).

Kant's general view about the relations between intuitions and concepts offers one argument: Any intuition that can participate in cognition must be understood by theorists and cognizers as standing in relations of necessary connection to the conceptual representations abstracted from them. But what is the basis of the relation of necessarily belonging to a common subject of two particular and unrelated intuitions? Suppose that I am enjoying a peaceful day at the beach and I suddenly hear a gunshot. Both the states of, say, contemplating the sea and hearing the shot belong to a common 'I', but *prima facie* at least, they stand in no relations of rational connection to each other. Alternatively, how can my account of necessary connection to each other and so to an 'I' be correct for the flow of perceptions as well as for the train of thought? Even if thinking consists of one rational cognition after another, how can that be the case for perceiving? Examining last few sections of the B deduction will permit me to offer an answer to this important objection.[42]

A second and related worry is how the argument I see in the deduction for the unity of apperception can supply an appropriate premise for the argument from the unity of apperception to the necessary applicability of the categories to anything that a cognizer can sense. No interpretation of the argument for

apperception can be deemed acceptable unless the argument so understood can be seen as part of an overall argument for the ubiquity of the categorial principles in experience. Kant sketches an argument for the category of causation in the last substantive section of the B deduction (§26). I analyze that argument in order to show that and how the argument for apperception that I extract from early parts of the texts is part of an overall argument for the category of causation. Several caveats are needed. Most importantly, I don't think his defense of the causal principle succeeds. So I do not try to show that the argument for apperception that I present is part of a sound argument for the necessity of using causal concepts. Rather, I try to show that it is the sound part of a larger argument that he took to be sound.

Further, even Kant's ambitions for the proof of the categories were less than they appear in prospect. The argument for apperception in the early sections of the B deduction abstracts from space and time. It applies to any sort of rational cognizer. By contrast, the later arguments of the Principles chapter and the argument sketch in §26 do not make this abstraction. So there is no argument from the requirements of rational cognition to the unity of apperception and from that step to the categories *per se* as necessary rules of synthesis. Even if it were successful, showing that temporally interpreted categories are necessary and sufficient for the unity of apperception of spatiotemporal cognizers could be nothing more than a demonstration of the possibility that the categories *per se* are necessary and sufficient for the unity of apperception *per se*. No recourse is available for a defense of the general thesis that cognition requires

the categories, regardless of the particular form of intuition, except an appeal to the table of judgments and its link to the table of categories. For reasons I give in the Appendix to Chapter 13, I don't think this strategy works. To sum up: The goals of the next section are to extend my account of the relations of necessary connection across mental states to the case of intuitions and to show how that account can be understood as part of an overall argument for that categories that Kant understood to be sound.[43]

7. Arguing from the Unity of Apperception to the Necessary Applicability of Categories to Intuitions

Section 20 of the B deduction offers the hypothesis that the categories are just the rules or functions needed to make the unity of apperception possible. Above I suggest that categorial principles might furnish templates for the construction of *a posteriori* rules about objects that make the unity of apperception possible. But neither Kant's hypothesis nor my suggestion bears any resemblance to a demonstration. He seems to believe that he can take the argument no further unless he reintroduces the forms of human intuition in the second part of the argument. His comment in §21 is a useful reminder of the abstract level at which the discussion has so far been carried out. It is a mistake—that I've tried to make less likely by my wording—to think of some representations, intuitions or partial cognitions, as coming *before* others, conceptual representations, judgments, or whole cognitions. The relation between them is one of partial cognition to whole cognition or element to combination of elements, not past to future. Avoiding this confusion may be

Kant's reason for ceasing to talk of the 'reproduction' of representations in the B edition (see above p. nn).

Although Kant is not explicit about the point in the *Critique*, we know from the *Duisburg Nachlaß* that he takes the states of a spatiotemporal cognizer to be subject to a second necessary condition. Not only must such states be represented as belonging to a common subject, they must also be represented as following one another in time. Having reintroduced the forms of human sensibility in §24, he sketches an account of how categories of understanding might apply to anything that can be sensed via the introduction of a *synthesis speciosa*. This synthesis is supposed to apply the rules of the understanding in directing a synthesis of the imagination on sensory materials.[44] He offers the example of the rule for lines directing the construction of a line in thought. Section 26 illustrates the *synthesis speciosa* with the central category of causation. In this crucial section, previous elements of the argument are assembled in a final effort to show that and how the categorial principles must apply to all objects of the senses [of spatiotemporal cognizers].

