

THE UNITY OF APPERCEPTION IN THE CRITIQUE OF PURE REASON

José Luis Bermúdez

Kant's account of the unity of apperception is at the heart of the notoriously opaque argument of the Transcendental Deduction of the Categories, and the Transcendental Deduction is itself the heart of the Critique of Pure Reason. The tantalizing prospect which Kant offers his readers in the Deduction is of a constitutive connection between some form of self-consciousness, on the one hand, and experience of an objective world, on the other. From the simple fact that we are aware of ourselves in the way that he tries to convey with the notion of the unity of apperception there follows the objective validity of the categories, and hence the existence of a causally structured spatio-temporal world of mind-independent objects. In this sense, then, the unity of apperception bears the weight of the Critical Philosophy. But the subject is of considerably wider interest than Kant exegesis. The problems which Kant attempts to solve with the unity of apperception remain of central interest to philosophers concerned with self-consciousness and personal identity. Here, as in so many other areas of philosophy, attention to what Kant has to say is an indispensable starting-point.

I begin by briefly discussing certain very general constraints which any satisfactory reading of the unity of apperception must obey. These constraints will be expanded upon in the main part of the paper. I then introduce a reading of the unity of apperception that I term the Substantial Ownership Reading. This interpretation is rejected and an alternative put forward, the Formal Ownership Reading. It is argued that the Formal Ownership Reading satisfies the general constraints, and in so doing provides a plausible premise for the Transcendental Deduction.

Constraints on Interpreting the Unity of Apperception

1. The interpretation must make the unity of apperception a plausible premise.

The Transcendental Deduction of the Categories takes the general form of a transcendental argument. Such arguments start from a basic fact or set of facts about experience, and attempt to

establish the necessary conditions which must hold if that fact or set of facts is to hold. A classic example is provided by the Second Analogy, where Kant takes it as a premise that he is capable of distinguishing the subjective order of experiences from the objective order of events, and argues that this can be possible only if all alterations take place in accordance with the law of cause and effect.¹ For a transcendental argument to be effective, the premise must be a relatively uncontentious feature of experience - uncontentious enough, for example, to be accepted by a sceptic. This emerges particularly clearly in Kant's most explicitly anti-sceptical transcendental argument, the Refutation of Idealism which he added to the second (1787) edition to rebut the charges of idealism levelled at the first (1781) edition. There he takes as a premise the undeniable fact that he is aware of his own existence as determined in time, something no sceptic could take issue with. It seems reasonable to expect a satisfactory interpretation of the Transcendental Deduction to be consistent with this aspect of the structure of the other transcendental arguments in the Critique, and that means that the unity of apperception must be interpreted in a way that does not make it immediately open to sceptical objection.

2. The interpretation must not conflict with the Paralogisms

In the section on the Paralogisms of Pure Reason in the Transcendental Dialectic Kant discusses the fallacies into which philosophers are drawn when they attempt to draw substantive conclusions from the Cartesian cogito. He ridicules arguments purporting to demonstrate the existence of a simple and substantial soul from the cogito. Such arguments conflate the unity of experience with the experience of a unity, and he makes it very clear that if such philosophers (the rational psychologists) had a proper understanding of the unity of apperception they would not have been drawn into such wild speculation. Clearly, then, the unity of apperception must be interpreted so that it does not come out as a Paralogism, or to be making the sort of claims that would feature in the conclusion of a Paralogism, because Kant is emphatic that the conclusions of the Paralogisms are as false as the arguments are fallacious.

3. The interpretation must be supported by Kant's arguments for the unity of apperception

Although Kant takes the unity of apperception as a premise, he does not think it sufficient just to draw attention to it as a fact about our experience, any more than he is prepared to accept the

spatiality of our experience as a brute fact. The Transcendental Deduction attempts to establish the necessary conditions that must hold for any experience to satisfy the unity of apperception, but this should not obscure the fact that the unity of apperception is itself a necessary condition of other features of our experience. The unity of apperception is hypothetically necessary and, as such, is grounded by transcendental arguments. One such argument is Kant's frequently repeated claim that the unity of apperception is implied by what it is to be conscious of a representation.² Clearly, then, any convincing interpretation of the unity of apperception should explain why Kant should have thought that transcendental arguments like that one were convincing.

4. The interpretation must respect the crucial steps of the Transcendental Deduction

The precise details of the argument of the Transcendental Deduction are very much a matter for dispute among Kant scholars, and they are likely to remain that way. In both the first and second editions, however, there are two crucial steps in the argument. The first is the claim that a consciousness for which the unity of apperception holds can know a priori that this is so. The unity of apperception is not just a requirement on consciousness. It is a requirement that can be known to hold a priori. The second crucial step in the argument is the move from the unity of apperception to the need for a priori synthesis. As he puts it in the second edition ". . . the analytic unity of apperception is possible only under the presupposition of a certain synthetic unity".³ It is this that makes possible the introduction of the categories as the appropriately fundamental rules of synthesis. In which case, putting the point in very general terms, the unity of apperception must be interpreted so as to make it plausible that it is underwritten by a priori synthesis. This provides a useful counterbalance to the first constraint. It was suggested earlier that the unity of apperception should be interpreted weakly enough to get past a sceptic, but we can see now that there are limits to how weakly it can be interpreted. It must be sufficiently rich to support the demand for a priori synthesis of the sort that brings in the categories.

Preliminary Remarks: Unified Consciousness and Self-Ascription

The unity of apperception is introduced in the first edition version of the Deduction as follows:

There can be in us no modes of knowledge, no connection or unity of one mode of knowledge with another, without that unity of consciousness which precedes all data of intuitions, and by relation to which representation of objects is alone possible. This pure original unchangeable consciousness I shall name transcendental apperception.⁴

A little further on Kant makes clearer what he means by a suitably unified consciousness:

There is one single experience in which all perceptions are represented as in thorough-going and orderly connection, just as there is only one space and one time in which all modes of appearance and all relation of being or not being occur. When we speak of different experiences, we can refer only to the various perceptions, all of which, as such, belong to one and the same general experience.⁵

The parallel with the forms of intuition is instructive. Just as all the events which an individual experiences must take place in a single spatio-temporal order (as argued in the relevant parts of the *Transcendental Aesthetic*), so too must all the experiences he has of them fall within a single experiential continuum, within a single unified consciousness. The general outlines of the claim are clear enough. The unity of apperception depends upon experiences being connected up in such a way as to form a single series. But before evaluating any such claim we need some understanding of the connections that have to hold between experiences within a unified consciousness.

One might hope for illumination from Kant's frequent and celebrated mentions of the 'I think' in the context of the unity of apperception. The following is the best known passage:

It must be possible for the 'I think' to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me. . . I call it pure apperception, to distinguish it from empirical apperception, or again, original apperception, because it is that self-consciousness which, while generating the representation 'I think' (a representation which must be capable of accompanying all other representations, and which in all consciousness is one and the same), cannot itself be accompanied by any further representation.⁶

But what is it for the 'I think' to be capable of accompanying all my representations? One plausible view here is that this is just an elaborate way of saying that it must be possible for me to ascribe all my representations to myself. This is the view taken by Strawson, which has been justly influential.⁷ Putting it together with the previous thought yields the idea that in the unity of apperception my experiences are connected up to form a single series in virtue of it being possible for me to ascribe them all to myself. This leaves open two very important questions, however. First, it

is not clear what the relation is between the capacity to ascribe experiences to oneself and the unified consciousness mentioned in the passages quoted earlier. And secondly there is the important question of who or what is the 'I' to which these experiences are being ascribed.

