

FIRST PERSON ILLUSIONS: ARE THEY DESCARTES', OR KANT'S?

Christopher Peacocke
Columbia University and University College, London

I am going to bring to bear a contemporary conception of the first person and of the self upon Kant's discussion of Descartes and other 'rational psychologists' in the *Critique of Pure Reason*. My aim in doing so is to understand better the nexus of relations between the first person concept, the self, and their epistemology and metaphysics, in the light of the contributions made by these two great philosophers. Kant accuses rational psychologists, amongst whom he cites Descartes as a representative, of a mistaken conception of this nexus of relations. I will be arguing that plausible theses in the philosophy of mind and the theory of intentional content imply that not all the problems lie on Descartes' side of the argument.

1. The Issues

Kant's treatment of rational psychology in the section entitled 'The Paralogisms of Pure Reason' in the *Critique of Pure Reason* is philosophical writing of such brilliance and depth, admired by successive generations of philosophers, that several disclaimers about my aims are required right at the outset. I have no dispute with Kant's critique of the idea that the subject of consciousness must be a simple, non-complex entity. Kant's writing on that topic looks extraordinarily prescient, even in an age of cognitive science. More than two centuries later it could hardly be bettered in so short a compass. Nor do I think that any of the other characteristic metaphysical doctrines of Descartes and the rational psychologists are true. Those doctrines hold that the subject of consciousness is immaterial and non-spatial, and not only persists over time, but also never ceases to exist. My dispute with Kant is not about many of his conclusions. The disagreement concerns rather some crucial points in his reasoning, and his corresponding diagnoses of Cartesian thought.

Kant's section on the Paralogisms has also been the focus of some important critical commentary in recent years, and I will at some points draw on some of

this work in comparing his position with one that seems to me more defensible.¹ Kant's treatment of the Paralogisms involves the intertwining of many strands in his thought, and a full textual discussion would merit a book-length examination.

The contemporary position on which I want to draw in developing my argument can be specified by four theses: two in the philosophy of intentional content, one in the philosophy of mind, and one in the metaphysics of subjects. The first thesis in the philosophy of intentional content concerns the individuation of the first person concept, and it is itself twofold. Its first part states that

What makes something the reference of the concept *I* in any particular event of thinking is that it is the agent, the producer, of that thinking.

Its second part states that the displayed condition is not merely a constitutive condition for something's being the reference of *I* in a particular thinking. The second part holds that the displayed condition, the fundamental reference rule for the first person concept, is also what makes something, what individuates, the first person concept (see further Campbell (1994) and Peacocke (2008, 2012)). This second part of the thesis can be seen as an instance of a broadly Fregean claim. It is an application of the idea that what individuates a sense is the fundamental condition for something to be the reference of the sense (Dummett 1981, Peacocke 2008).

The other thesis in the theory of intentional content applies to first person, or *de se*, contents more generally. These contents may occur in perception, in memory, in action awareness, or even in sensation (you may have a pain experienced as occurring in *your* arm). This second thesis states that

For a mental state or event to have a *de se* content is simply for it to have a content that, *de jure*, concerns the subject who enjoys that state or event.

This applies both to conceptual and to nonconceptual contents. Again, it is plausible that this is what makes *de se* content the content it is, both in the conceptual and in the nonconceptual instances. In the special case in which the mental event is a mental action of making a judgement with a first person content, there is no conflict between this second thesis and the first thesis. In the case of a mental action that is the judging of a first person content, the agent making the judgement is identical with the conscious subject who enjoys the action awareness of making the judgement.

The contemporary thesis in the philosophy of mind states that

A subject who possesses the first person concept has a mental file on himself, a file for the first person concept. When all is functioning properly, the present tense contents of this file are updated noninferentially, as time passes, to the appropriately corresponding past tense contents. So, for instance, if on one day the subject's file contains something with the content *Today I am in Oxford*, then

the next day the file will contain something with the content *Yesterday I was in Oxford*, without any personal-level inference on the subject's part.

The idea that mental files are important for understanding first person thought is found in several writers of rather different stripes (Perry 2002, Peacocke 2012). This thesis about mental files can also be applied to nonconceptual *de se* contents, in accounting for the way in which a present tense perception with a *de se* content can generate a later memory with a correspondingly appropriate past tense *de se* content (Peacocke 2012).

Lastly, the contemporary thesis in the metaphysics of subjects states that

The identity of a conscious subject over time is constitutively dependent on the identity of an integrating apparatus that integrates the precursors of perceptions, sensations, thoughts and emotions.

This is not a conception on which identity of subject is dependent on a subject's having a body. The conception does, however, involve the identity of a material entity, the integrating apparatus (Peacocke 2012).

So much by way of very brief statement of a contemporary position. I now turn to the application to Kant's critique of Descartes and the other 'rational psychologists'. When Descartes presents his arguments about what he is, and his essential properties, it is overwhelmingly plausible that he is using the first person in its ordinary sense. Descartes' arguments simply employ the first person pronoun, or first person case, in natural language (Latin or French), without any stipulation of a special sense. The sense expressed by these uses of the first person is the way of thinking of himself employed in Descartes' ordinary first person thoughts. We are meant to assess his arguments as sound or otherwise taking them to involve this normal sense. When you adapt Descartes's arguments to your own case, as when you think through the *Cogito* and refer to yourself, you will be using the ordinary first person way of thinking that you employ in your other first person thoughts.

Kant himself relies on a series of highly substantive theses in the philosophy of mind, the philosophy of thought, and the theory of meaning and intentional content in his critique of a Cartesian-style position. Here is a selection of these theses:

The rational psychologist is making a mistake that involves a failure to appreciate the importance of the point that "not the least intuition is bound up with" first person representations (A350). A genuine and non-empty singular concept of an object that can be given (or at least an object that is a substance) must be based on an intuition of that object (B412–3).

There is a positive account to be given of the role of the first person in the rational psychologist's premise "I think", an account that involves the idea that the **I** in 'I think' is "rather purely intellectual, because it belongs to thinking in general"

(B423). Kant says that if he has called “I think” an empirical proposition, that is because “without any empirical representation, which provides the material for thinking, the act I think would not take place, and the empirical is only the condition of the application, or use, of the pure intellectual faculty” (B423).

“rational psychology has its origin in a mere misunderstanding. The unity of consciousness, which grounds the categories, is here taken for an intuition of the subject as object” (B421).

The rational psychologist’s reasoning as captured in the First Paralogism “passes off the constant logical subject of thinking as the cognition of a real subject of inherence” (A350). “Meanwhile, one can quite well allow the proposition **The soul is substance** to be valid, if only one admits [. . .] it signifies a substance only in the idea but not in reality” (A350–1).

The apparent awareness of we have of our own identity over time should be given a construal that does not involve awareness of the genuine identity of anything over time (A362–4). “The identity of the consciousness of Myself in different times is therefore only a formal condition of my thoughts and their connection, but it does not prove at all the numerical identity of my subject” (A363).

What exactly these claims mean, and what they are targeted against, we will consider below. My general position is that while some of these points involve incisive and original insights, which is why Kant’s discussion has been so admired, what is right in these points cannot be soundly applied against significant parts of Descartes’ reasoning. Others of these theses I will be arguing to be untrue. Any case against Descartes’ metaphysics of the mind needs to be differently organized, and in part differently based. I divide the relevant issues into three clusters.

(i) Kant’s Complaints About the Cartesian Use of The First Person Concept:

The first cluster of questions concerns some plausible claims that Kant makes about the first person concept. Does what is right in Kant’s points about the first person concept as it occurs in certain premises used by the rational psychologist really undermine the early stages of Descartes’ argument? Kant seems to suggest that an appreciation of his points about the relevant occurrences of the first person concept, when taken together with other theses held by the rational psychologist, will show that the rational psychologist’s position involves the violation of constraints on significance, or at least the violation of constraints required for a singular concept to refer to a ‘substance’. What, if anything, is the connection between what is right in Kant’s points about the first person, in its relevant uses, and the alleged violation of a principle of significance? Is Kant’s target the transitions made in Descartes’ thought from certain states of consciousness, or is it something else?

(ii) Kant's Positive Account:

The second cluster of questions centres on Kant's positive account or conception of how the 'I' in 'I think' operates. What is Kant's own positive account, and with which parts of Descartes' position is that positive account incompatible?