Before looking at that argument, it is helpful to recognize that it must have a somewhat different form from the reasoning of the earlier sections. In the first portion of the B dedution, Kant is concerned with cognizers' consciousness of themselves as the same subject across representations. So it is an analysis of, roughly, how cognition looks to cognizers (if they think about it.) The discussion of the syntheses involved in perception (the *synthesis speciosa*) must take a different point of view, that of the theorist. This is not an account of what

cognizers are or can be conscious of, but of how perception must be understood in light of the theoretical claim that time cannot be perceived. Ordinary cognizers are similar to his predecessors in believing that they are simple aware of the succession of their mental states.

Kant's first point in § 26 is that humans do not just have forms of space and time (forms that organize materials in a certain way). They also have intuitions or perceptions of determinate spatial and temporal arrays, where properties, objects and events are represented as standing in determinate spatial and temporal relations with each other. Intuitions can be part of cognition only if they can be brought under concepts. Chapter 8 traced Kant's evolving views that the intuitions that are combined in or under a concept must be suitable to that concept. In the *Duisburg Nachlaß* he considered the possibility that the intellectual representations had to be tailored to the sensory or intuitive ones and also the reverse relation. But in the *Critique* and particularly in the discussion of the *synthesis speciosa*, he is clear that the suitability or isomorphism between conceptual representations and intuitive ones that is necessary for cognition could only be brought about through the understanding directing the construction of perceptual representation in accord with its own intellectual rules.

Since §26 offers only a sketch of the argument to come in the Second Analogy, I fill in a few points from the more extended treatment to make it more comprehensible. The argument does not regress all the way from RE cognition, but only from the requirement (for cognition) that a spatiotemporal cognizer can know herself to be the same spatiotemporal thinker. In the second edition of the

Second Analogy, Kant is explicit that time cannot be perceived (B233). So intuition cannot supply the need for a cognizer to be able to represent her states as standing in determinate temporal relations. To represent her states S_1 and S_2 as standing in temporal relations—as a temporal array—it is necessary to represent the synthetic or combined whole of which they are parts and hence to represent S_1 as either following or preceding S_2. But how is it possible to represent them in one relation or another, as preceding or as following?

Reading this background into the preview offered in §26, the reasoning goes as follows. It is a requirement of cognition (and also widely believed) that cognizers can know that their states belong to a single I and to stand in determinate temporal relations. Given the inability to sense either an 'I' or time, how is such cognition possible? When a cognizer thinks of something in causal terms—here, the freezing of water—he thinks of the temporal order as determined. Lowering the temperature necessitates the state of solidity, so he must represent the substance as a solid *after* it was in a fluid state. As Guyer has argued in detail (1987, 241-49), for all the controversy surrounding the interpretation of the Second Analogy, the text is clear that the order of the subject's states is derived from that of the object's states:

> In our case, therefore, I shall have to derive the **subjective succession** of apprehension from the **objective succession** of appearances; for otherwise the subjective succession is entirely indeterminate (A193/B238).

At this point the cognizer can judge that the fluid state was followed by the solid state of the substance and, *via* the causal theory of perception, that his

apprehension of the fluid state preceded his apprehension of the solid state. But the latter is a judgment that the subject's states must have come in a determinate order; it is not a perception of temporal order. And what is to be proved is that the categories must apply to anything that is represented in perception.

To move from the judgment of ice after water to a perception of water followed by ice, Kant introduces in the *synthesis speciosa*. Perceptual representations must be suitable to conceptual ones. So when the understanding represents something causally, it directs the imagination or *synthesis speciosa* to construct a sensory representation that is suitable to the causal relation. (At this point, the account may seem hopelessly idealistic; I discuss the role of sensory data in limiting cognition in Chapter 13.) In the case of a temporal cognizer, '*propter hoc*' has a suitable sensory representation in '*post hoc.*' The relation of the lowering of the temperature *causing* water to become ice is represented perceptually as the cognizer apprehending the substance as ice *after* apprehending it as water. So the cognizer not only judges ice after water but constructs a perceptual representation of water followed by ice.

Kant compresses the reasoning as follows (with interpolations):

> When I perceive the freezing of water, then I apprehend two states [of the object] (fluidity and solidity) as states that stand to each other in a relation in time. Since the appearance [the subject's perceptual representation] is [belongs to] inner **intuition**, I lay time at the basis [the representation must be understood as containing temporal relations]. But in time I necessarily represent synthetic **unity** of the manifold [to represent a temporal array the cognizer must combine representations, since time

> cannot be perceived]; without this unity [this combination by the *synthesis speciosa* directed by the understanding as it grasps causal relations], that relation [of temporal sequence] could not be given **determinately** (as regards time sequence) in an intuition [and the state could not be understood as being the state of a spatiotemoporal cognizer] (B162-63).