Had Kant made the stronger claim that the 'I think' must accompany all my representations it would be easier to see a connection between the self-ascription of representations and the sort of unity of consciousness that permits him to refer to 'one and the same general experience'. On such a view the necessary self-ascription of all experiences might be thought to generate automatically an appropriately unified consciousness. In such a situation all one's thoughts and experiences would be self-ascribed, and there could be no question (at least from a first person perspective) of whether or not they fell within a single unified series, because they all come with the 'marker' of self-ascription. But this does not seem to hold when what is at stake is just the possibility of self-ascription. Why should such a possibility be thought to explain the unity of consciousness? If there is a representation that is not self-ascribed, then it cannot be the possibility of ascribing it to oneself that explains why it is part of a unified consciousness. On the contrary, surely it is only possible to ascribe it to oneself because it is connected up with one's other thoughts and experiences so that they are all part of 'one and the same general experience'? The possibility of ascribing experiences to oneself presupposes a unified consciousness, and so cannot be appealed to explain how experiences are connected up in a unified consciousness.⁸ Any initial puzzlement one might feel about the very idea of a unified consciousness will not be dispelled by the simple fact that the 'I think' must be able to accompany all my representations.

Nonetheless, even if the possibility of self-ascription does not explain what it is for a consciousness to be unified in the appropriate manner, there does seem to be a very close connection between these two components of the Kantian notion of the unity of apperception.⁹ It is a natural thought that when one claims that a set of experiences are part of a unified consciousness one is effectively claiming that they are all experiences of a single subject (that they are all 'owned' by a single subject), and one can see Kant's emphasis on the 'I think' as sharing this intuition. So, retaining the thought that the unity of consciousness is presupposed by the possibility of self-ascription, it might be suggested that there must be more to the unity of consciousness than

experiences being connected up with each other in a certain way. They must also be unified in virtue of being owned by a single subject. If this is right, then the possibility of ascribing experiences to oneself would follow easily from the fact that those experiences are part of a unified consciousness. (At least, it would follow easily provided that it is made a condition of possessing a suitably unified consciousness that one should grasp that one is oneself the owner of one's experiences. Kant does in fact impose such a condition, because it follows from his claim that a subject must be able to know a priori that his experiences are subject to the unity of apperception.)

Once again, though, the significance of this depends upon how the 'owning' subject is understood, and this brings us back to the earlier query about the nature of the 'I' to which experiences are ascribed. One point which Kant continually stresses here is the identity over time of the 'I'. For example (from the second edition):

I am conscious of the self as identical in respect of the manifold of representations that are given to me in one intuition, because I call them one and all my representations, and so apprehend them as constituting one intuition.¹⁰

And even more explicitly in these three passages from the first edition:

All possible appearances, as representations, belong to the totality of a possible self-consciousness. But as self-consciousness is a transcendental representation, numerical identity is inseparable from it, and is a priori certain.¹¹

We are conscious a priori of the complete identity of the self in respect of all representations which can ever belong to our knowledge, as being a necessary condition of the possibility of all representations.¹²

The abiding and unchanging 'I' (pure apperception) forms the correlate of all our representations in so far as it is to be at all possible that we should become conscious of them.¹³

These four passages all seem clearly to suggest that, when experiences are self-ascribed at different times, they are ascribed to a numerically identical 'I'. This is a crucial feature of the unity of apperception, although everything depends upon just how this ascription to a numerically identical self is understood.

We seem, then, to be in a position to extract the following points from this preliminary survey of what Kant has to say about the unity of apperception.

(1) The unity of apperception is a complex notion. In particular, two central elements need to be distinguished - the unity of consciousness and the possibility of ascribing experiences to a numerically identical 'I'.

(2) One cannot expect the possibility of ascribing experiences to a numerically identical 'I' to explain the unity of consciousness. On the contrary, the former seems to presuppose the latter.

The Substantial Ownership Reading of the Unity of Apperception

The first interpretation of the unity of apperception that I shall consider seems in many ways to follow naturally from some of the points raised in the previous section. It has been canvassed in the literature principally by Dieter Henrich and Paul Guyer, although to radically different ends. Henrich uses it in his attempt to reconstruct the argument of the Transcendental Deduction, claiming not only that it is a plausible premise but that a plausible argument can be developed from it towards Kantian conclusions.¹⁴ Guyer, although he too puts it forward as a correct interpretation of the unity of apperception, thinks that it is fatally flawed and employs it in support of his thesis that the importance of the Transcendental Deduction has been overrated.¹⁵

As a way into the Substantial Ownership Reading, consider the 'I think' that Kant mentions so frequently. An initial question that arises is whether the 'I' is functioning as a referring term. The Substantial Ownership Reading holds that it is indeed a referring term, and, moreover, that it refers to a single numerically identical self. It is this single numerically identical self which owns all my experiences in such a way that they form a single unified series of experiences. And whenever experiences are self-ascribed it is to this single numerically identical self that they are ascribed. If pressed on the question of quite what numerical identity amounts to here the Substantial Ownership theorist will reply that a numerically identical self should be understood as an enduring subject. Henrich puts the point succinctly when he states that ". . . identity belongs to self-consciousness in so far as through self-consciousness one and the same subject is conscious of itself in all these thoughts".¹⁶

Henrich's position is based upon ". . . the formal implications of an individual 'I think'-instance" (1989, p.278). His suggestion is, in brief, that every self-conscious thought contains an implicit reference to all the other self-conscious thoughts that that same thinking subject can think:

The subject is not just the subject of actual self-consciousness. Rather it is only the subject of consciousness in so far as it is actually thought, in the thought being thought on a particular occasion, as the subject of indefinitely many other 'I think'-thoughts. (1989, p.270)

The numerical identity of the self is, on Henrich's view, a function of the fact that the same subject is thought in all these 'I think'-instances. He ascribes to Kant a conception of numerical identity derived ultimately from Christian Wolff, according to which it is a property that belongs to an entity when one can distinguish a range of different instances in which that identical thing can come to consciousness. And in this sense of numerical identity it is clearly the case that all the relevant instances of 'I think'-consciousness will involve consciousness of the same subject, as evidenced in Henrich's insistence that all actual and potential self-conscious thoughts form a totality in which ". . . precisely the same reference from them to the system as a whole, and to every other individual 'I-think' instance in it, is thought as possible" (1989, p.273).

A closer understanding of the Substantial Ownership Reading can be gained from seeing how the interpretation of the unity of apperception which it provides fits into the argument of the Transcendental Deduction. In the first section we identified two crucial features of the argument of the Deduction. The first was that the unity of apperception should be a source of a priori knowledge because the fact that it holds can be known by the subject a priori. As Kant puts the point in the first edition, "we are conscious a priori of the complete identity of the self in respect of all representations which can ever belong to our knowledge. . .".¹⁷ On the Substantial Ownership Reading this a priori knowledge is knowledge that there is a persisting identical self, the subject of all my experiences. We know, in a way that is logically independent of any experiences (and indeed is presupposed by the very possibility of having experiences) that all our experiences are states of a single self. Henrich is very clear about this:

In precisely this sense we have a priori knowledge of our identity, in so far as this consists in the sameness of the subject referred to by 'I' in indefinitely many instances of 'I think'-consciousness.¹⁸

On the Substantial Ownership view the unity of apperception comes out as a close parallel to the Cartesian cogito. Both seem to hold that a priori knowledge of the existence of a persisting subject of experience is a correlate of there being any experiences at all. We shall return to this.

The second crucial stage in the Deduction is the claim that the very possibility of such a priori knowledge of the unity of apperception depends upon the activity of a priori synthesis, and it is not difficult to see how the Substantial Ownership Reading fleshes this out. We can take Guyer's interpretation as a useful guide. He attributes to Kant the following argument.¹⁹ We have a priori knowledge that all our experiences are experiences of a single persisting self, and that depends upon a priori knowledge that our experiences are connected up in ways that are (a) independent of any particular empirical contents, and (b) that make it certain that all our experiences will be experiences of a single self. Given the further premise that experiences can be connected only by acts of synthesis, it follows that we know a priori that we perform acts of synthesis on our experiences. But since such acts of synthesis cannot depend upon the particular contents of those experiences, it follows that they must be a priori. In which case, then, a priori knowledge of the unity of apperception is possible only on the assumption of a priori synthesis. The desired conclusions will follow if satisfactory reasons can be given for thinking that the categories are the only possible rules for a priori synthesis. This argument is obviously not going to convince everybody, but it certainly provides a reasonable case for moving from the unity of apperception to the necessity of a priori synthesis.