(iii) 'A Merely Logical Subject'

The third cluster of questions concerns Kant's frequently expressed idea that there is in first person ascriptions of conscious states a 'merely logical subject'.² Kant's oft-expressed idea is that the rational psychologist is mistaking this 'merely logical subject' for something more substantial. What is a 'merely logical subject'? How does this notion relate to the concerns in the question clusters (i) and (ii)? Is it really true that there are uses of the first person that are correctly described as involving 'merely logical subjects', and produce illusions of reference to something more substantial? If we question the idea of a 'merely logical subject', is there any good reason for a subject not to take her apparent awareness of her own identity over time at face value?

2. Some Replies to Kant's Objections

I will be arguing that a contemporary account of the first person and the self that endorses the theses displayed above suggests answers to some of the questions raised in these clusters (i)–(iii), answers that diverge from Kant's own position.

(i) Kant's Complaints About the Cartesian Transitions Involving the First Person Concept:

In developing his case against rational psychology, Kant makes the point that there is a use of the first person which does not involve its reference being given in an intuition: "For the I is, to be sure, in all thoughts; but not the least intuition is bound up with this representation, which would distinguish it from other objects of intuition. Therefore one can, to be sure, perceive that this representation continually recurs with every thought, but not that it is a standing and abiding intuition, in which thoughts (as variable) would change" (A350). That the first person 'representation' about which Kant is saying this is something he takes to be genuine, and not something illusory postulated by the rational psychologist, is evident in several other passages. In the following passage, from the B edition, he emphasizes that this representation is used in

describing what the thinker can really become conscious of, and again it does not involve any intuition of the subject: “this identity of the subject, of which I can become conscious in every representation, does not concern the intuition of it, through which it is given as object” (B408). Later in the second edition he writes of “the empirical but in regard to all kinds of intuition indeterminate proposition ‘I think’” (B421).

Kant emphasizes the absence of any intuition associated with this use of the first person, because the rational psychologist is concerned to establish the existence of purely mental, non-material substances. For Kant, any genuine thought about substance must involve an intuition of the substance. He writes of “the category of substance, which always presupposes a given intuition” (B422). He repeatedly asserts this link between substance and intuition. “The concept of substance is always related to intuitions, which in me cannot be other than sensible” (B408). We must “ground the persistence of a given object on experience if we would apply to that object the empirically usable concept of a substance” (A349). “Thus if that concept, by means of the term “substance”, is to indicate an object that can be given, and if it is to become a cognition, then it must be grounded on a persisting intuition as the indispensable condition of the objective reality of a concept, namely, that through which alone an object is given” (B412–3). So the argument seems to be: the use of the first person in the rational psychologist’s premise “I think” does not have the required connection with an intuition that it would have to have if the rational psychologist’s reasoning is to establish what he wants it to establish. Kant summarizes his discussion of the paralogisms in the second edition discussion: “From all this one sees that rational psychology has its origin in a mere misunderstanding. The unity of consciousness, which grounds the categories, is here taken for an intuition of the subject as object, and the category of substance is applied to it” (B421–2).

Kant’s starting point in this reasoning, that there is no intuition ‘bound up with’ the first person seems to me entirely correct. It is supported by everything in a good account of the first person concept. No demonstrative, either of the form ‘that’ or ‘that F’, in perception or sensation or any other form of consciousness, is equivalent to a representation of something as oneself. Even when a perceptual experience has a genuine first person content—as in the example of seeing something as coming towards oneself—there does not need to be an intuition of oneself in the experience which gives it that first person content. Kant’s premise seems to be true and important. It is in part a Humean insight. The point is completely independent of Kant’s phenomenal/noumenal distinction, and of other contentious theses in his theoretical thought. That is, dialectically, part of the strength of this premise.

The explanation of the truth of this point about the absence of any intuition “bound up with” the first person lies in the difference between the fundamental reference rule for any demonstrative and the fundamental reference rule for the first person (in both its conceptual and its nonconceptual forms). Any intuition-based demonstrative picks out the object experienced, either perceived or sensed.

That is, it refers to the object standing in a certain relation to what Kant would call the intuition. The first person by contrast always refers *de jure* to the subject of the experience or conscious event in question. These are utterly different conditions. The condition for the first person to refer does not even require an intuition of what is in fact the subject, for the first person may feature in the content of an experience in which the subject is not given in any intuition. The first person may also feature in an intuition-free thinking. This difference between any intuition-based demonstrative and the nature of the first person is brought out clearly by the first two theses above of the contemporary conception.

I pause the argument briefly to note that this account of what is arguably defensible in Kant's thought makes the point one that is quite specific to the first person concept, rather than something applicable to any singular concept that can feature in subject position in the content of a thought. In his highly illuminating article 'Kant's First Paralogism', Ian Proops (2010) gives a construal of Kant's treatment of that paralogism. On his construal, the rational psychologist rightly points out that the first person concept cannot feature in predicative position. Proops' point is that nothing follows from this about whether the subject referred to by the first person concept is something whose existence depends on the existence of other things. As Proops notes, a singular concept picking out something whose existence is obviously dependent upon the existence of other things, such as your right fist (Proops' example), is equally incapable of featuring in predicative position (2010: 475–7). I think that Proops is certainly identifying one strand in Kant's thought here. But this particular strand is not something unique to the first person, as the example of the fist shows. We need to cite something distinctive of the first person, amongst singular notions of persisting entities, if we are to contribute to an explanation of why there should be a special tendency to illusions about the self. The non-equivalence of *I* with any demonstrative concept of the form *that F*, where this demonstrative is some perceptual or sensorily-based singular concept, is certainly a feature of much more restricted application. The feature integrates well with some of Kant's adumbrations and analysis. In particular, if the *F* that occurs in a perceptual demonstrative is something true of material extended things, then if *I* were equivalent to such a demonstrative *that F*, then the subject would even be given as something material and extended; and in fact the subject is not so given.³ In the first edition's treatment of the First Paralogism, the absence of any "intuition bound up with" the first person is certainly playing a role in the argument (A350). By contrast, there would, of course, be an intuition bound up with the perceptual demonstrative "this right fist".

To return to the main argument: I have argued that Kant is recognizing an important point about the distinctive nature of the first person concept *I*, viz. that its fundamental reference rule does not mention or require any relation of the reference of the concept, on a particular occasion, to any experience enjoyed by the thinker. This elaboration of the point on which I want to agree with Kant about the first person does, however, undermine part of the use Kant wants to

make of the very same point. The explanation just given of what seems right in his point is entirely at the level of modes of presentation, at the level of intentional content rather than the level of reference. The explanation is quite neutral on the nature of the ontology of subjects. It does not advert to the metaphysics of subjects at all, but only to distinctions drawn between ways in which something can be represented.

It is, then, entirely consistent with acceptance of Kant's point about the first person to insist that in a thinking of 'I think', the first person concept does refer. It just does not refer by means of a relation to some particular intuition. It would even be consistent with such a position to hold that, if the first person is to refer, it must be to something that could be an object of intuition (though Kant would not agree, in this particular case). Such a constraint is not at odds with the explanation at the level of modes of presentation.

There are other genuinely indexical concepts that refer on particular occasions of their use, but do not do so via a relation to a perceptual intuition. If we prescind from Kant's own very special and complex views about time, the present tense concept *now* is plausibly a salient example. The present tense concept can feature in thoughts ("I wonder what is happening in Congress now") without there existing any perceptual intuition of which it is true that it is an intuition of the time referred to by *now* in the thought in question, and is an intuition whereby *now* as used on this occasion gains its reference. Perceptual experience represents events as occurring in the present, and one's thinkings are experienced as occurring in the present, but none of this means that the present is itself given in a perceptual intuition. Also, the way the present is given in experience is not something that makes the present time something that is available as an object of attention. A detailed discussion of time in experience is a topic for some other occasion, but still it should be clear a broad parallelism is present. The fundamental reference rule for *now* is that in any mental event in whose content it occurs, *now* in that event refers to the time of occurrence of that event. Since any event occurs time, its present tense component will refer to a particular time, even if there is no such thing as a perceptual intuition of a time. If it is equally in the nature of mental events that each such event must be enjoyed by a subject, then equally in any mental event with a first person content *I*, that use of *I* will refer, under the fundamental reference rule we have been discussing, and it will refer to the subject of that event.