Since only the causal relation can determine temporal relations (or so Kant believes), it follows that the contents of any sensory state that can be understood as such (as belonging to a subject and as standing in the following relation to other states) must be such that it can be understood through the category of causation. The perceptual array acquires its needed temporal relations in being constructed to be suitable to causal relations, so this result is inevitable for all perceptions that involve temporal relations—which they all do—and so for all perceptions. In particular the case of *prima facie* 'unrelated' intuitions, perceiving the calm sea and hearing a gunshot, can be understood in the same way as the water to ice case. In this case a larger object, the overall scene I am contemplating, is altered by the firing of a shot and, again, I derive the order of my states from my rational appreciation of the causal relation among objects: An area changes from being quiet to containing a piercing noise when a gun is fired in it.

I do not endorse Kant's theory of the perception of time or his argument for the causal principle. My purpose in running through the argument sketch in §26 is to defend my interpretation of the first half of the deduction, the argument from cognition to apperception, by showing how it fits into the argument *from* apperception. It is easy to read the Second Analogy as pursuing a different and

simpler line from that of the deduction chapter(s). Cognizers must be able to understand their mental states as standing in relations of following; allegedly, only a causal law can determine necessary succession; one mental state can be understood as coming after another, because the order of objects represented by the states is determined through the causal law. This approach ignores material that is crucial to both the argument for the causal principle and the doctrine of the unity of apperception. The important feature of causal laws is not just that they refer to temporal succession, but also that they are rules. It is the latter feature that makes the operations performed on representations by the imagination to be syntheses. These are not accidental groupings of representations, but combinations of them according to rules.

The *synthesis speciosa* is not understood as such by the cognizer. But for the theorist it is clear that the relation between intuitions in a temporal array—the representation of perceiving ice after perceiving water—depends on rational considerations. The perception of ice is not rationally dependent on the perception of water; but the perception of ice after the perception of water depends on a rational appreciation of the relations among the representations of 'water,' 'ice' and a causal rule. The first part of the B deduction argues that it is only the use of an object rule that permits the I-rule to be satisfied in the case of judging; §26 and the Second Analogy argue that it is only the use of an object-rule that permits the construction of representations of temporal arrays that meet the second condition for apperception, the requirement that the states of a spatiotemporal subject be represented as following one after another. One

advantage of my interpretation of the first part of the deduction argument is that it enables us to appreciate that and how object-rules are necessary for representations to meet both the 'belonging to a common subject' and the 'following' requirements on representations. It offers a univocal account of how the two key elements that cannot be acquired through the senses are supplied through the activities of the mind. Intuitions that can stand in determinate temporal relations must be constructed through the use of object rules, and any intuition that can be thought must be thought under a concept whose use sets up a relation of rational dependence across different representations that makes them states of a common 'I.' Since any intuition that is anything to a thinker must be able to be thought, it follows that it must be possible not just to order representations in time, but to connect them to a common subject.

Kant claims that the bond created across states through the action of synthesis is the same whether the synthesis is conscious of not. It is the same in being the result of the mind's rational activity. At least from a theoretical perspective, representations must be understood as connected to each other in the flow of perceptions as well as in the train of thinking through the mind's rational activities. At first glance, my interpretive claim that the necessary connection across mental states of a single cognizer is one of rational dependence looks implausible for the case of intuitions. When we consider the role of the *synthesis speciosa* in the construction of temporal arrays and so in the argument of §26, however, we can see that Kant is committed to the view that an

appreciation of rational relations is the source of all the connections—conceptual and temporal—that bind mental states to each other.

Matters do not appear that way to cognizers. For them, the transition of their minds from perceiving a ship upstream to perceiving it a bit downstream, for example, appears to be a single experience in the specious present. Because they grasp this experience as a whole, they take themselves to perceive the connection between the two representations as well as their order. Arthur Melnick has argued that the denial of the awareness of succession in the specious present is a serious weakness in the argument of the Second Analogy (1973, e.g., p. 85). As far as I can tell, however, the specious present is a consequence of Kant's theory rather than a problem for it. He agrees with his predecessors that it seems to cognizers as though they are aware of a succession of representations and aware of those representations as belonging to a common self. Those are the data that his theory must explain in light of what he takes to be the philosophical discoveries that humans can sense neither time nor a common self.