The Substantial Ownership Reading, then, has no difficulty in accommodating these important features of the argument of the Transcendental Deduction. Nonetheless, I shall argue in the next section that there are good reasons not to accept it.

Problems with the Substantial Ownership Reading

In the first section of this paper some constraints were put forward which any satisfactory interpretation of the unity of apperception ought to meet. We have already noted that the Substantial Ownership Reading has little difficulty in accommodating the contours of the argument of the Transcendental Deduction. The other three constraints do present serious problems, however.

It was suggested that, as a transcendental argument, the Transcendental Deduction should start from a feature of experience which is uncontentious enough to be acceptable to a sceptic. There is, after all, little point in examining the necessary conditions for something whose very existence is dubious. But this is exactly what the Transcendental Deduction seems to be doing on the Substantial Ownership Reading. That reading interprets the unity of apperception in such a way that the Transcendental Deduction requires a priori knowledge of the existence of a persisting and enduring self, and this is something that no sceptic and few philosophers would be prepared to accept. Far from being an innocuous feature of experience the unity of apperception becomes a very contentious claim when interpreted in this manner. And even if true it should surely only be accepted at the conclusion of an argument, rather than as a premise.

Guyer himself argues that on his interpretation the principle of the unity of apperception is false, and he concludes that the Transcendental Deduction is a doomed project:

The assumption that even the candidacy of any representation for empirical knowledge presupposes that it already be assigned to the unity of apperception seems to be the deepest reason for Kant's assertion of the a priori certainty of apperception. . . Nevertheless, this assumption still does not yield a philosophically sound argument for the premise which Kant requires to ground his direct inference from apperception to a priori synthesis, and thus to a priori categories. This is because Kant has failed to establish that I must in fact know - a fortiori, be certain - that I have really had all of a putative series of representations through some period of my continued existence in order to investigate their possible empirical significance.²⁰

Guyer's objection is that Kant has not ruled out - and indeed cannot rule out - the possibility that when a subject starts empirically investigating the connections between a series of representations he might find it necessary to reject his putative ownership of one or more of them.²¹ This is rather obscure as it stands, principally because we are not offered any account of what it would be like to discover that one did not really have an experience that one thought one had. The only example Guyer gives of how this might occur is that ". . . I might be able to make sense of the observations I

have recorded in a lab book only if I admit that one of my notes describes an observation I could never have made, thus a representation I could never have had".²² This does not seem an appropriate illustration, however, because the example as described is too indirect to be really a case of rejecting the putative ownership of a representation. It would be closer to the matter in hand if what was at stake was making sense of the observations in the lab book by denying that one or more of them was actually in the lab book, and this does not seem to be coherent. Clearly cases of hallucination or extreme perceptual error do not fit the bill either, because the discovery that one is hallucinating does not threaten one's putative ownership of the hallucinatory experience.

The details of Guyer's argument against the principle of apperception (as he interprets it) are bizarre, but the conclusion is surely correct. It is indeed hard to see how there could be a priori knowledge of a persisting numerically identical self of the sort that the Substantial Ownership Reading requires. Better grounds for this negative conclusion have been elegantly put forward by Quassim Cassam (1989). As Cassam notes, the inference to a persisting numerically identical self can seem unproblematic if the situation is described in a particular (question-begging) way - ". . . if R1, R2 and R3 are all representations given to me, then it is indeed the numerically identical self which is conscious of these representations".²³ But, as he points out, when the matter is described in a way that does not presuppose the persistence of a numerically identical self, the fallacious nature of the inference becomes clear. If a subject ascribes representation R1 to himself at one time, a subject ascribes R2 to himself at a later time, and a subject ascribes R3 to himself subsequently, each subject will think of himself as 'I', but there is no reason to conclude from that that it is the same subject on each occasion.

The argument which Cassam puts forward is surely correct. What is particularly interesting about it is that it bears a very close resemblance to an argument which Kant himself puts forward against the rational psychologist in the Paralogisms (which brings us on to the second constraint discussed earlier). Discussing the Third Paralogism in the first edition Kant states:

Since we reckon as belonging to the identical self only that of which we are conscious, we must necessarily judge that we are one and the same throughout that whole time of which we are conscious. We cannot, however, claim that this judgement would be valid from the standpoint of an outside observer. For since the only permanent appearance which we encounter in the soul is the representation 'I' that accompanies and connects them all, we are unable to prove that this 'I', a mere

thought, may not be in the same state of flux as the other thoughts which, by means of it, are linked up with one another.²⁴

Put in a more contemporary idiom, Kant's point seems to be that constancy in the sense of the expression 'I' can be no guarantee of constancy in its reference.²⁵ On the previous page Kant supports this with a suggestion very much like that put forward by Cassam:

Despite the logical identity of the 'I', such a change may have occurred in it as does not allow of the retention of its identity, and yet we may ascribe to it the same-sounding 'I', which in every different state, even in one involving change of the subject, might still retain the thought of the preceding subject and so hand it over to the subsequent subject.²⁶

It is perfectly conceivable, Kant is saying, that there could be a switching of subjects masked by the logical identity of the 'I think'. Moreover, not only is this possible, it also seems to be compatible with the principle of the unity of apperception, in the sense that a series of experiences pertaining to such a series of switched subjects could nonetheless correctly be described as subject to the unity of apperception.

When he refers to "the logical identity of the 'I think'" in this passage Kant means several things. As his mention of the "same-sounding 'I'" reveals, one aspect of this logical identity is simply that even if subjects are switched the same form of words will be employed in each case of self-ascription. More significant, though, is the point that the expression 'I' has the same sense even if it is being employed by what ultimately turn out to be different switched subjects. Roughly, irrespective of who is employing it, the sense of the expression 'I' will always be that any token of it picks out the individual who produces it. But perhaps the most important point here is that even if subjects are switched so that different subjects are referred to over a period time this is not necessarily something that any of those subjects will be in a position to detect. This emerges particularly clearly in the footnote that Kant appends to the end of the quoted passage:

If, then, . . . we postulate substances such that the one communicates to the other representations together with the consciousness of them, we can conceive a whole series of such substances of which the first transmits its state together with its consciousness to the second, the second its own state with that of the preceding substance to the third. . . The last substance would then be conscious of all the states of the previously changed substances, as being its own states, because they would have been transferred to it together with the consciousness of them. And yet it would not have been one and the same person in all these states.²⁷

The point here is that there is nothing in the content of such communicated memories that will reveal the change of subject, because each subject will interpret the communicated memories as if it had itself experienced them. The content of such memories does not individuate the subject who originally had the experiences. It just makes it natural to assume that the subject who had the experiences is the same as the subject who is having the memories. And, as the work of Parfit and Shoemaker has brought out, this does not guarantee the truth of such an identification.²⁸

It seems clear that a situation in which such switching of subjects occurred could not correctly be described as involving the persistence of a single numerically identical self. We can illustrate this by considering an example of such switching much discussed in contemporary treatments of personal identity. This is the case of fission, made popular by Wiggins (1967, p.50). Suppose that while he is asleep a person's brain is divided into two identical halves, each of which is successfully transplanted into a body very similar to the body that he had before fission. Each of the resulting individuals, when they are woken up, believes that he is the original person, seems to remember living that person's life and has all and only that person's character traits. Suppose, moreover, that just after the fission the two individuals are kept apart and neither told that fission has taken place. This seems to be an example of the sort of switching situation that Kant is discussing in the context of the Third Paralogism, because it incorporates the communication of "representations together with the consciousness of them" in such a way that "it would not have been one and the same person in all these states". Each individual would think that they really were the original person (and hence there would be 'logical identity'). Nonetheless, it obviously cannot be the case that the original person continues in both, and, since they are ex hypothesi indistinguishable to all intents and purposes, it seems implausible that one would be the original person but not the other.²⁹ So, despite the continuity of psychological characteristics there is no continuity of the person. And nor, of course, could there be numerical identity, for the simple reason that numerical identity is a one-one relation. As a one-one relation it would have to hold between the original person and one of the two fissioned individuals. But as we have seen they are indistinguishable, and hence there would be no reason for choosing one rather than the other.³⁰ So, in the case of neither of them would there be numerical identity with the original person.