We can now ask a pivotal question about Kant's intended target in the argument of the Paralogisms. When Kant makes what we have endorsed as the sound point that "no intuition is bound up with" the first person as it occurs in the rational psychologist's reasoning, what is his aim? Is his argument aimed

- (a) against the transitions that lead the rational psychologist to his conclusion *I think*, as inadequate to ground knowledge about the thing that is the reference of *I*? This we can call "the inadequate-grounds interpretation". Or is his argument aimed

- (b) to establish that the rational psychologist is presupposing an incorrect theory of the nature of the first person concept? This is the “misunderstood-concept interpretation”. Or is his argument
- (c) that if the rational psychologist’s general metaphysics of the subject is to be correct, the first person concept together with its reference would have to have various properties that the rational psychologist’s reasoning has in no way established that they have? This is the “missing properties” argument.

Kant might of course have had more than one of these aims simultaneously. But I will argue that the first two of these interpretations take as their target something to which a rational psychologist such as Descartes has a good defence, while the third interpretation, when considered in more detail in the light of Kant’s text, has problems as an interpretation of Kant’s intentions.

The inadequate-grounds interpretation, (a), has Kant as insisting that the transitions that, for example, Descartes makes from thinkings or experiencings or imaginings or willings to *I think* are fallacious. But there is a strong case to be made that such transitions, taken simply at face value, are not fallacious. What makes it the case that some event or state is a conscious event is that there is something it is like for the subject of that event (Nagel 1974, Peacocke 2012). If that is correct, then a transition in thought from enjoyment of a conscious state to self-ascription of a state, of the conscious kind it is, will always yield a true judgement. If the conscious kind is thought of in a way that has constitutive links to self-ascription when the subject enjoys the conscious state or event, then there is a good case to be made that the transition not only yields a true judgement, but one to which the thinker is entitled, and which can be knowledge (Peacocke 2012b). There is not required to be any intuition of the subject in the conscious state from which the transition to *I think* is made. Even if *per accidens* there is an intuition of the subject in the initial conscious state, that is not what makes the transition truth-yielding, entitled, and knowledge-yielding.

These points do not rest on or presuppose any questionable Cartesian metaphysics of subjects of consciousness. They stand independently of any such metaphysics. Nor do they presuppose some bleached-out or *ersatz* notion of reference for the concept *I* as it features in the conclusion *I think*. Whatever is wrong with a Cartesian metaphysics of subjects, it is not these transitions.

The misunderstood-concept interpretation, (b), has Kant protesting that the rational psychologist is presupposing an incorrect theory of the nature of the first person concept. I suspect this interpretation does an injustice to both Descartes and Kant. Of course neither Descartes nor Kant, to the best of my knowledge, formulated explicit semantic theses about the first person concept. A fortiori, neither connected semantic theses with the epistemology and metaphysics of subject. (Much as one may wish to the contrary: the question of what Kant might have said had he had Frege’s sense/reference apparatus to hand is intriguing here as in other areas too.) The reason this interpretation (b) seems to be an

injustice is that some of the most striking points that each of these philosophers makes about the first person is vindicated under the supposition of correctness of the first thesis in the philosophy of mind we mentioned at the outset: that what makes something the reference of the concept *I* in any particular event of thinking is that it is the agent, the producer, of the thinking. This reference rule contributes essentially to the validation, both semantic and epistemic, of the *Cogito* transitions in Descartes (again, see Peacocke 2012b). If Descartes had thought some intuition had to be associated with the first person for it to gain reference, he would never have accorded to the *Cogito* the status he actually gave it. Nor could the *Cogito* have turned back the doubt for him. Equally, there are comments in Kant that would seem to be best explained by his having some tacit appreciation of the above reference rule for *I*, or at least the independence of any correct rule from intuitions of the thing that is its reference. Kant acutely observes that “even if we supposed” “secure observation” of a soul (whatever that might involve), that would not establish the persistence of what is referred to by the first person, precisely because “not the least observation is bound up with this representation, which would distinguish it from other objects of intuition” (all quotes from A350). More generally, Kant comments: “I would not say by this that the **I** in this proposition [viz, **I think** – CP] is an empirical representation; for it is rather purely intellectual, for it belongs to thinking in general” (B426). If Kant had held that an intuition of an object were required for *I* to refer to the object, he would after all be holding that the **I** in **I think** is an empirical representation, which is precisely what he is denying.

Kant does repeatedly emphasize in his discussion of the Paralogisms that one cannot obtain from a concept alone “the usual conclusions of the rationalistic doctrine of the soul” (A350–1); “here as elsewhere we can have little hope of broadening our insight through mere concepts” (A361). His point here concerns what conclusions can be drawn from various phenomena involving the first person concept, not that Descartes or other rational psychologists have mistaken the nature or fundamental identity of that concept.

The missing-properties interpretation, (c), has Kant saying that the rational psychologist has failed to establish that the first person concept and its reference has certain properties they would have to have if the rational psychologist’s conclusions are to be true. Certainly anyone who agrees that subjects do not exist for ever, and are not immaterial, is committed to saying that Descartes’ arguments must be unsound somewhere. His arguments must either have false premises, or rely on invalid transitions. Now in the *Meditations*, the stage of Descartes’ reasoning at which he concludes that he is purely a thinking thing, and not an extended thing, is to be found in Meditation VI. His reasoning there is complex and articulated. It involves the notion of a clear and distinct idea, it involves a conception of natures of things, the notion of a ‘complete’ thing, and his own supposed capacity to conceive of thought without extension. (For a particularly good discussion of the structure of Descartes’ argument there, see Wilson 1978, Ch. VI, esp. p.197 ff.) So one might have expected Kant to

address Descartes' reasoning involving these various notions in his section on the Paralogisms. In fact he does not. Rather, what Kant does is to imply that Descartes had already gone astray in his conception of conscious mental events and states, and of their relation to what contents we can know that contain the first person concept in their content. The implication is that he is under some illusion about the role of the first person and its relation to consciousness, and about what conscious states can justify.

Here are some of Kant's statements. One of his formulations is that the rational psychologist has confused a 'formal condition' of the identity of his thoughts with the numerical identity of a subject (A363). "I relate each and every one of my successive determinations to the numerically identical Self in all time, i.e. in the form of the inner intuition of my self. On this basis the personality of the soul must be regarded not as inferred by rather as a completely identical proposition of self-consciousness in time, and that is also the cause of its being valid *a priori*. For it really says no more than that in the whole time in which I am conscious of myself, I am conscious of this time as belonging to the unity of my Self" (A362). Most striking, in discussing the concept of personality, conceived of as applying to subjects who are aware of their identity over time, he writes that: "we can never boast of it [the concept of personality – CP] as an extension of our self-knowledge through pure reason, which dazzles us with the uninterrupted continuous duration of the subject drawn from the mere concept of the identical self, since this concept revolves in a circle around itself and brings us no farther in regard to even one single question about synthetic cognition" (A366). At the start of the B edition discussion of the Paralogisms, he emphasizes, clearly intending his remarks in opposition to the rational psychologist, "I cognize myself not by being conscious of myself as thinking" (B406). It is, on Kant's view, the philosophical scrutiny of consciousness, its nature, and its relations that reveals the errors in the rational psychologist's thought. Presumably that is why he thinks there is no need to go into the theory of clear and distinct ideas, complete natures, and the rest, in the case of Descartes. "Thus through the analysis of the consciousness of myself in thinking in general not the least is won in regard to the cognition of myself as object" (B409). Kant's summary diagnosis of the errors of rational psychology mentions a mistake about consciousness, and not anything about the (highly problematic) argumentative apparatus of Meditation VI. In his summary, Kant writes, "From all this one sees that rational psychology has its origin in a mere misunderstanding. The unity of consciousness, which grounds the categories, is here taken for an intuition of the subject as object, and the category of substance is applied to it. But this unity is only the unity of **thinking**, through which no object is given; and thus the category of substance, which always presupposes a given **intuition**, cannot be applied to it, and hence this subject cannot be cognized at all" (B421–2).