Having shown (I hope) how my account of the argument for apperception fits into the overall deduction, as Kant understood it, and how it can be applied to the case of intuitions, I return to an issue put aside in section 5, that of mental transitions (for spatiotemporal cognizers). Leibniz took awareness of transitions from one state to another to be essential to the consciousness of an enduring subject. Both Henrich and Carl read Kant as basing the argument for the categories on the need to guarantee that transitions from one mental state to

another can be understood as such (Wunderlich, 2007, 228). In section 5, we considered one sort of transition in the counting example: the counter's transition from designating a stroke symbol as '1' to designating the next stroke symbol as '2.' These transitions—representing '2 after 1'—are driven by and grasped through the intention to follow the counting rule. The argument of §26 (and the Second Analogy) provides an account of other transitions, those brought about through actions of external objects on the perceptual system. In both cases, the understanding determines the order of the contents of 'inner sense' and so the transition from state to state. It does so either intentionally, through a construction of representations in thinking, or in response to the actions of external objects on the senses, where it directs the imagination to build a suitable sensory representation of causal relations using the contents of the representations, e.g., a sensory representation of 'water followed by ice.'

Kant's theories of mental unity and of mental transition are remarkably uniform: The active power of understanding that combines materials from sensory data according to rules accounts for both the unity and the transitions. His theories are as uniform as the Empiricist and Rationalist theories that they were intended to replace. For Locke, an unusual sort of sense, inner sense or reflecting consciousness, informs subjects of both their transitions from state to state and of their continuing identity through those transitions. Leibniz's metaphysical theory also offered a unified account. He took the mind, as substance, to have both an intrinsic unity and an internal principle that accounted for the unfolding of its perceptions in a particular order. Kant saw this theory as

paradigmatic of the sort metaphysical theory that blocks progress in science. He does not offer a scientific hypothesis in its place, but a theory of how the mind must function to meet the demands of RE cognition. A cognizer *per se* must have unity, but not a substantial unity. Rather it is a transcendental (*a priori* and required for cognition) unity—a relation of necessary connection across representations that is understood as such—and that is produced in the course of cognition by faculties that combine representations according to rules. It is only his theory of this strange non-substantial unity that must be produced through acts of combination involved in thinking that makes it possible for him even to try to argue from the necessity of apperception to the necessary use of object rules for combining representations.

8. Transcendental Apperception, Empirical Apperception and 'Mineness'

Despite all that has been accomplished to establish the unity of apperception and the necessary applicability of the categories, there is still a large hole in the deduction's theory of the thinker. For all Kant's efforts to show that the activity of thinking or synthesizing brings about a relation to a common 'I-think,' he still needs to explain how an individual cognizer is able to say '*I*-think.' The obvious place to look for an answer is in his theory of inner sense or empirical apperception. Empirical apperception is supposed to explain how a particular thinker self-ascribes particular representations at particular times (B158). Kant also introduces the notion of 'empirical apperception' to provide a clarifying contrast for the 'transcendental' unity of apperception. We've already considered some important aspects of that contrast in relating his early view

about ‘inner sense’ to his later conception of ‘apperception’ as an active self-conscious faculty. I conclude my study of the argument for apperception in the B deduction by looking a several other systematic discussions of empirical apperception and by considering again the ‘mineness’ problem.

Kant’s strategy of clarification would be more successful if his notion of ‘empirical apperception’ were clearer. As we look at some key texts, however, it will be hard to disagree with the interpretive consensus that he left this notion too undeveloped to do any serious work in his theory. His main point about empirical apperception is that it is dependent upon transcendental apperception. Since ‘apperception’ refers both to a faculty and to the unity of consciousness (produced by that faculty), the contrast can involve four different items: the faculty and unity of transcendental apperception and the faculty and unity of empirical apperception. I have deferred a systematic discussion of the faculty of transcendental apperception until the next chapter, but discussing the relations between empirical and transcendental apperception will require me briefly to consider the faculty as well as the unity of transcendental apperception.

The introduction of ‘inner sense’ or ‘empirical apperception’ in the *Critique* seems inconsistent with later claims. ‘Inner sense’ is introduced as the faculty

> by which the mind intuits itself, or its inner state (A22/B37).

This description and later ones (B158) suggest that empirical apperception provides representations of a state of the mind at a time: I am tasting something sweet now; I was thinking about the properties of a square earlier. But this claim is inconsistent with a central thesis of the Transcendental Aesthetic that time

cannot be sensed and with a central thesis of the deduction that the 'I' cannot be sensed.