Nonetheless, Kant is very definite that the unity of apperception would hold in such a switching case. There would be two unified consciousnesses subject to the unity of apperception after fission. One would be the psychological characteristics of the original person, together with the thoughts and experiences of one of the fissioned individuals. The second would be the psychological characteristics of the original person, together with the thoughts and experiences of the other fissioned individual. Each of these unified consciousnesses would be subject to the unity of apperception, although neither of them would be numerically identical with the original person. This is a very good reason to think that the unity of apperception cannot straightforwardly be understood in terms of the numerical identity of the self, as the Substantial Ownership Reading claims.

Another reason for being dubious about this reading is the fact that it gives the Transcendental Deduction a very dubious premise. Closer attention to Kant's discussion of the Third Paralogism should bring home why the Substantial Ownership Reading is unacceptable within the parameters of the Critique of Pure Reason. In the First Paralogism Kant discusses the following mistaken inference about the substantiality of the soul:

That, the representation of which is the absolute subject of our judgements and cannot therefore be employed as determination of another thing, is substance.

I, as a thinking being, am the absolute subject of all my possible judgements, and this representation of myself cannot be employed as predicate of any other thing.

Therefore I, as thinking being (soul) am substance.³¹

What is wrong with this, Kant claims, is that it misuses the concept of substance. Like all the other categories, the pure concept of substance can be applied only empirically. As a category it is a principle for the synthesis of intuitions, thus clearly requiring empirical data on which it can be exercised to yield knowledge, and in the absence of such data there is only the illusion of a legitimate application of the category.³² But it is precisely such an illegitimate application of the category that is being undertaken by the rational psychologist, because he is undertaking to deduce the substantiality of the self from the category alone. As Kant points out, the category of substance is applicable only when an object is given in experience as permanent.³³ No such object is given to inner sense, however.³⁴ There is no introspective awareness of an enduring self, and inner sense yields only a flux of appearances to which the category of substance cannot be applied.³⁵

There is a general point to be extrapolated from this. When the unity of apperception is interpreted according to the Substantial Ownership Reading, the argument of the Transcendental Deduction demands a priori knowledge of a persisting numerically identical self. As the discussion of the First Paralogism makes clear, however, there can be no such a priori knowledge, because the unity of apperception does not involve the sort of intuition of an enduring self that would be required to support any claim to knowledge of such a self. To the philosophical objections to which we have already noted that the Substantial Ownership Reading is susceptible, then, must be added the fact that it is highly implausible that Kant would have held a view which, as he expressly argued in the Paralogisms, contravenes perhaps the most basic principle of the Critical Philosophy - the principle that all cognition involves both the receptivity of intuition and the spontaneity of the understanding.

The Formal Ownership Reading

We seem, then, to have reached a position where we need either to abandon the Transcendental Deduction because its fundamental premise seems so flawed, or to find a new way of interpreting the principle of the unity of apperception. In this section I will take the latter path. My strategy will be to begin by examining two central passages which seem initially to support the Substantial Ownership Reading, and to argue that these passages will actually support a weaker interpretation, which I term the Formal Ownership Reading. In the next section I go on to argue that the Formal Ownership Reading also respects the arguments which Kant offers in support of the principle of the unity of apperception, and preserves the unity of apperception from the attacks on rational psychology in the Paralogisms.

I begin, then, by citing the following passages, one from the first edition and one from the second. The first occurs in the so-called Objective Deduction (Section 3 of the first edition version). It reads as follows:

- (1) We are conscious a priori of the complete identity of the self in respect of all representations which can ever belong to our knowledge, as being a necessary condition of the possibility of all representations. For in me they represent something only in so far as they belong with all others to one consciousness, and

therefore must be at least capable of being so connected. This principle holds a priori, and may be called the transcendental principle of the unity of all that is manifold in our representations, and consequently also in intuition.³⁶

Guyer puts this passage forward as incontrovertible support for the Substantial Ownership Reading. He thinks the same of the next passage, which is from the second edition:

(2). The thought that the representations given in intuition one and all belong to me, is therefore equivalent to the thought that I unite them in one self-consciousness, or can at least so unite them; and although this thought is not itself the consciousness of the synthesis of the representations, it presupposes the possibility of that synthesis. In other words, only in so far as I can grasp the manifold of the representations in one consciousness do I call them one and all mine. For otherwise I should have as many-coloured and diverse a self as I have representations of which I am conscious to myself. Synthetic unity of the manifold of intuitions, as generated a priori, is thus the ground of the identity of apperception itself, which precedes a priori all my determinate thought.³⁷

These two passages are amongst the most important and frequently quoted discussions of the unity of apperception in the Critique, and they are certainly touchstones for any adequate interpretation of Kant's pronouncements on the unity of apperception.

The first thing to notice about them is that neither passage directly asserts that the self is numerically identical in the way that the Substantial Ownership Reading suggests, although passage (1) states that we are 'conscious a priori' of the identity of the self. On the Substantial Ownership Reading, of course, this is taken to imply the fact of numerical identity. I shall suggest that this is unwarranted, but before doing so there are a couple of other points to be made. In passage (1) Kant states that the principle of the unity of apperception holds a priori, but on this occasion he glosses the principle in terms of the hypothetical necessity of all my representations being capable of being connected up in one consciousness. When glossed in this way the principle comes out as a claim about the unity of consciousness, rather than about the identity of the self. And even if the principle's holding a priori involves, as we suggested in the first section, the subject's knowing a priori that it holds, this still comes out as a claim about the unity of consciousness, because what the subject knows a priori is that it holds a priori that all his representations are capable of being connected up in one consciousness. The same seems to be true of passage (2). The crucial sentence there is the final one:

Synthetic unity of the manifold of intuitions, as generated a priori, is thus the ground of the identity of apperception itself, which precedes a priori all my determinate thought.

Two different elements are explicitly distinguished here. The first is the unity of the manifold of intuitions which is, Kant states, generated a priori. This is presumably equivalent to the claim in passage (1) that all the subject's representations are capable of being connected up in one consciousness. The second, which Kant describes as 'the unity of apperception itself', is an a priori awareness of that connectibility. Both of these are claims about the unity of consciousness.

So, at least as far as passages (1) and (2) are concerned, the weight of the case for the Substantial Ownership Reading is born by the putative equivalence of being conscious a priori of the numerical identity of the self, and the actual existence of a numerically identical self. The key to the interpretation of the unity of apperception that I am putting forward is the suggestion that these are not equivalent at all. In particular, whereas it is clear that a numerically identical self cannot be said to exist in the switching cases which Kant discusses a propos the Third Paralogism, it does make sense to speak of being conscious a priori of the numerical identity of the self in such a situation. Alternatively put, the a priori consciousness of numerical identity that is implicated in the unity of apperception does not entail the existence of such a numerically identical self.³⁸

But what can a priori consciousness of numerical identity be, if it does not require the actual existence of a numerically identical self? Illumination can be gained here if we reflect on the parallel between Kant's treatment of the Paralogisms and his discussion of the Antinomies. As is well-known, Kant's resolution of the Antinomies involves a distinction between regulative and constitutive conceptions of the Principles of Reason.³⁹ Constitutive principles are principles which yield knowledge about the ultimate nature of reality, in contrast to regulative principles which cannot give such knowledge but instead serve a crucial role in governing enquiry. So, in the case of the principle that there must be a completed totality of conditions for every phenomenon, this should not be taken as a fact about the world (because it makes claims beyond the scope of any possible experience) but rather as ". . . a rule, prescribing a regress in the series of the conditions of given experiences, and forbidding it to bring the regress to a close by treating anything at which it may arrive as absolutely unconditioned".⁴⁰ The cosmological principle of totality offers a rule which all

empirical enquiry must obey (the rule that one must always seek the conditions of any given phenomenon), but it does not license the sort of speculation about the world as a spatio-temporal totality that is involved in, for example, the cosmological argument for the existence of God.