If Kant's arguments are indeed confined to the domain of the consciousness/thought relations, where does this leave him if we accept the positions in the philosophy of mind, thought, and metaphysics that I mentioned at the

outset? As far as the first paralogue is concerned (I will return to the third), it looks as if an option has been overlooked. It is possible to maintain that the *Cogito* transitions are sound, that they yield knowledge, and that it is simply open to further investigation what other kind of properties are possessed by the entity to which the ordinary first person in the *I think* refers. It is consistent with the soundness and knowledge-yielding status of *Cogito* transitions that the first person refers to something, perhaps must refer to something, that could be an object of intuition. There is no obvious requirement that if *I* is to refer to something that could be an object of intuition, it must be an object of intuition in the very mental states or events that are the starting point of a *Cogito* transition. Under this overlooked option, there is no reason for saying that *I* is not being treated as fully referential in *I think* when that conclusion is reached by a *Cogito* transition. Attempts in the spirit of Lichtenberg to say that what is supported by the starting point of a *Cogito* transition can be explained in subject-free terms have proved difficult to sustain (see Williams 1978 Chapter 3, Peacocke 2012, 2012b). Though I have no brief for the full Cartesian metaphysics of the subject, it is not clear to me that Descartes has made any mistake within the sphere of consciousness/thought relations when he concludes that one can have knowledge of an entity in making a *Cogito*-like transition in thought.

This objection to an element of Kant's line of thought, like the other objections I will raise, does nothing to support a rational psychologist's conception of an ontology of purely mental subjects that never go out of existence. As I said, the objection in this case is to Kant's reasoning, not to his conclusion. Moreover, it is hard not to sympathize with Kant's view that a Cartesian conception of egos violates some form of a principle of significance. No possible experience has been associated, even indirectly, with the distinction between one Cartesian ego persisting and its replacement by another with the relevantly same mental states. That application of a principle of significance operates, however, at the level of reference and objects. It is an objection to the Cartesian conception of a certain kind of object, Cartesian egos. The objection does not fundamentally have to be formulated at the level of sense, at the level of the first person concept; though it will of course have ramifications for the truth-conditions of first person thoughts.

(ii) Kant's Positive Account:

The second cluster of questions asks whether Kant's own positive account of the 'I' in 'I think', or the underlying motivations for it, are incompatible with Descartes' position. The first task here is to say exactly what Kant's positive account is. On this, Béatrice Longuenesse (2008) has some important suggestions. Her view is that according to Kant, "In referring his thoughts to 'I', the thinker (perceiver, imaginer) is doing nothing more than committing himself to the unity and consistency of his thoughts, and committing himself to obtaining a unified standpoint that could be shared by all" (2008: 17). Following her description

of the role of 'I' in relation to these unifying functions, Longuenesse writes that "There is nothing more to be known of 'I' in the context of this argument" (2008: 17); and "It is thus apparent that the function of 'I' in this argument is quite different from what it was in Descartes' *cogito* argument" (2008: 17). "'I' in 'I think'" does not refer to a permanent *object* whose properties *change*. 'I' is just the term to which we refer our thoughts in order to think of them as unified by one standpoint and bound by rules that commit us ("me") to bring about unity and consistency under a unifying standpoint" (2008: 22) "... of course what we refer to by 'I' has to be one and the same through the whole time of our experience. And of course this identity is prior to and different from the identity of any object identifiable and reidentifiable *in* time, although it may readily be mistaken for such an identity" (2008: 23). Of such a statement as that the reference of "I" in "I think" is a mind distinct from a body, on this reading of Kant, "when we make this kind of statement we make a category mistake. For we compare the certainty of the pre-categorical existence contained in "I think" and the certainty of the *actuality* of objects given, identified, and reidentified in space and time" (2008: 25).

There is considerable textual support for some aspects of this reading, even beyond the passages Longuenesse cites. One such passage is a note Kant made about "I think" in his own copy of the first edition of the *Critique of Pure Reason*: "[This] is a proposition *a priori*, is a mere category of the subject, intellectual representation without anywhere or at any time, hence not empirical. Whether the category of reality lies in it, whether objective inferences are to be drawn from it" (1998: 413 note *a*). The reading gives further significance to such passages as A346/B404: "consciousness in itself is not even a representation distinguishing a particular object", and to the idea of what Kant calls *modi* of self-consciousness, which he describes in the second edition as "mere functions, which provide thought with no object at all" (B407). There is further support in these passages: "I would not say by this that the I [in "I think" – CP] in this proposition is an empirical representation; for it is rather purely intellectual, because it belongs to thinking in general"; "I think of myself, in behalf of a possible experience..." (B426).

The first set of points I want to make about this interpretation is that, under this reading of Kant, the "I" in "I think" must still have a very close connection with the ordinary first person concept, for three reasons.

First, to engage with such a rationalist as Descartes, these Kantian points will be relevant to Descartes' reasoning only if they employ the same first person concept as Descartes was employing in reasoning. But it was the ordinary first person that Descartes was employing. This is consistent with Longuenesse's main points if we regard Kant as making a contribution to a constitutive account of what is involved in grasp of the first person. That account can be developed consistently with the first person being governed by the fundamental reference rule I have been defending. The Kantian account aims to say something philosophically informative about grasp of the same first person concept mentioned in the fundamental reference rule.⁴

Second, it is very plausible that the constraints of consistency and unity that Longuenesse very properly emphasizes must involve consistency and unity of predications about objects in the world. These predications will include both contents that are not about oneself, and also some that are about oneself (including one's spatial, temporal, and causal properties and relations). Having consistent and unified predications about the wrong thing would not be enough. How is the reference that is the target of these consistency and unity requirements determined when the contents involve the first person? The plausible answer is that the reference must be the thinker who is thinking the "I" in "I think". But that rule is precisely the rule that individuates the normal first person concept, on the contemporary account outlined at the start of this paper.

Third, Longuenesse emphasizes that " 'I think' is a *universal* form of thought, which can be attributed to *any* thinker; on the other hand, this universal form is necessary for *particular, empirically determined perceivers and thinkers* to come up with thoughts about the world that *are independent of their own particular standpoint in the world*" (2008: 17). I agree. This interplay between the general requirements that apply to an arbitrary thinker in the world, and those that apply to each one of us thinkers individually, is entirely consistent with taking the constraints with which Kant is concerned in the "I" in "I think" as applying to its normal sense when used by any one of us. The first person concept as employed by any one thinker is an instance of a general type instances of which are employed by other thinkers. This is the type labeled the '[self]' type in the kind of neo-Fregean terminology account I would favour myself (1981). The constraints of unity and consistency can be formulated for an arbitrary thinker, for anyone who is employing an instance of the [self] type; the constraints are then given for arbitrary instances of that type. That is consistent with use of any instance of the type, so constrained, being the ordinary first person concept as used by a particular thinker.

It would, however, be a mistake to think that Kant's points about 'I' apply only at the level of the type of concept. On the contrary, it is clear that in many passages, Kant is talking about his own uses of 'I'; and it is clear that anyone else is meant to appreciate that the same points apply to his own uses (involving that instance of the [self] type available to him). This point applies to many of the passages quoted or cited above. As we saw, Longuenesse wrote " 'I' in "I think" does not refer to a permanent *object* whose properties *change*" (2008: 22). But if Kant's points apply at the level of instances of the first person type, those are instances that do refer—not indeed to a permanent object, but certainly to an object that changes. What is true is that in thinking of something in the first person way, one is not thereby thinking of it as an object identified in a particular way in relation to space, nor as something temporally reidentified in a certain way, nor as an object given in an intuition. These are all points at the level of sense, not of reference. As far as I can see, accepting these points and acknowledging their philosophical importance is consistent with the soundness and knowledge-yielding character of *Cogito*-like transitions, when taken at face value. The points

elaborate constraints on what is involved in grasping the first person that features in the *Cogito*-like transitions. The constraints do not undermine those transitions, taken at face value. In particular, they do not undermine that part of Descartes' discussion that is confined to the nexus of consciousness/judgment relations involving the first person, as opposed to the further metaphysics of Meditation VI, and Descartes' views about the nature of the reference of the first person.

(iii) 'A Merely Logical Subject'

The third cluster of questions concerns Kant's idea that the rational psychologist mistakes what is a 'merely logical subject' for something more substantial, a mistake that involves confusing the unity of experience with an experience of unity. The first paralogism, he writes, "passes off the constant logical subject of thinking as the cognition of a real subject of inherence" (A350). Kant's view seems to be that all conscious states have a certain common form; this form is captured by the characterization "I think"; "consciousness in itself is not even a representation distinguishing a particular object" (A346/B404), but the "I" in "I think" is still a logical subject, and the fact that the "I think" can accompany all my representations may, according to Kant, lead to the mistaken thesis that from all this, with no other premises, one can conclude that some entity stands in a relation to all one's own conscious states.