Chapter 2 proposes handling the apparent discrepancy between Kant's claims that time cannot be perceived and that it is the form of inner sense by making a distinction between inputs to and outputs from inner sense (p. nn). We now have a somewhat fuller account of how temporal relations get "added" to the raw materials of inner sense. The understanding, which he equates with the 'faculty of (transcendental) apperception,' directs the *synthesis speciosa* to produce a sensory representation suitable to the intellectual representation of a causal relation. In this way, the faculty apperception 'affects' the outputs of inner sense (B153-54). Time cannot be sensed, but it does not seem that way to cognizers. Due to the complex processing needed to produce cognition, a subject is aware of his states through inner sense as following one another in time. On Kant's account, cognizers do not intuit their states as states of a continuing subject through inner sense either. Again, it doesn't seem that way to cognizers, however, which may be why so many (including himself at an earlier time) take the 'I' to be an intuition. As he now understands the situation, a cognizer is aware of states through inner sense as belonging to a common subject only by virtue of his activities of thinking about objects.

When read against the background of the theory he goes on to develop, Kant's introduction of the faculty of inner sense or empirical apperception can be understood as non-paradoxical (though, perhaps, somewhat misleading.) Further, as we reconcile the apparent inconsistencies, we come to appreciate his

central claim that the ability of the faculty of inner sense to provide representations of the mind and its states is thoroughly dependent on the faculty of understanding or transcendental apperception.

In his focal discussion of the 'subjective' or 'empirical' unity of apperception in §18 of the B deduction, Kant argues that it depends on the 'objective' unity of [transcendental] apperception. He contrasts the two unities in three different and not obviously compatible ways. The subjective unity is contingent as opposed to the necessary objective unity; the objective unity is related to determining the time sequence whereas the subjective is not; the subjective unity of empirical apperception is derived from the objective unity in particular conditions (B139-40). If the subjective unity is just a special case of the objective unity, as the third condition suggests, then it seems that it cannot be so different from it in the first two respects.

One clue to resolving the tensions in §18 is that the 'subjective' unity rests on the law of association (Allison, 1983, 156). Kant accepts that even conscious representations are subject to the law (A100), so he wants to explain how representations that must belong to the unity of apperception can be understood as also subject to irrational forces. Locke's original use of 'association' was to explain superstitions such as the "connection" between goblins and darkness (ECHU, 2. 33.10, 397-98). A surfeit of fairy tales as children lead adults to associate darkness with danger, but this "relation" is entirely contingent and cannot be used to determine the order of anything in time. The linkage between darkness and danger is subjective in a number of ways. There is no logical

connection between the ideas; their connection depends on a contingent history; because of these factors, there is no place for inter-subjective agreement. That my history leads me to make the connection cannot be checked by asking you whether you also make the connection. However you answer it is irrelevant to the existence of my association. Despite the thorough contingency of this connection, it would be impossible if I could not represent darkness. For the law of association to operate on conscious representations (as well as on unconscious ones which aren't Kant's concern here) the cognizer must have conscious representations. This is the basic sense in which the subjective unity is derived from the objective unity and is a special case of the objective unity. The implied contrast is that the relation could never go the other way. No amount of irrational or accidental association among representations can produce the necessary connections across representations that are required for RE cognition. In §18, the B deduction finally makes good on the project attempted in the Second Synthesis in A: It explains how the law of association can hold while still being irrelevant to cognition and while presupposing the active faculties that make cognition possible.

So the unity of empirical apperception depends on the transcendental unity of apperception as much as the operation of the faculty of empirical apperception depends on the activity of the faculty of understanding or transcendental apperception. Kant never entertains the possibility of the reverse dependence. But since RE cognition requires that the cognizer know the basis of his judgments and that basis is his representations, however, isn't empirical

apperception also necessary to RE cognition? How can the cognizer be aware of combining some of his representations in others unless he is aware of the representations? Kant seems committed to giving empirical apperception or inner sense this crucial role of enabling cognizers to be aware of their representations as such. In that case, however, he should regard the operation of empirical apperception as necessary for the operation of transcendental apperception. He never concedes this point, so perhaps he held the semi-official view already discussed. Perhaps he believed that the actions of the mind in thinking not only create relations across representations, but also require a cognizer to use her capacity to understand representations as such. But he doesn't say this.