To return to the principle of the unity of apperception, we clearly want to retain the idea that the unity of apperception involves the potential ascription of all the members of a suitably unified set of experiences to a single numerically identical subject. This is to go some of the way with the Substantial Ownership Reading, but, as mentioned earlier, there are good reasons for not wanting to go as far as claiming that this involves affirming the existence of a single numerically identical subject. The difficulty is to see how one can have the former without being *ipso facto* committed to the latter. But we can do this if we view the principle of the unity of apperception as a regulative principle. On this view it must be possible for all members of a suitably unified set of experiences to be ascribed to a single numerically identical subject, just as the regulative principle of totality demands that one must be capable of seeking the conditions of any phenomenon one might encounter. But this does not entail that there really is a single numerically identical subject, any more than the regulative principle of totality means that there really is a completed series of conditions. It might well be the case, as Kant hypothesises, that what seems to be a single numerically identical subject turns out to be a causally connected series of distinct subjects, but this is beyond the ken of the subject(s) and does not contravene the principle of the unity of apperception, provided that it is possible for all the relevant experiences actually to be ascribed to a single subject. The switching case can be fatal to the principle of the unity of apperception only if the principle is interpreted constitutively. Moreover, the reason for the principle of the unity of apperception being a regulative principle is very much parallel to the reason for the cosmological principle of totality being regulative. In both cases the crucial point is that, if interpreted constitutively, they contravene Kant's fundamental principle that knowledge is constrained by the bounds of possible experience.

On this interpretation, then, what the unity of apperception requires is simply the possibility of all experiences being ascribed to a single numerically identical subject. It does not require that such ascription be true. The persistence of an enduring subject is postulated in a regulative manner, as a

presupposition rather than a claim to knowledge. Certainly, the 'I' is functioning here as a referring expression. Moreover, on each occasion of its use 'I' will indeed pick out the subject who thinks or utters it. But it will not necessarily be true that all the 'I'-thoughts in a suitably connected series will refer to the same numerically identical subject. The unity of apperception is perfectly compatible with there being an unnoticed succession of subjects. With this in mind, then, we can return to some of the comments about ownership raised in section 2. It was suggested there that it is Kant's view that there is more to the unity of apperception than experiences being connected up in certain ways. Instead they must be 'owned' by a single subject. We are now in a position to see what this amounts to. All it means is that a single numerically identical subject must be posited as the owner of experiences. When a set of experiences forms a single series all its component experiences should be ascribable to a single subject.⁴¹

There may be a query, however, about the way in which the Formal Ownership Reading has been formulated. It has been suggested that what is important is that all the members of a suitably connected series of experiences should be ascribable to a single numerically identical subject, and that this is possible both when there really is such a single numerically identical subject, and when there is the sort of undetected switching of subjects that Kant discusses in the context of the Third Paralogism. In both of these cases the personal pronoun functions as a referring expression, the difference between them being that in the latter there is no constancy of reference across the switching of subjects, because different occurrences of the 'I think' within a single suitably connected series of thoughts and experiences can refer to different subjects. But this seems to neglect one very important possibility, which is that there might not be a subject there at all. This possibility which we can term the No Ownership situation, seems as admissible on the Formal Ownership Reading as is the possibility of switched subjects. In such a situation, although a single numerically identical subject would be posited as the 'owner' of experiences, there would in reality be no such subject. As Strawson puts it, in such a situation: "It is only a linguistic illusion that one ascribes states of consciousness at all, that there is any proper subject of these ascriptions, that states of consciousness belong to, or are states of, anything at all" (1959, p.94).

The crucial issue posed by the possibility of the No Ownership Situation is this. If, as has been suggested, it is possible for the unity of apperception to hold even when there is a switching of subjects, then it is compelling that it must also be possible for it to hold when there is no subject at all. This seems to come into conflict, however, with the thought that 'I' is functioning as a referring expression. How can 'I' function as a referring expression, if there is nothing to which it refers? Certainly, if it is true that the Formal Ownership Reading forces us to reject the view that 'I' is a referring expression, then this will be a weighty consideration against it, for two reasons. The first is exegetical. Kant seems to have held that 'I' does in fact refer. That seems the most natural way to understand this important footnote to the second edition Paralogisms:

The 'I think' expresses an indeterminate empirical intuition, ie perception (and thus shows that sensation, which as such belongs to sensibility, lies at the basis of this existential proposition). . . An indeterminate perception here signifies only something real that is given, given indeed to thought in general, and so not as appearance, nor as thing in itself (noumenon), but as something which actually exists, and which in the proposition "I think" is denoted as such.⁴²

Any view, then, which makes it a Kantian principle that 'I' cannot be a referring expression departs significantly from the textual evidence. The second point is that it just seems overwhelmingly plausible that 'I' is a genuine referring expression. Any view which denies this (as Anscombe 1975 does) forces us to revise some very basic beliefs about how we think about ourselves.

So, the crucial question is whether the Formal Ownership Reading of Kant can retain the idea that 'I' is a referring expression while countenancing the possibility that the No Ownership situation might hold. There seem to be two broad strategies available here. The first would be to give a theory on which 'I' could be a referring expression even though there was never anything to which it referred. The second would be to suggest that, even if the No Ownership situation did hold, 'I' could still be a successfully referring expression because what it refers to is still there even when there is no subject of experiences at all.

The first strategy seems doomed from the start. Although in the case of most referring expressions it is possible to make a distinction between successful instances of reference and unsuccessful ones, this is so only because the latter are parasitic on the former. Referring expressions just are expressions employed to identify objects, actions, events and so on. Although

we can in most cases make sense of the idea that a referring expression fails to perform such an identification this is because we know what it is that that expression usually serves to identify. In contrast, the suggestion that there might be a referring expression which never picks anything out, because there is nothing for it to pick out and there never will be anything for it to pick out, just seems to be incoherent. In such a situation we would have a referring expression only in the highly misleading sense of an expression that shares certain grammatical features with genuine referring expressions (those that are used to identify objects, actions and events). There seems, then, no possibility of dealing with the possibility that the No Ownership situation might hold by arguing that 'I' could be a referring expression even if there could be nothing to which it referred.

Fortunately, the second strategy is more promising. We can approach it by considering the interpretation of the unity of apperception put forward by Patricia Kitcher (most comprehensively in her 1990, but also in her 1984). On Kitcher's interpretation, Kant's principal concern in the Transcendental Deduction is to answer Hume's attack on the notion of personal identity. She sees Kant as responding to Hume's denial that there could be any real or necessary connections between experiences with the counter-claim that the attribution of experiences and mental states in general requires understanding those mental states as connected together in virtue of certain synthetic relations of contentual dependence. As part of this interpretation Kitcher rejects traditional understandings of the unity of apperception. In particular she rejects the thesis that the unity of apperception involves the ascription or potential ascription of mental states to a subject or self that exists over and above the contentually connected set of mental states. Clearly, then, she is outlining a version of the No Ownership situation, and as such she has to explain what it is that 'I' refers to.