I agree that there is a phenomenology of all of what Kant would call one's 'representations' being one's own. This is a feature of consciousness, of its form if you will. Nonetheless, Kant's argument as summarized in the preceding paragraph is open to objection if the points I made on the second cluster of questions are correct. If "I" in "I think" does not have a reference at all, then of course mental states with an intentional content "I think . . ." cannot be a source of knowledge about the normal reference of "I" in "I think", since it has no such normal reference. But if it does have its normal reference, the possibility is reopened that such mental states can, on occasion, be a source of knowledge about the thing that is its reference. That it does have a normal reference in Descartes' thought is implied by what I have been arguing. If Descartes' *I* has its normal sense, it must also have its normal reference, however mistaken Descartes may have been about the properties of that normal reference.

If a thinker, including a rational psychologist, is clear-sighted about how the reference rule for *I* works, then is it quite unfair to accuse him of mistaking the unity of experience for an experience of unity. The clear-sighted thinker will not hold that an experience of unity (as opposed to unity of experience) is in any way necessarily involved in a correct, referential use of *I*. Neither the rational psychologist, nor we, in judging "All these experiences are mine", need to be committed to thinking that, in making this attribution of common ownership of the experiences to a single subject, there needs to be an experience or intuition of that subject. According to the fundamental reference rule, a thinking "All these

experiences are mine” is true iff all the experiences in question belong to the subject doing the thinking. The experiences can all seem to be the thinker’s, and the thinker can be entitled to take this seeming as correct, without the thinker needing to have an intuition of himself. In its seeming to the thinker that all the experiences are his, he enjoys a seeming with a primitive first person content. The correctness condition for that content is as given in the thesis at the outset of this paper, that all the experiences do in fact belong to the subject enjoying that seeming. (Of course we also have to give an account of the entitlement to take the seeming at face value.) If Descartes did ever say that he has an experiential (Kantian-style) intuition of himself that is involved in such self-ascriptive uses of the first person, it is not something he ever needed to say. An easy acceptance of the stylish formulation “unity of experience, not experience of unity” should not lead us into thinking that this was a mistake Descartes had to make.

In his discussion of the third paralogism in the first edition, Kant makes a further point. He rightly and forcefully observes that it’s seeming to oneself that one was thus-and-so in the past does not suffice to make it the case that one was thus-and-so in the past (A362–5). His discussion is cast in terms of consciousness of the identity of oneself at different times.

Now it cannot follow from Kant’s true observation that there is a seems/is distinction in this area that there is no such thing as genuine awareness of one’s identity over time. To argue from that premise alone would be to apply the fallacious argument from illusion to the case of apparent awareness of identity over time.

The thesis in the contemporary philosophy of mind displayed at the outset of this paper serves further to illustrate why a fallacy would be involved. A temporally sensitive mechanism that noninferentially updates a subject’s file on itself, a file of the sort described in that thesis, can generate knowledge-sustaining awareness of identity over time. The mechanism will not be infallible, any more than perception needs to be infallible. But when the temporal updating mechanism is working properly, it can give a subject genuine awareness of its identity over time, an awareness that he did or was thus-and-so in the past, just as perception can give genuine awareness of physical objects and events. The operation of the updating mechanism also does not at all require that the subject who is aware that he was thus-and-so in the past have a current intuition of himself.

An adequate defence of the claim that such memories give genuine awareness of identity must involve an account of the kind of thing of whose identity over time the thinker is aware (even though of course a thinker may be unaware of his or her own kind in making ascriptions of individual identity over time). It is very plausible that Descartes, given the metaphysics of Meditation VI, has no good account of the matter. On the particular account I have been developing, the identity of the subject over time involves identity of integrating apparatus. The view that identity of subject in the most fundamental case involves identity of embodied subject also has an account, a different one, of the matter. Insofar as Kant’s point was merely that the rational psychologist has no good account

of the kind of entity the awareness of identity is awareness of, I am in agreement. But again, this is a point having to do with the level of reference, and the nature of the entities referred to, rather than one turning on the nature of the intentional content of mental states. It does not rule out genuine, knowledge-sustaining awareness of one's own identity with the person who was so-and-so in the past.

At this stage in the discussion, we reach a point at which I find myself in agreement with John McDowell in his book *Mind and World* (1994). Although a quick review of the literature will rapidly show that McDowell and I differ on the nature of subjects of consciousness, and differ on the correct account of the nature of the first person concept, we are nevertheless in agreement on a further point. McDowell writes that "we make room for supposing that the continuity of 'I think' involves a substantial persistence, without implying that the continuant in question is a Cartesian ego", once we discard a certain assumption (1994: 101). The assumption is "that when we provide for the content of this idea of persistence, we must confine ourselves within the flow of 'consciousness'" (1994: 101). If "confining ourselves to the flow of 'consciousness'" means confinement to what is common to both genuine and illusory experiences as of the identity of the subject over time, then I agree.⁵ In good cases, there is no more obstacle to saying that a subject is genuinely aware of his having been so-and-so in the past than there is to saying that perceptual experience is, in good cases, of an external material object. If I have understood him aright, McDowell and I agree that there can be genuine awareness of one's own identity over time.

This general approach to Kant's treatment of the third paralogism is not at all a complete dismissal of his thought. What it suggests rather is that Kant's insights should be put in a conditional form, and regarded as part of a *reductio* argument against the Cartesian conception. If the first person refers to something like a Cartesian ego, then grasp of the (alleged) distinction between one continuing ego and the replacement of one by another similar ego is something that would not show up in the thinker's psychology. In those circumstances, we could not properly make the distinction between mere apparent awareness of one's own identity over time and illusions of identity. The argument could be elaborated within the non-verificationist framework I developed in 'The Limits of Intelligibility' (1988). The truth of this conditional formulation of Kant's insight is, however, entirely consistent with the distinction between apparent and genuine awareness of identity being applicable when subjects are something other than Cartesian egos. Further, on this construal of what is valuable in Kant's discussion of the third paralogism, the argument does not turn on any kind of misinterpretation of the first person concept, nor on misunderstanding of the consciousness/judgement nexus for first person contents. It turns rather on the consequences of the Cartesian metaphysics of the self for the intelligibility of certain distinctions in memory and awareness.

We can contrast this treatment of the Paralogisms with that of Peter Strawson in *The Bounds of Sense* (1966: 163–9). Strawson's position and my own agree that the 'I' in the Kantian 'I think' does in fact refer, and refers to an entity

for which there are empirical conditions of identity. For Strawson, that entity is a person conceived of as fundamentally embodied: “our ordinary concept of *personal* identity does carry with it empirically applicable criteria for the numerical identity through time of a subject of experience (a man or human being) and that these criteria, though not the same as those for bodily identity, involve an essential reference to the human body” (1966: 164). “. . . the notions of singularity and identity of souls or consciousnesses are conceptually dependent upon, conceptually derivative from, the notions of singularity and identity of men or people. The rule for deriving the criteria we need from the criteria we have is very simple. It is: *one* person, *one* consciousness; *same* person, *same* consciousness” (1966: 168–9). On an account endorsing the contemporary theses displayed above, the entity is a subject whose identity over time turns on the identity of an integrating apparatus. This produces several divergences from Strawson’s position. In one direction, the integrating apparatus may or may not be embedded in a body which is the body of the subject in question (see further below). In the other direction, we can apparently, in certain conceivable split brain cases, make sense of there existing two subjects in the same body. In such cases there will, functionally, be two pieces of integrating apparatus. In those cases, Strawson’s rule “*same* person, *same* consciousness” is incorrect if there is meant to be only one person per body at a given time. I acknowledge that this position has a commitment to, and an obligation to justify, the proposition that the nature and identity of an integrating apparatus over time can be explained independently of embodiment.⁶

Strawson usually formulates the conceptual issues in terms of language rather than thought. For the most part, this makes no substantive difference. Strawson regards part of Kant’s crucial insight as the recognition that when someone “ascribes a current or directly remembered state of consciousness to himself, no use whatever of any criteria of personal identity is required to justify use of the pronoun “I” to refer” to the subject of that experience (1966: 165). That is, such ascriptions are, as we have come to say, immune to errors of misidentification. I agree that they are so immune, and that there is some more or less explicit recognition of the point in Kant. The immunity does, however, need some explanation. There are explanations of the immunity, explanations that go hand-in-hand with the contemporary conception outlined above (see Peacocke 2012a). A crucial part of such explanations is the account of the soundness of the transition, noninferentially, from a conscious event or state enjoyed by a subject to the subject’s self-ascription of an event or state of the relevant kind. The account emphasizes that a subject who moves from a mental state or event to the judgement that he is enjoying a state of event of the relevant kind will not, for reasons in the nature of the case, thereby be led to a false judgement on the contemporary conception outlined above. That explanation does not, apparently, require that subjects are fundamentally Strawsonian persons. Nor does it require that they be subjects whose experience is in part as of a route through an objective spatial world.