Finally, as noted at the end of Chapter 2, empirical apperception or inner sense seems necessary to solve the problem of how humans use the terms 'I' and 'my.' They call a sensation 'mine' or claim that '*I* think that water turned into ice,' because they are aware of their sensations and judgmental states through inner sense. Given his theory, Kant should acknowledge that rational cognition also depends on empirical apperception. But he doesn't concede this point either. Konrad Cramer has suggested that Kant is able to extract the '*I* exist' from 'I think' because thinking is a self-conscious act (1987, 171). The fact that an act directs a series of sub-actions is, however, insufficient to establish the 'togetherness' of the representations involved. Cramer well understands that there is a great difference between establishing that a representation must belong to an I-think and showing that different ones must belong to the same I-

think (1987, 174). Still even if the act-awareness on its own could not establish the 'togetherness' of representations, it seems well-suited to provide an account of their mineness.[45] Representations are mine because I am conscious of creating some from others—conscious in making judgments. Since this is not Kant's view official theory of 'mineness', but an emendation of it, I postpone further discussion until Chapter 15 where I propose a Kantian alternative to current views on the self-knowledge of belief.

9. Summary

The argument of the transcendental deduction for the unity of apperception consists in an examination of the necessary conditions for RE cognition. Alternatively, it involves a regression from the assumption that humans have such cognition, which sketches how such cognition is possible. It is possible because humans can be aware of their representations as such and can combine representations through self-conscious acts that enable them to grasp the relations of necessary connection across their representations. The deduction also argues for RE cognition from the assumption that humans know themselves to be an identical self across different representations. The back and forth character of these arguments is possible, because a central thesis is that RE cognition and self-consciousness mutually imply each other.

Since the argument proceeds by an analysis of the normative requirements of RE cognition, the resulting theory of the 'I-think' is not metaphysical but epistemological. It establishes only weak and conditional metaphysical claims, e.g., if RE cognition is possible, then some creatures must

have the capacity to recognize their representations as such and to engage in conscious synthesis of representations. Given the result that it is a requirement of cognition that different representations be understood as belonging to a common subject, however the deduction offers a defense of the alleged metaphysical principle that all representations inhere in a common substance. Since this principle cannot be established on metaphysical grounds, the deduction provides the only way of defending what seems to be the obvious truth that different mental states necessarily belong to a single thinker.

The transcendental deduction reveals several unusual features of thinkers. Since the necessary connection across diverse states specified by the I-rule can only be established through engaging in thinking, the thinker is, in a real sense, created by his acts of thinking. Absent these, the various capacities required for thinking would lie dormant and never produce thought or the necessary connections that thought requires. Further, thinking and the self-consciousness it enables require an act-awareness in combining representations that is hard to fit into either Kant's system or contemporary theories of mind. In a passage that will be discussed at greater length in Chapter 10, he can only offer a negative characterization of the awareness at issue:

> when the understanding is considered by itself alone, then its synthesis is nothing but the unity of the understanding's act: of which the understanding is conscious as such even apart from sensibility … (B153, amended translation).

Finally, despite Kant's enormous efforts to limn the contours of the thinker, the resulting theory contains a large hole. The deduction is all about the 'I-think,' but

it does not work out a good theory of how cognizers are able to use the representation '*I*.'

[1] At B139, Kant characterizes the synthetic unity as the first principle of human understanding; he also characterizes it as the 'highest principle' of cognition. As we see below, these characterizations are consistent with his view that it is none the less dependent on the use of categorial concepts and principles.

[2] I. e., Chapter 1 of Book 1 of Division 1 of Part 2, "The Clue to the Discovery of All Pure Concepts of Understanding."

[3] Longuenesse offers a 'logical' gloss of these texts (1998, 81).

[4] The agreement is not universal, because some take the argument for the categories to come in the so-called 'metaphysical deduction.'

[5] See, for example, Howell's helpful presentation (1992, 61ff.).

[6] I'm grateful for Christopher Peacocke for urging me to remove any suggestion that Kant takes cognizers to have a meta-cognitive vocabulary of 'marks' and 'representations.'

[7] I'm grateful to a comment by Stephanie Grüne that led me to see that I should discuss a wider range of cases of concepts.

[8] In a note to the B Paralogisms, Kant also raises the contrast between differentiating and being aware of the basis of the differentiation:

> A clear representation is, rather, one in which the consciousness suffices for **being conscious of the distinction** between this representation and others. If the consciousness suffices for distinguishing between them but not for being conscious of the distinction, then the representation would still have to be called obscure. Hence there are infinitely many degrees of consciousness, down to it vanishing (B414-15).