Her answer is this:

According to Kant's doctrine of apperception, we must acknowledge the existence of a thinking self in that we recognize that all representations, which are possible as representations, must be regarded as belonging to a contentually interconnected system of mental states. . . His reflections on the nature of mental states show that we can acknowledge something to be a mental state only if we can acknowledge the existence of, not a mere bundle, but a synthetically connected set of mental states (1984, p.118)

Kitcher's position, then, seems to be that we can retain the idea that 'I' is a referring expression in the No Ownership situation by holding that 'I' refers to a contentually connected set of mental states. If

she is right then this provides one way of getting round a major difficulty for the Formal Ownership Reading, because it allows the reading to accommodate the possibility that there might not be a subject at all.⁴³ In such a situation a token utterance of 'I' would refer to the suitably unified set of thoughts and experiences of which it is a member.

One misgiving one might have with this suggestion, however, is that reference to a contentually connected set of mental states is not what we are trying to achieve when we employ the first person pronoun in ascribing mental states to ourselves. It seems fair to assume that most people think that 'I' refers to a thinking self (of some description) that is the subject of all their thoughts and experiences, and that exists independently of those thoughts and experiences. So, one might think, either 'I' does refer to such a thinking self or it does not refer at all. Given that, on the No Ownership situation, there would be no such thinking self, it seems to follow that 'I' cannot be a referring expression. But this view is not compelling. It rests on the assumption that (in this case, at least) we can refer only to what we intend to refer to, where the content of our intended reference is given by all or most of the beliefs that we have about the intended referent. This might be true in some cases of reference, particularly those where reference is fixed by definite descriptions. But there are plenty of cases of reference for which it does not hold, and the first person pronoun is one of these. The reference for the first person pronoun is given by something like the token-reflexive rule that any token of 'I' refers to the individual that produces it. And the reference of the first person pronoun will be fixed in this way whether or not the beliefs that one might have about the relevant individual are true or false. So, if as Kitcher suggests the individual who produces a given token of 'I' turns out to be just a suitably connected set of thoughts and experiences, it is to that suitably connected set of mental states that a given token of 'I' refers.

To tie the threads of the argument together, then, the Formal Ownership Reading that I am proposing interprets the unity of apperception in a regulative manner, as imposing the following two requirements. The first is that the set of thoughts and experiences subject to the unity of apperception be appropriately unified. The second is that it be possible for all the members of that set of suitably connected experiences to be ascribed to a numerically identical subject. It is vital to the Formal Ownership Reading that this second requirement does not require that any such

ascriptions be true. It does not require that all such ascriptions should be to the same subject, or even that there should be a subject there at all. Nonetheless, this does not mean that we have to revise the idea that the 'I' in the 'I think' that must be able to accompany all my representations is a referring expression. On the contrary, even in the situation which I have termed the No Ownership situation, where there is no subject of thought or thinking self over and above the relevant thoughts and experiences, the 'I think' can still be construed as referring to the appropriately connected set of thoughts and experiences.

In Defence of the Formal Ownership Reading

In this section I will argue that the Formal Ownership Reading is the right way to understand the unity of apperception as developed in the Critique of Pure Reason. In particular, I shall argue that it offers a way of understanding the unity of apperception that respects all four of the constraints discussed earlier. I begin by arguing that the criticisms of rational psychology in the Paralogisms rule out the two alternatives to the Formal Ownership Reading. These two alternatives are the Substantial Ownership Reading, which we have already discussed, and its opposite, the No Ownership Reading which denies that the unity of apperception involves any potential ascription of thoughts and experiences to a numerically identical self because no thinking self or subject exists over and above the suitably unified set of thoughts and experiences.⁴⁴ I then go on to discuss the remaining three constraints.

Two vital points emerge from the Paralogisms. The first is that Kant holds that an undetected switching of subjects is compatible with a set of experiences and thoughts that conform to the unity of apperception. The second is that the self-knowledge implicated in the unity of apperception must obey the 'highest principle of synthetic judgements' that all cognition involves both the receptivity of intuition and the spontaneity of the understanding. As we have seen the first of these points has serious consequences for the plausibility of the Substantial Ownership Reading, because it entails that the unity of apperception can hold of a set of thoughts and experiences even when there is an undetected succession of subjects, as opposed to the persistence of a single numerically identical

self. Nonetheless, it still leaves the No Ownership Reading in play, because the compatibility of the unity of apperception with an undetected succession of subjects suggests that it would be equally compatible with there being no independently existing subject there at all.

What we have to ask of the No Ownership Reading, however, is what ground it has for the claim that there is no independently existing subject over and above the set of contentually connected thoughts and experiences. Presumably this is known. It is not merely a conjecture. But, as the second point emerging from the Paralogisms makes clear, any such knowledge claim requires a supporting intuition, or intuitions. What intuition or intuitions could support the claim that there is no independently existing subject? The only possible candidates are the introspective intuitions of inner sense, about which Kant makes the following celebrated comment:

Consciousness of self according to the determinations of our state in inner perception is merely empirical, and always changing. No fixed and abiding self can present itself in this flux of inner appearances. (A107)

Clearly, then, introspection both does not and cannot provide any evidence of the existence of independently existing subject. But this hardly warrants the conclusion that there is no such fixed and abiding independently existing self. Quite the contrary. Surely the conclusion it warrants is the rather different one that, whether there is or is not an independently existing 'fixed and abiding self', it will not be the sort of thing that we could possibly encounter in introspection. Of course, for Hume the fact that a self is not the sort of thing that we could possibly encounter in introspection is a good reason for denying that there could be such a thing. But Kant is not a Humean empiricist, and Hume's reasoning is unacceptable on Kant's principles. Kant is entitled to conclude only that, since we cannot encounter such a fixed and abiding self in introspection, we cannot in principle have knowledge of it. And this, of course, is not the same as saying that we can know that no such self exists.

It would seem, then, that there can be no more epistemic warrant for the No Ownership Reading that there is no independently existing self than there is for the Substantial Ownership Reading that there is a single such self. Both claims are illegitimate. The conclusion to be drawn from this is that, on Kant's principles, there can be no epistemic warrant for any definite claims about the existence or non-existence of such a self. What is permissible, however, is the agnosticism of the Formal

Ownership Reading which avoids substantive claims altogether, because all it requires is that it should be possible for any member of a suitably connected set of experiences to be ascribed to a numerically identical subject.

This shows, I think, that the Formal Ownership Reading does not fall foul of the Paralogisms in the way that the Substantial Ownership and No Ownership Readings do. It remains to be seen, however, whether it satisfies the remaining three constraints upon a satisfactory interpretation of the unity of apperception. In the remainder of this section I shall discuss the connected questions of whether the unity of apperception under the current interpretation is a basic enough feature of experience to be plausibly taken as a premise, and whether it is supported by the defence which Kant puts forward for the principle of the unity of apperception. I then go on to discuss the relation between the unity of apperception and the overall argument of the Transcendental Deduction.

How basic is the unity of apperception on the interpretation I am putting forward? Clearly, it is not so basic that it is inconceivable that a creature could have something which might pass as experience and yet fail to satisfy it. There is no need for it to be that basic, however. Transcendental arguments like the Transcendental Deduction attempt to establish conclusions which hold of necessity, but this does not mean that their premises have to be necessary. The form of necessity which transcendental arguments confer upon their conclusions is hypothetical necessity, as opposed to absolute necessity.⁴⁵ The Second Analogy, for example, attempts to establish the necessity of the law of cause and effect by arguing that having causally structured experience is a necessary condition of being capable of objective time-determination. This does not imply the impossibility of experience which is not causally structured, but it does mean that the universal applicability of the law of cause and effect is a hypothetically necessary feature of any form of experience that supports objective time determination. Arguments establishing hypothetically necessary conclusions do not require absolutely necessary premises, and so the fact that the unity of apperception does not seem to be absolutely necessary is not an obstacle.