Strawson also says, and I agree, that in the criterionless self-ascription of mental states, “reference to the empirically identifiable subject” is “not in practice lost” (1966: 165). On my account, a public utterance of “I have an experience as of a lawn in front of me” will be produced by a subject, and this subject’s identity depends on the identity of an integrating apparatus whose persistence and identity over time is an empirical matter. The experience itself makes rational the self-ascription, when thought of under psychological concepts of experience. The experience could not exist without a subject, and its subject must be the subject for whom the ascription is thus made rational. For Strawson, such an utterance of “I” can still refer to a subject “because—perhaps—it issues publicly from the mouth of a man who is recognizable as the person he is by the application of empirical criteria of personal identity” (1966: 165). This indeed explains how the term refers, but it does not yet explain the immunity to error through misidentification.

Strawson could, of course, endorse the explanation of immunity to error through misidentification that I offered. It would then need much further argument that the subject for whom such self-ascriptions are thus made rational must also be fundamentally embodied, rather than meeting the weaker condition of having an identity dependent upon a continuing integrating apparatus. Strawson does give part of an argument for that view. He argues that subjects with a unified consciousness must have experiences that are unified, in being (in part) experiences having a certain relation to a path through an objective world. His argument is of great interest, given our present concerns, for it would question the adequacy of a contemporary conception based on the theses displayed above, and I turn now to consider Strawson’s argument.

3. Strawson’s Neo-Kantian Conception of Subjects

Nothing in the four contemporary theses displayed at the start of this paper implies that there are not also further true theses that equally constrain the properties and relations of conscious subjects. Nothing that has been argued so far rules out the possibility that further investigation of the nature of conscious subjects may not reveal that there are plausible principles implying that conscious subjects must in some central cases be embodied, and must have perceptions that involve a route through the spatio-temporal world. In *The Bounds of Sense*, Strawson argues precisely that there are such principles.

We can distinguish various strengths of a proposed requirement on conscious subjects concerning either embodiment or experienced spatio-temporal history. In the case of embodiment, the requirement might be

(B1) that the subject actually possess a body;

(B2) that the subject once had a body;

(B3) that it must be for the subject as if it has a body; or

(B4) that it must be for the subject as if it at least once had a body.

There is a similar range of possible proposals for a requirement of experienced spatio-temporal history. The requirement might be

(S-T1) that the subject has throughout its history a location the subject either perceives, or is capable of perceiving, as its own;

(S-T2) that the subject in an initial segment of its history had such a perceived or perceptible location;

(S-T3) that throughout the subject's history it is for the subject as if it has such a perceived or perceptible location; or

(S-T4) it is for the subject as if in some initial stage of its history it had such a perceived or perceptible location.

My own view is that not even the weakest of these proposed requirements is correct, either for embodiment, or for spatio-temporal location and history. There is in my view no conceptual or metaphysical obstacle to the possibility of a subject having only a sequence of monaural sound experiences, of a sort that do not represent objective events. This possibility is arguably that of Strawson's own sound world in *Individuals* (1959). Again, the subject may have a sequence of visual experiences with only what I called sensational properties in *Sense and Content* (Peacocke 1983), or olfactory experiences, or any combination of the preceding, non-representational experiences.

In *The Bounds of Sense*, Strawson sets out to expound a Kantian argument intended to prove the opposite. He considers "the thesis that for a series of diverse experiences to belong to a single consciousness it is necessary that they should be so connected as to constitute a temporally extended experience of a unified objective world" (1966: 97). This formulation seems to be of the form (S-T1) of the spatio-temporal view. Strawson writes "of a unified objective world", not "as of a unified objective world". The requirement is also stated for the whole sequence of experiences, not just for some initial segment thereof. Strawson, rightly in my view, takes this thesis to be in conflict with the view that there could be a subject of consciousness whose experiences consist solely of "a succession of items such that there was no distinction to be drawn between the order and arrangement of the objects (and of their particular features and characteristics) and the order and arrangement of the subject's experiences of awareness of them (1966: 98–99). "Such objects might be of the sort which the earlier sense-datum theorists spoke of—red, round patches, brown oblongs, flashes, whistles, tickling sensations, smells." (1966: 99).

Strawson's argument in support of his thesis can be divided into the following steps. I label them "Points", because they are set out here in not exactly the order

in which he expounds them, to bring out the direction in which the argument is proceeding:

Point (1): "Unity of the consciousness to which a series of experiences belong implies, then, the *possibility* of self-ascription of experiences on the part of a subject of those experiences" (1966: 98).

Point (2): Nothing but "a form of words" is added to the hypothesis of a "succession of essentially disconnected impressions" by saying that they all belong to a single subject, or that a unitary consciousness is aware of them all (1966: 100).

The remainder of Strawson's argument from here on argues for a necessary condition for a series of experiences to belong to a single subject.

Point (3): There is a necessary condition for a series of experiences to belong to a single subject, a condition without which "even the basis of the idea of the referring of such experiences to an identical subject of a series of them by such a subject would be altogether lacking" (1966: 101).

The argument for the necessary condition proceeds from a claim about the nature of experience:

Point (4): "There can be no experience at all which does not involve the recognition of particular items *as* being of such and such general kind. It seems that it must be possible, even in the most fleeting and purely subjective of impressions, to distinguish a component of recognition, or judgement, which is not simply identical with, or wholly absorbed by, the particular item which is recognized, which forms the topic of the judgement." (1966: 100)

Strawson immediately goes on to say there is an apparent difficulty here:

Point (5): "Yet at the same time we seem forced to concede that there are particular subjective experiences (e.g. a momentary tickling sensation) of which the objects (accusatives) have no existence independently of the awareness of them." (1966: 101)

Strawson says the tension he sees between Points (4) and (5) can (actually, he writes "must") be regarded by Kant, or a Kantian, as resolved by this observation:

Point (6): "the recognitional component, necessary to experience, can be present in experience only because of the *possibility* of referring different experiences to one identical subject of them all. Recognition implies the *potential* acknowledgment of the experience into which recognition necessarily enters as

one's own, as sharing with others this relation to the identical self." (1966: 101)

Point (7): The potentiality of such an acknowledgment implies that "some at least of the concepts under which particular experienced items are recognized as falling should be such that the experiences themselves contain the basis" (1966: 101) for a seems/is distinction; and "collectively, the distinction *between* the subjective order and arrangement of a series of such experiences on the one hand and the objective order and arrangement of the items of which they are experiences on the other" (1966: 101).

I am not at all sure, well over forty years after first reading it, that I understand the conception of the subject matter that generates this bold and fascinating argument. But I do think the argument as formulated here is vulnerable to the following objections.

(a) There is a fundamental distinction between giving a theory of the relation

experience e is owned by subject s

and giving a theory of the different relation

the subject s is capable of self-ascribing experience e in thought.

Having an experience e is one thing; having the mental capacities to be able to self-ascribe e is another. *Prima facie*, many animals and young children have the former and lack the latter. Some passages from Strawson show him using now a thesis about ownership of experiences, now a thesis about using some form of first person representation and self-ascribing experiences. At the start of the argument, he is apparently writing about first relation, that of ownership: "First, we ask: how can we attach a sense to the notion of the single consciousness to which the successive "experiences" are supposed to belong?" (1966: 100). Yet the argument later mentions what's "in fact self-ascribed": "Not all the members of such a series [of "experiences which belong to a unitary consciousness" – CP] are in fact self-ascribed: a man may be more prone to forget himself in contemplation of the world . . . than he is to be conscious of, or to think of, himself as perceiving" (1966: 104). If it is the relation of ownership itself that is in question, no 'in fact' qualification necessary. For self-ascription, the qualification surely is required. The starting point of the argument concerns the unity of the consciousness to which a series of successive experiences belongs. Any move from this to consideration of self-ascription needs justification.