Although a human's consciousness of her representations may have degrees, from the obscure to the clear, Kant does not believe that the difference between human and animal cognizers is one of degree. Sometimes humans can act 'instinctively,' with little awareness of what they do. Animals are incapable of being conscious of their representations as such, *a fortiori*, incapable of being conscious of them as the basis of the way they distinguish things (cf. 25.1033).

[9] This is not meant to be an appallingly bad pun on 'recognition.'

[10] I made this over-hasty assumption in Kitcher (2008).

[11] My understanding of the status of this principle was clarified by a workshop hosted by Tobias Rosefeld at the University of Konstanz in June of 2008.

[12] Béatrice Longuenesse used this helpful formulation in conversation.

[13] In the B edition Paralogisms Kant characterizes the principle that different states belong to a single subject as 'analytic.' See Chapter 11, section 4 for discussion.

[14] I make no effort to defend Kant's appeal to rules associated with concepts. On Kripke's (1982) interpretation of Wittgenstein's discussion of rule-following (1953), it makes no sense to say that concept-users follow rules for or associated with concepts. As far as I understand that argument, Kant's position would be vulnerable to Kriptensteinean skeptism about rules. Whether this skepticism can be mitigated has been a subject of intense debate. Paul Boghossian (2008) offers a recent endorsement and expansion of the argument for rule skepticism.

[15] For further discussion see my 1990, Chapter 8 and Philip Kitcher (1982).

[16] I offer this more precise formulation after Béatrice Longuenesse and Robert Howell questioned how '2' could be a partial representation of '4' at a meeting of the German Idealism Workshop in November of 2009. I use the term 'group' rather than 'set' in the text, because Kant pre-dates any systematic study of sets. On the other hand, his logic notes contain proto-Venn diagrams indicating the relations of set inclusion for different types of judgments. See, e.g.,16.627.

[17] Jen Saugstal pointed out that because of numbers can be used in two ways, as ordinals and as cardinals, there is some ambiguity about how exactly to understand Kant's argument. In the text, I assume the 'cardinal' reading, because Kant uses the term '*Menge*'. Still the argument can also be run on an ordinal reading. In this case, it is not that a set of stroke symbols exemplifies the '4' rule, but that the counter has kept track of where he is in number sequence. He designates the next stroke symbol as the '4th' not on the basis of the '4' rule, but on the basis of his being aware of where he is in the sequence. In either case, Kant's point is that the use of number concepts in adding or in counting is possible only through an awareness of the judgmental state '4' or '4th' as dependent on representations contained in other states.

[18] Chapter 15 spells out the problems in greater detail (section 4).

[19] I owe this formulation to a very helpful discussion with Sebastian Rödl. Even though the awareness of a mental action cannot be separate from it, when the awareness occurs, it might seem that there are mental actions that are not accompanied by awareness. Chapter 15's discussion of the possibility of rational cognition in the absence of self-awareness also considers the possibility of rational mental action in the absence of awareness.

[20] Because he wants to argue that cognition of succession (and so change) in one's mental states requires cognition of change in the world, Kant focuses on a principle associated with 'substance,' *viz*., that in all alteration, substance is permanent. But he takes the category to be that of the inherence of properties in objects (and ultimately substances), A80/B106.

[21] See Chapter 7, p. nn.

[22] Wunderlich (2005, 56, n. 193) suggests that Eberhard's claims may have been a target of the second Paralogism in A.

[23] The text I elide concerns the relation between the unity of apperception and that of space and time. This was a doctrine Kant sometimes offered, but did not in the end maintain.

[24] See Chapter 7, pp. nn, Chapter 8, pp. nn.

[25] Besides the account of the difference between innate representations and *a priori* ones in the Eberhard polemic, Kant apparently also considered the priority of the faculties and the necessity of data in his Metaphysics L_1. In reply to the question of whether pre-birth souls have cognitions, he explains that

> the soul ... already possessed all abilities and faculties; but such that these abilities developed only through the body, and that it acquired all the cognitions that it has of the world only through the body ... . **The state of the soul before birth was thus without consciousness of the world and of itself** (28. 284, CMeta, 93).

Although the issue concerns the necessity of embodiment, the position is the same as that offered in the Reply to Eberhard. Humans come equipped with faculties (and so with innate tendencies to operate in various ways), but without sensory data they cannot have cognition or even consciousness of themselves. On his view, faculty of apperception and its the I-rule precede the receipt of sensory data, but the consciousness of unity of consciousness and perhaps the unity itself, arise only through the receipt of data—and only through the receipt of data that can be handled in such a way that the I-rule can find application.