If absolute necessity is not a prerequisite for premises in transcendental arguments, however, a certain immunity to sceptical challenge definitely is. If the Transcendental Deduction of the Categories is to provide a convincing argument for the a priori necessity of synthesis according to

the categories by arguing that such synthesis is required by the possibility of experience which is subject to the unity of apperception, then the existence of the unity of apperception must be uncontroversial in the following important sense. If the unity of apperception appears to hold, then it must hold. It should not be possible for a determined sceptic to undercut the whole argument by denying the truth of the premise. We have already noted that such a sceptical line of attack is available on the Substantial Ownership Reading. It does not seem to be possible, however, on the Formal Ownership Reading. The fact that all the members of a suitably connected series of experiences can be ascribed to a single subject is all that is required for the unity of apperception to hold. They might be incorrectly ascribed to a single subject, but there is no room to take issue with the fact of the ascription (and, as was stressed in the previous section, the existence of such error is compatible with 'I' being a referring expression).

Nonetheless, as stressed above, the fact that the principle of the unity of apperception is a crucial premise in the Transcendental Deduction does not mean that it is a brute fact not susceptible to any further explanation. On the contrary, Kant is at pains to establish that it is itself hypothetically necessary if experience is to take a certain form and he provides at least two reasons for thinking so. I will discuss only the most prominent of these. We find it in a famous passage from the second edition already quoted:

It must be possible for the 'I think' to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying that the representation would be impossible, or at least would be nothing to me.⁴⁶

Similar considerations are adduced in the first edition:

What is first given to us is appearance. When combined with consciousness, it is called perception. (Save through its relation to a consciousness that is at least possible, appearance could never be for us an object of knowledge, and so would be nothing to us; and since it has in itself no objective reality, but exists only in being known, it would be nothing at all.)⁴⁷

And even more clearly in an important footnote a few pages earlier:

All representations have a necessary relation to a possible empirical consciousness. For if they did not have this, and it were altogether impossible to become conscious of them, this would practically amount to the admission of their non-existence. But all empirical consciousness has a necessary relation to a transcendental consciousness which precedes all special experience, namely the consciousness of myself as original apperception.⁴⁸

These three passages seem to add up to the following line of argument from the possibility of consciousness to the necessity of apperception. What makes a representation 'cognitively significant' is its being potentially conscious. Without this it would be "nothing to us".⁴⁹ But a representation is only potentially conscious in so far as it can play a part in empirical knowledge. Empirical knowledge, however, never involves single representations but rather combinations of representations. It is this that brings in the unity of apperception, because representations can be combined only if they are elements of a single unified consciousness, where this involves the possibility of ascribing experiences to a single numerically identical self. Now, this is not the place to argue about the soundness of this argument. It does seem clear, however, that if it does actually provide good grounds for believing in the necessity of the unity of apperception, it will do so only on the Formal rather than the Substantial Ownership Reading.⁵⁰ It is wildly implausible to suggest that the very possibility of combining representations rests upon the existence of an enduring self, but there does seem to be scope for arguing that it rests upon the presupposition of an enduring self.

The crucial issue that remains to be discussed is what happens to the argument of the Transcendental Deduction when the principle of the unity of apperception is read according to the Formal Ownership Reading. In section (1) it was suggested that there are two central features of the unity of apperception which underpin the argument. The first is that it must be known to hold a priori, and the second is that it should be possible to argue that the unity of apperception must be underwritten by an a priori synthesis.

The first of these is not a problem. The Formal Ownership Reading can construe a priori knowledge of the unity of apperception in terms of knowledge that all the members of a suitably unified set of experiences must be ascribable to a single enduring subject. This is a fairly plausible premise in itself, as well as meshing well with the various passages quoted in section 2. The second is a little more difficult, given the implausibility of expecting all experiencing subjects to have views about the necessary conditions of their having experience at all. Even so, there is a case to be made for saying that this sort of a priori knowledge does not have actually to be possessed (any more than one needs to be au fait with Euclidean geometry just because that is a priori), but it can be

acquired, perhaps through apprehending a transcendental argument like that put forward at the end of the previous section, and in acquiring it no recourse to experience is demanded.

This last point could be expanded further. The principal difficulty encountered by the Substantial Ownership Reading is that it demands a form of knowledge which contravenes 'the highest principle of all synthetic judgements' that knowledge claims must be confined within the bounds of possible experience. This does not arise, however, on the Formal Ownership Reading. The fact that all the members of a set of experiences are ascribable to a single enduring subject is accessible to inner sense, and the fact that this is necessarily so can be known through a transcendental argument. Both of these are perfectly respectable sources of knowledge on Kant's principles.

The crucial step in the Transcendental Deduction, however, must be the claim that the unity of apperception requires a priori synthesis. As a first step to seeing how the Formal Ownership Reading might cope with this we need to return to one of the points made in the preliminary discussion - the importance of distinguishing between the unity of consciousness, on the one hand, and the possibility of ascribing experiences to a single persisting self on the other. It was suggested that the possibility of ascribing experiences to a single persisting self cannot explain the fact that experiences are unified in such a way as to make it appropriate to talk of a unity of consciousness. Instead, the possibility of such ascription seems to presuppose such a unity. Now, it is this presupposition that I think serves as the linchpin of the Transcendental Deduction. It allows the broad outline of the Transcendental Deduction to be reconstructed as follows. The premise is the principle of the unity of apperception, or, more specifically, a priori knowledge that a single numerically identical subject is a necessary presupposition of experience. But experiences can only be ascribed to a numerically identical subject if they are connected up in certain ways that create a unified consciousness. This is not a sufficient condition for the unity of apperception, but it is certainly a necessary condition. The next stage of the argument would be to note that all connections between experiences are products of synthesis. This is one of Kant's fundamental premises and receives its clearest exposition in the Preliminary Remarks that precede the Subjective Deduction. The argument would be complete, then, if it can be shown that such synthesis must be a

priori. Suppose we accept that the connections which must hold within a unified consciousness should not be contingent upon the specific details of a subject's experience, but instead must hold in virtue of more general structural features of experience, such as the fact that experience is structured according to the law of cause and effect. Then all that needs to be shown is that experience can only be structured in these general ways as a consequence of a priori synthesis. It is arguments in this general direction that Kant offers in the *Analytic of Principles*.⁵¹

Of course, it is a consequence both of my interpretation of the unity of apperception and of my reconstruction of the Transcendental Deduction that much of the real work is done by the connections that hold between experiences, by the fact that a set of experiences is 'suitably unified', and it might be felt that this is a cheat. On the contrary, however, I think it is an advantage, precisely because it focusses attention on the detailed arguments in the *Analytic of Principles*. There is no getting round the fact that the possibility of ascribing the members of a set of experiences to a single enduring self cannot explain the unity of that set of experiences. We must look to the internal connections of that set to explain the unity. For example, we will have to explore the conditions of time determination, considering what it is for a set of experiences to yield an awareness of an objective temporal order. It would be too much to expect all this to take place in the short span of the Transcendental Deduction. On the contrary, the Transcendental Deduction merely sets the scene for this more ambitious enquiry, to which Kant first points the way in the *Schematism*, and which he then undertakes in the *Analytic of Principles*.⁵²

References

References to the *Critique of Pure Reason* are to the translation by Norman Kemp Smith (London, MacMillan, 1929) of Immanuel Kant, *Kritik der reinen Vernunft*.

Page references are given in the usual manner, 'A' referring to the first (1781) edition and 'B' to the second (1787) edition.