Strawson almost certainly thought there is a connection between the two relations, a connection that can be formulated in two propositions. First, he likely thought that a subject enjoys a token experience only if the subject can

self-ascribe it. Thus, “The notion of a single consciousness to which different experiences belong is linked to the notion of self-consciousness, of the ascription of an experience or state of consciousness to oneself” (1966: 98). It is not only the animals and young children that make this problematic. The passage also seems to imply that consciousness requires a certain kind of self-consciousness (a form that is sometimes called reflective self-consciousness). The second connecting proposition that Strawson likely accepted is that a subject can self-ascribe an experience only if he enjoys a sequence of interconnected experiences of an objective world—though this of course would have to be the conclusion of the argument, not its premise.

Even if we restrict our attention to the case of subjects who have the concepts required to self-ascribe experiences, we need to draw a further distinction. The fact that, for such subjects, when *s* owns *e*, it is possible for *s* to self-ascribe *e* does not imply that that fact is what *makes e* one of *s*'s experiences. For there may be some philosophical explanation of why that possibility always holds that traces the possibility to a quite different constitutive origin. That seems to me to be the case here. *e* as a conscious event is in an element of its subject's total subjective state. Thereby *e* can make rational for the subject various judgements, including a judgement of it, demonstratively given, to the effect “That's mine”, or “I am having that experience of such-and-such kind”. It is entirely consistent with this non-Strawsonian, non-Kantian direction of constitutive explanation that whenever there is ownership of a conscious event, there is the possibility of self-ascription. But this ‘whenever’ claim is true not because ownership consists in the possibility of such self-ascription. Rather, it is ownership that grounds the possibility of the self-ascription. In short, in the order of philosophical explanation, Strawson has not established that an “ascription-first” view of these matters is correct, rather than an “ownership-first” view.

An ownership-first view can be correct, and is a rival to such neo-Kantian views as Strawson's, only if there is a background account of the nature and identity over time of the subjects who are said to enjoy the experiences. But that is precisely what the contemporary conception, outlined in the earlier displayed theses, offers. It presents an account of conscious subjects capable of representing themselves that involves identity of integrating apparatus and the possession of files on themselves that does not involve the neo-Kantian materials.

(b) Points (4) through (6) of the argument raise other issues. Strawson says that in even the most fleeting of impressions there is a “component of recognition, or judgement” (1966: 100). Judgement has conceptualized content. Not every subject who has experiences or impressions has to possess mental states with conceptualized content. Even for a subject who possesses concepts, the occurrence of a mental event need not engage or involve the subject's concepts. Once we distinguish judgement from awareness, then even when a subject has a rich conceptual repertoire and conceptualizes his mental events under various kinds, there is no incompatibility between this conception of awareness and the fact that

for pains, for instance, there is no distinction between the experience of the pain and the pain event itself. There certainly would be a problem if awareness were to consist in judgement; but it does not. So the “yet” that starts off the quotation in Point (5) (“Yet at the same time we seem forced to concede”) does not seem to be well motivated; there is no apparent need to see the acknowledgment of such subjective sensations as any kind of concession.

Why is the possibility of self-ascription thought to solve a problem that is alleged to exist for sensations, but which is apparently not thought to exist for objective perceptions of the spatio-temporal world? One can imagine an argument that is motivated as follows. If one fails properly to distinguish awareness and judgement, and one thinks, as apparently Strawson (or at least Strawson’s Kant) does, that all awareness has to involve judgement, then it may indeed seem that in sensation we have collapsed the item/recognition distinction. Such a combination of views may lead one to think that a sensation being of a certain kind cannot come apart from the subject’s thinking that it is. By contrast, it may be said, for objective perception, since it is always possible that some objective state of affairs exists without experience as of it, and conversely, there is no such (alleged) danger of the so-called item/recognition distinction collapsing. Then, this argument may continue, genuine recognition in the sensation case is secured because the possibility of self-ascription means that this is a subjective item. Presumably on this view that means that we can thereby have recognition of kinds without the kind of seems/is distinction that is present for experiences with objective content. The ascription of purely subjective experiences to a subject is not vacuous, on this view, only because the subject is an independent entity that traces a path through the spatio-temporal world.

I confess to being utterly unclear how the argument of this reading is meant to work. Recognition needs to be recognition of the subjective item as of a particular kind—as a pain, as a particular kind of smell or sound, and so forth. Consider the self-ascription to which appeal is made in this version of the argument: does that self-ascription specify the kind of the experience, or is it not? If the kind of the experience is not included, the argument has not provided for what is involved in recognition.

If the kind is included, how has the specific kind of the subjective experience been secured merely by the possibility of self-ascription of a particular experience? Isn’t self-ascription of an experience of a specific kind, and ascription as that specific kind, possible, and made rational, by the experience’s already being of the kind in question? Then it seems that the experience’s being of the kind it is is causally and explanatorily prior to the possibility of proper self-ascription. Without that basis, it is not clear how a specific kind is available for self-ascription on the Strawsonian conception.

Perhaps the argument is meant to be that a conception of the experience and its properties as causally and explanatorily prior, as in the ownership-first view, is some kind of illusion. We say that the experience makes reasonable the self-ascription, but this is, it may be said, a way of speaking that should not be

confused with constitutive or metaphysical priority in this area. The problem is that we have been given no real reason for taking these appearances as illusory. If we combine the conception of subjects outlined in the contemporary conception with an insistence on the awareness/judgement distinction, there will indeed always be the possibility of self-ascription of a sensation as being of a particular kind, when the sensation occurs to a conceptually competent subject. There is no obvious obstacle to the naïve view that the kind of the sensation, at an appropriate level of kind, contributes to the rational explanation of the subject's self-ascription of a sensation of that kind.

There is a parallelism here between this criticism and the earlier criticism of Strawson's discussion of the way links are not "in practice" severed with a bodily subject in 'criterionless' mental self-ascription (in language). In both cases, Strawson's argument aims to succeed in securing an embodied subject, as, respectively, owner or ascriber of the mental state; but in both cases, the rationality and entitlement to the transition involved in each case seems to be left unelucidated in his account.

(c) Point (2) of Strawson's argument, taken in the context of his reasoning, suggests that there is some vacuity in saying "all the experiences have the same owner" when the experiences are purely subjective, and that the vacuity disappears when some at least of the series are perceptions of an objective world, through which the subject traces a route. But having a continuous spatio-temporal route and point of view on an objective world is neither sufficient nor necessary for identity of subject over time or for identity of owner of the experiences. We can conceive of inserting a new brain, of a new subject, into one identical continuing body that traces a continuous path in space. This possibility shows that even if the perceptions of the new subject trace a continuous series of points of view with experiences of the previous subject on that route, that continuity is not sufficient for identity of subject over time. A further condition is needed: continuity of integrating apparatus over time.

This further condition is not vacuous. Moreover, this further condition can also hold without the condition of an experienced spatio-temporal route condition being met. The condition of an experienced spatio-temporal route is not met in Strawson's own sound world in *Individuals*, but the relevant auditory experiences could all be enjoyed by a subject with continuity of integrating apparatus over time.

I am inclined to think that what objective perceptions add to the purely subjective ones is not by itself enough to make sameness of subject intelligible if it were not so already. And when we add what is enough for identity of subject, viz. identity of integrating apparatus, then tracing a continuous spatio-temporal route through an objective world is not obviously required. Indeed it is not met in some cases even when there are objective perceptions. Such is the case of Daniel Dennett's subject in his paper "Where Am I?" (1978), in which a subject whose physical integrating apparatus is in Houston may control, and perceive

from, different bodies at different times. The location of the body controlled by this subject does not need to possess spatio-temporal continuity over time. It may jump around as the subject controls different bodies at different times.

We can also add that if some split-brain cases are possible in which there are two conscious subjects in the same body at the same time, then it is possible that two subjects have exactly the same genuinely perceived spatio-temporal path through the world, yet remain distinct subjects. The two subjects could share a visual perceptual apparatus in such cases. A spatio-temporal path through the world does not fix identity of the subject of experience either at a given time, or over time.

There is, I agree, an important psychological and representational distinction between a subject whose experiences are interconnected because they are of an objective world, and a subject whose experiences are not so interconnected. The former subject, but not the latter, is able to locate itself as an element in the objective order of things and events. On my view, that is a difference between a subject that has an applicable conception of itself as having a location in the world, and a subject who does not have such a conception. That is a distinction in respect of the conception available to two subjects who differ in that respect. It is not a distinction in point of existence of the two subjects. There is further discussion of these points in Peacocke (2012).