[26] I use the counting example as the parade case of Kant's theory, because he does in the A deduction and because it is simpler. In the case of 'body,' for example, for the judgment to be rational, it might not be based on a partial representation 'extended,' but on some other partial representation such as 'shape.' Or it could even be used as a simple concept such as 'orange' is used, where the rationality of the cognition does not rest on any particular partial cognition, but on the presumption of a set of similar representations from which the concept was 'abstracted.' In the latter case, the cognition is rational—is a case where the subject knows the basis or ground—because she takes the present judgment to be based on the similarity of the present instance to

the instances from which she abstracted the concept. Here, too, in making the judgment, the subject is aware of its rational dependence on and so necessary connection to, the earlier states.

[27] Several critics saw Wolff's appeal to differentiation as involving too much complexity to provide a basis for self-consciousness—because differentiation implicitly involved judging (Wunderlich, 2005, pp. 41-42). If the critics were right, and Wolff saw object differentiation as requiring judgment, then his claim that cognizers were aware of themselves through being aware of their acts of differentiating or judging would be in substantial agreement with the doctrines of the *Critique*. On that reading of Wolff, Kant would not be criticizing his predecessor's view, but elaborating it.

[28] The claim that such possible unconnected sense data would be nothing to me repeats the assertions that any representation that could not participate in cognition would be impossible as a representation (A116, B132) or would be nothing to me (B132).

[29] Carl made this point in conversation.

[30] As noted in Chapter 1, it is not necessary that I know him to be here on this basis; I might have some other reason for the judgment. The point is that since I do know him to be here on the basis of seeing him by the window, the rationality of my judgment in this case depends on that observation.

[31] In Chapter 15, I argue that Kant's view raises a significant obstacle to Evans's widely accepted approach to self-knowledge.

[32] See note 34 for defense.

[33] Kant elaborates by suggesting the fact representations that have been synthesized can be analyzed. Presumably the idea is that analysis is more obviously a single (type) of operation. Still, I'm not confident I see his point about analysis.

[34] I'm grateful to a referee for OUP for pressing me on how to read this passage. Although '*einig*' strongly suggests a single act, I don't think '*einig*' can be handled separately from '*ursprünglich*.' But reading Kant as suggesting that all of a human's cognition comes out of a single original act moves him too close to a Leibnizean view that the course of (cognitive) life is just the unfolding of a single principle.

[35] Cf. the claim about spatial representation reflecting the peculiar character of the receptive faculty in the citation from the Reply to Eberhard, above, p. nn.

[36] I'm grateful to Bernard Thöle for raising this objection in response to an earlier description of what the counter does. The relation between these two rules will be crucial in Kant's diagnosis of the errors of the Rational Psychologists in the Paralogisms.

[37] Hence I think Allison errs in dismissing this material (1983, 137).

[38] See Chapter 5, pp. nn. Henrich argues in a well-known essay that Kant follows many predecessors in adopting an impossible 'reflection' theory of consciousness (1966, 191-93). The criticism of Locke (or of someone who holds Locke's position and/or of Wolff or someone who holds his position) at the pinnacle of the B deduction shows that he well understood that cognizers cannot use reflection, or mere consciousness of representations, to boot-strap their way either to having a self or to being self-conscious.

[39] I'm grateful to Robert Howell for stressing the importance of making clear Kant's own difficulties with reference to an 'I.'

[40] I argued in section 5 that there is no inconsistency between the claim that the 'I-think' is an *a priori* representation and so, in a sense, 'precedes' actual thought and the claim that the unity of apperception is made possible through the unity [or combinability] of representations.

[41] In the set-up to this discussion, Kant mentions the other two sorts of judgments, the hypothetical and the disjunctive. Presumably these would work in the same way. In a hypothetical judgment, the relation is not between concepts, but between judgments (B141). These are linked not by the copula 'is' but by connective, 'then' (cf. 9.106-106, CLog 599-600). So in the judgment, 'If cold it applied to water, *then* it freezes' the connective 'then' would indicate a reference to the *objective* unity of apperception. In this case, the object rule 'events or changes are brought about by causes' would be the *a priori* template underlying the judgment, 'if cold is applied to water, then it freezes.'

[42] Bernard Thöle pressed this objection forcefully—and helpfully—at a discussion at the Max Planck institute in Berlin in January of 2008.

[43] I'm grateful to an anonymous referee for OUP for leading me to clarify exactly what I aim to show.

[44] Chapter 8, p. nn.

[45] My thinking on this issue has been much influenced by discussions with Christopher Peacocke and by the chapter on Act-Awareness in his 2008.