- Anscombe, Elizabeth (1975), 'The First Person', in Samuel Guttenplan (Ed.) Mind and Language. Oxford: Oxford University Press.
- Bermúdez, José Luis (forthcoming), 'Scepticism and the Justification of Transcendental Idealism', Ratio.
- Broad, C. D. (1978), Kant. Cambridge: Cambridge University Press.
- Cassam, Quassim (1989), 'Kant and Reductionism', Review of Metaphysics 72, pp.72-106.
- Cassam, Quassim (1993), 'Inner Sense, Body Sense, and Kant's Refutation of Idealism', European Journal of Philosophy 1, pp. 111-127
- Guyer, Paul (1979), review of Henrich 1979, The Journal of Philosophy 86, pp.151-167.
- Guyer, Paul (1980), 'Kant on Apperception and A Priori Synthesis', American Philosophical Quarterly 17, pp.205-212.
- Guyer, Paul (1987), Kant and the claims of knowledge. Cambridge: Cambridge University Press.
- Guyer, Paul (1992), 'The transcendental deduction of the categories', in Paul Guyer (Ed.), The Cambridge Companion to Kant. Cambridge: Cambridge University Press.
- Harrison, Ross (1982), 'Transcendental Arguments and Idealism', in G. Vesey (Ed.), Idealism Past and Present. Cambridge: Cambridge University Press.
- Henrich, Dieter (1976), Identität und Objektivität: eine Untersuchung zu Kants transzendentaler Deduktion. Heidelberg, 1976.
- Henrich, Dieter (1989), 'The Identity of the Subject in the Transcendental Deduction', in Eva Schaper and Wilhelm Vossenkuhl (Eds.) Reading Kant. Oxford: Basil Blackwell.

- Hurley, S. L. (1994), 'Kant on Spontaneity and the Myth of the Giving', *Proceedings of the Aristotelian Society* 94, pp.137-164.
- Kitcher, Patricia (1984), 'Kant's Real Self', in Allen W. Wood (Ed.) Self and Nature in Kant's Philosophy. Ithaca: Cornell University Press.
- Kitcher, Patricia (1990), Kant's Transcendental Psychology. Oxford: Oxford University Press.
- Parfit, Derek (1971), 'Personal Identity' in Philosophical Review 80, pp.3-27.
- Parfit, Derek (1984), Reasons and Persons. Oxford: Oxford University Press.
- Sellars, Wilfred (1979), '... this I or he or it (the thing) which thinks. . .', in Wilfred Sellars, Essays in Philosophy and its History. D. Reidel.
- Shoemaker, Sydney (1970), 'Persons and Their Pasts' in American Philosophical Quarterly 7, pp.269-285.
- Strawson, P. F. (1959), Individuals. London: Methuen.
- Strawson, P. F. (1966), The Bounds of Sense: An Essay on Kant's Critique of Pure Reason. London: Methuen.
- Wiggins, David (1967), Identity and Spatiotemporal Continuity (Oxford: Oxford University Press)

¹ B 232-256.

² See, for example, B 131-2. This argument is discussed further below.

³ B 133.

⁴ A 107.

⁵ A 110.

⁶ B 131-2.

⁷ See Strawson 1966, pp.72-117. Also relevant is his discussion of persons in Chapter Three of his 1959.

⁸ Compare Hurley 1994, pp. 141-3.

⁹ Certainly, there is something very implausible about Kitcher's claim (1990, pp.92-4 and 125-7) that the principle of the unity of apperception has nothing at all to do with the possibility of self-ascribing experiences.

¹⁰ B 135.

¹¹ A 113.

¹² A 116.

¹³ A 123.

¹⁴ See Henrich 1976. An English presentation of his position will be found in his 1989.

¹⁵ See Guyer 1987, particularly pp.132-149. He discusses Henrich's position in his extended review of Henrich's book (1979).

¹⁶ Henrich 1989, p.255.

¹⁷ A 116.

¹⁸ Henrich 1989, p.272.

¹⁹ Guyer 1987, pp.137-8. See also his 1992.

²⁰ Guyer 1987, p.146.

²¹ Compare Guyer 1992, p.144.

²² Guyer 1987 p.147

²³ Cassam 1989, p.81.

²⁴ A 364.

²⁵ The importance of this argument for understanding the unity of apperception is stressed by Sellars (1979, p.70).

²⁶ A 363.

²⁷ A 363a-364.

²⁸ See Shoemaker 1970 and Parfit 1971.

²⁹ See Parfit 1984 pp.254-261 for a more detailed argument for this conclusion

³⁰ It is worth pointing out in this context that Kant's conception of numerical identity, unlike Leibniz's, does allow for two objects that are qualitatively indistinguishable to be numerically distinct. Difference of spatial position is sufficient for numerical distinctness (A 263-264, B 319-320).

³¹ A 348.

³² I have discussed this at greater length in Bermúdez forthcoming.

³³ Compare the discussion of the category of substance in the First Analogy (A 182-189, B 224-232) and in the Refutation of Idealism added in the second edition (B 274-287). There is an excellent discussion of Kant's argument in the Refutation of Idealism in Cassam 1993.

³⁴ See A 107 where Kant famously remarks that "No fixed and abiding self can present itself in this flux of inner appearances." The parallel with Hume is clear enough, and quite possibly a case can be made for direct influence. Kitcher points out that Kant was familiar with Hume's discussion of personal identity via the passages quoted in the German translation of James Beattie's Essay on the Nature and Immutability of Truth - see her 1990, pp.97-102.

³⁵ This diagnosis of the Paralogisms is repeated in the second edition. See B 408-9. Cassam 1993 explores the possibility that one's body might function as such a subject of introspective awareness.

³⁶ A 116.

³⁷ B 134.

³⁸ This is intended to hold also when Kant refers, as he often does, to a priori certainty. A representative passage is A 113 where he states: "But as self-consciousness is a transcendental representation, numerical identity is inseparable from it, and is a priori certain". A priori certainty does not imply truth any more than a priori consciousness does.

³⁹ See, for example, A 508-509, B 536-537. The distinction between regulative and constitutive principles is actually introduced in the discussion of the Analogies of Experience (A 179-180, B 222-223), but in a slightly different sense. Broad 1978, pp.228-9 has an interesting suggestion about the two senses might be related.

⁴⁰ A 508-9, B 536-7.

⁴¹ This is not to claim, however, that ascribing experiences to a single subject is what makes them a single series. Quite the contrary, as suggested above, the self-ascription of experiences presupposes and does not explain the unity of consciousness. I return to this at the end of the paper, where I

suggest that it is in the *Analytic of Principles* that we find Kant's account of the conditions of the unity of consciousness.

⁴² B423a.

⁴³ Kitcher's proposal is not the only way in which one might retain the idea that 'I' is a referring expression in the No-Ownership situation. It could be argued, for example, that it refers to the body, or perhaps to the brain. But such suggestions would not be very Kantian in spirit.

⁴⁴ The No Ownership Reading, as a possible interpretation of the unity of apperception, should be distinguished from the No Ownership situation discussed in the previous section. Of course, if the No Ownership Reading is true, then the No Ownership situation will hold (i.e. there will be no independently existing subject), but the converse might not hold. It is possible that the No Ownership situation might hold (i.e. there would be no independently existing subject), but that the No Ownership Reading might still be wrong, because it could be necessary for the unity of apperception to think in terms of an independently existing subject (even though there is no such thing).

⁴⁵ The distinction is very clearly made in Harrison 1982.

⁴⁶ B 131-132.

⁴⁷ A 120.

⁴⁸ A 117a.

⁴⁹ Of course, this does not mean that there could not be representations which are in principle unconscious. Kant recognises that such representations are possible in his famous letter to Herz in May 1789. But, as he stresses in the letter, such representations would have no part to play in cognition. Guyer's discussion of this argument (1987, pp. 140-147) is vitiated by his conviction that Kant is trying to establish the absolute rather than hypothetical necessity of the unity of apperception.

⁵⁰ Certainly, the No Ownership Reading does not seem at all appropriate here.

⁵¹ To this extent I agree with Guyer that the Transcendental Deduction does not succeed in providing a self-contained argument.

⁵² I have been helped by comments from an anonymous referee for the EJP. Work on this paper was made possible by the award of a British Academy Post-Doctoral Research Fellowship.