(d) Strawson's argument relies upon an alleged experience/concepts connection: "Certainly concepts, recognition . . . would be necessary to a consciousness with any experience at all" (1966: 99). There is by now a fairly extensive literature detailing what would be involved in a notion of nonconceptual content of experience.⁷ Perhaps Strawson's arguments should not be taken as required to dispute the possibility of nonconceptual content. Perhaps much or all of what he says should be translated to the level of nonconceptual content. There is a legitimate notion of nonconceptual recognition of shape, of nonconceptual correctness conditions, and, on my own views, a nonconceptual version of subject-representation that refers to a subject of consciousness. But I do not see how any such translation of Strawson's argument rules out the possibility of a subject of consciousness with a sequence of purely sound, visual-sensational, or olfactory experiences. The account of subjects with updateable subject-files on themselves helps to explain how such subjects could represent themselves as having been in certain states earlier, without having any conception of themselves as occupying spatial locations.

These points are all arguments that the conclusion of Strawson's argument is too strong. If the conclusion is too strong, then there is no point in trying to construct different arguments to the same conclusion (the reasoning is too strong however the conclusion is reached). It does not at all follow that there are not successor arguments to Strawson's that draw in interesting ways on the materials he deployed, to related, though different, conclusions. It would take us too far afield to develop successor arguments here. But I note that

consideration of what is involved in having experiences that do have objective spatio-temporal content, and consideration of the nature of subjective properties and their relation to the public world may each have interesting interconnections and consequences for self-representation in an objective world. The light cast by neo-Kantian arguments may come not from a Strawsonian conclusion, but from what they illuminate in adjacent areas.

I also note a final point of agreement with Strawson, though it has a twist. Strawson writes, at the end of his section 'Unity and Objectivity' in *The Bounds of Sense*,

memory is involved in experience, recognition, consciousness of identity of self through diversity of experience. But it is far too deeply and essentially involved to be capable of being safely handled as if it were a separable and detachable factor which can, say, be conveniently invoked to link up temporally successive or separated episodes into an experiential sequence. If experience is impossible without memory, memory is also impossible without experience. From whatever obscure levels they emerge they emerge together (1966: 111–12).

The non-Strawsonian view of subjects I have endorsed is committed to agreement with Strawson on his point in the final sentence of this most recent quotation. Part of what makes something a continuing and self-representing subject is its capacity for primitive self-representation in memory of its own earlier properties and relations. Self-representing subjecthood and memory really are essentially interdependent on the account I have been offering, and both involve the operation of the subject's file on itself. Neither self-representing subjecthood nor memory is something more primitive than the other: and so neither is available for reductive explication of the other. The twist in this point of agreement is, however, that on the treatment I propose, the way in which personal identity and memory emerge together, via identity of integrating apparatus and the subject's file on itself, is also a way that provides the materials for a non-Strawsonian, non-Kantian account of conscious subjects.⁸

Notes

1. In the past thirteen years, see especially Ameriks (2000), and the references therein to literature to that date, and more recently Longuenesse (2008) and Proops (2010).
2. A350, A363 (which speaks of the "the logical identity of the I"), B413 (which speaks of "a merely logically qualitative unity of self-consciousness in general"). Translations are from the Guyer and Wood edition (Kant 1998).
3. Of course, from the true premise that the subject is not so given in first person thought we should not move to the false conclusion that the subject is not in fact material and extended—the fallacy of moving from "not being given as having property P" to "not having property P". I do not in fact think that Descartes made any such fallacious inference. His arguments for the distinctness of mind and body involve other, further premises.

4. When Longuenesse writes, elaborating Kant, “‘I’ in ‘I think’ does not refer to a permanent *object* whose properties *change*” (2008: 22), we have to agree with the proposition thereby attributed to Kant, because “I” does not refer to a permanent object. But it does refer to an impermanent object whose properties change. That is something we are committed to by our ordinary views, in combination with the thesis that both Kant and Descartes are using “I” with its ordinary sense. This observation does not, however, damage the point that there are consistency and unity constraints on the first person concept that contribute to its having the sense (and thereby the reference) it does.
5. If I have understood him correctly, the quoted formulation from McDowell superficially seems to endorse a “highest common factor” conception of consciousness, a conception he strongly rejects for perceptual experience in its role of ordinary justification of perceptual knowledge. On that rejection, see for instance McDowell (1994: 113) and the references on that page. But his point about awareness of identity over time could be reformulated without such apparent endorsement of something he rejects elsewhere.
6. For further discussion see Peacocke (2012), especially the discussion there of cases described as of “Degree 0”, in which there are subjects who do not self-represent at all. The relation of such subjects to the physical object in which their integrating apparatus is housed is not at all the same as our relation to our own bodies.
7. Examples of my own efforts are in Peacocke (1992: Chapter 3) and Peacocke (2001). See also the important contributions in Burge (2003, 2010).
8. Earlier versions of this material were presented in one of my lectures in the ‘Context and Content’ series in 2010 at the Institut Jean Nicod, École Normale Supérieure, Paris, and in one of my Kohut Lectures at the University of Chicago in 2011, in addition to my regular seminars at Columbia University and University College London. I thank Jonathan Lear, Robert Pippin, François Recanati, and Georges Rey for valuable comments. I was stimulated to think about these issues again by a joint Columbia/NYU seminar ‘Kant and Contemporary Issues’ I gave with Béatrice Longuenesse in the fall of 2007. I thank her not only for many discussions over the years on these and related topics, but for highly illuminating comments on a late draft of this paper, and for further discussion of her own published views on these issues. This final version is significantly different as a result.

References

- Ameriks, K. (2000), *Kant’s Theory of Mind: An Analysis of the Paralogisms of Pure Reason* (Oxford: Oxford University Press, second edition).
- Burge, T. (2003), ‘Perceptual Entitlement’, *Philosophy and Phenomenological Research* 67: 503–548.
- (2010), *Origins of Objectivity* (Oxford: Oxford University Press).
- Campbell, J. (1994), *Past, Space, and Self* (Cambridge MA: MIT Press).
- Dennett, D. (1978), “Where Am I?”, in his *Brainstorms* (Montgomery, VT: Bradford Books).
- Dummett, M. (1981), *The Interpretation of Frege’s Philosophy* (London: Duckworth).
- Kant, I. (1998), *Critique of Pure Reason* tr. P. Guyer and A. Wood (Cambridge: Cambridge University Press).

- Longuenesse, B. (2008), 'Kant's "I think" versus Descartes' "I Am a Thing That Thinks"', in *Kant and the Early Moderns* ed. D. Garber and B. Longuenesse (Princeton: Princeton University Press).
- McDowell, J. (1994), *Mind and World* (Cambridge, MA: Harvard University Press).
- Nagel, T. (1974), 'What Is It Like to be a Bat?', *Philosophical Review* 83: 435–50.
- Perry, J. (2002), 'The Self, Self-Knowledge, and Self-Notions', in his *Identity, Personal Identity and the Self* (Indianapolis: Hackett).
- Peacocke, C. (1981), 'Demonstrative Thought and Psychological Explanation', *Synthese* 49: 187–217.
- (1983), *Sense and Content: Experience, Thought, and their Relations* (Oxford: Oxford University Press).
- (1988), 'The Limits of Intelligibility: A Post-Verificationist Proposal', *Philosophical Review* 97: 463–496.
- (1992), *A Study of Concepts* (Cambridge MA: MIT Press).
- (2001), 'Does Perception Have a Nonconceptual Content?', *The Journal of Philosophy* 98: 239–64.
- (2008), *Truly Understood* (Oxford: Oxford University Press).
- (2012), 'Subjects and Consciousness', in *The Self and Self-Knowledge* ed. A. Coliva (Oxford: Oxford University Press).
- (2012a), 'Explaining *De Se* Phenomena', in *Immunity to Error through Misidentification: New Essays*, ed. S. Prosser and F. Recanati (Cambridge: Cambridge University Press).
- (2012b), 'Descartes Defended', *Proceedings of the Aristotelian Society, Joint Session* 86: 109–125.
- Proops, I. (2010), 'Kant's First Paralogism', *Philosophical Review* 119: 449–495.
- Strawson, P. (1959), *Individuals* (London: Methuen).
- (1966), *The Bounds of Sense* (London: Methuen).
- Wilson, M. (1978), *Descartes* (London: Routledge & Kegan Paul).