

1 Kant: A Unified Representational Base for All Consciousness

Andrew Brook, Carleton University

Kant had the makings of an extremely rich and interesting version of the idea that self-presenting representations are the representational base of consciousness. Stripped down to its bare minimum, he held the view that common or garden representations present not just what they are about, not just their object. They also present themselves and they present oneself, their subject, that is to say, the thing that has them.

The situation with respect to the evidence that he held this view is complicated, however, more complicated than is usually the case even with Kant. Because of the way the mind fit, or did not fit, into his overall projects, everything he had to say about the mind is sketchy and incomplete. Also, Kant's route into his point of view is very different from any of the routes that researchers follow now. As a result of the latter, he does not consider many of the issues that are at the centre of discussion now. So let us start with the overall project and how the mind in general and consciousness in particular fit into it.

1. Kant's Overall Project

Kant's most famous work is the *Critique of Pure Reason (CPR)* of 1781/7 (two editions). For purposes of understanding his views on the mind and consciousness, this work and a small book

worked up from lecture notes late in life, *Anthropology from a Pragmatic Point of View* (1798) are the two most important works.¹ Since the *Anthropology* was based on popular lectures, it is often superficial compared to *CPR*, which therefore is the main work for our purposes. Until middle age, Kant was a conventional rationalist.² Then memories of reading David Hume “interrupted my dogmatic slumbers”, as he put it (1783, Ak IV:260). He called the new approach that ensued *Critical Philosophy*.

Two Main Projects

In the part of the critical philosophy of interest to us, two of Kant’s main projects were to:

- Justify our conviction that physics, like mathematics, is a body of necessary and universal truth (1781/87, B19-21³).
- Insulate religion and morality, including the possibility of immortality and of free will, from the corrosive effects of this very same science (Bxxx).

Kant accepted without reservation that ‘God, freedom and immortality’ (Bxxx) exist but feared that, if science, or any evidence or argument, were relevant to demonstrating or refuting their existence, it would provide reason to doubt their existence. So Kant wanted to insulate such matters from all evidence and argument. Fortunately, as he saw it, neither scientific evidence nor philosophical argument can touch these questions. If so, God, freedom and immortality could be accepted on the basis of faith (and Kant did so accept them) without being at risk from science. “I have found it necessary to deny *knowledge*, ... in order to make room for *faith*.” (Bxxx, his italics).

The first aim occupies much of the first big part of *CPR*, which Kant called the Analytic. The work of insulation comes in the second big part of *CPR*, which Kant called the Dialectic and which consists of a series of attacks on, as he saw them, unjustifiably grand aspirations of the metaphysics of rationalism. ‘Dialectic’ was Kant’s name for a certain kind of faulty reasoning in which one’s conclusions run out far beyond what one’s premises can support. The specific faulty arguments that concern us here are about the nature of the mind or soul and its possible immortality.

In the course of pursuing the first aim, putting knowledge in general and physics in particular on a secure foundation, Kant asked the following question: What are the necessary conditions of experience? More specifically, what must the mind be like for our knowledge to be as it is (A96-7)? Put simply, he held that for our experience, and therefore our minds, to be as they are, our experience must be tied together in the way that physics says it is. So the status of physics is secured. But this also tells us quite a lot about what our minds must be like. In particular, it tells us about how the mind must be able to function.

Notice how the mind enters Kant’s project here. It enters not as an object of interest in its own right but as a means to an end, the end being to justify our conviction that physics is a science. As a result, his explorations of the mind are sketchy and incomplete, carried as far as he needed to carry them for the purposes of his project but no further. As he put it, “This enquiry ... [into] the pure understanding itself, its possibility and the cognitive faculties upon which it rests ... is of great importance for my chief purpose, ... [but] does not form an essential part of it” (Axvii). It is also, he tells us, “somewhat hypothetical in character” (Axvii). (Kant did not retain this passage in the second edition of *CPR* but the sentiments it expresses continued.) There

is no sustained, focussed discussion of the mind anywhere in Kant's work except the *Anthropology*, which, as we said, is quite superficial.

In addition, the two chapters of *CPR* in which most of Kant's remarks on the mind occur, the chapter on the Transcendental Deduction (TD) and the chapter on what he called Paralogisms of Pure Reason (faulty arguments about the mind mounted by his rationalist predecessors), were the two chapters that gave him the greatest difficulty. Indeed, they contain some of the most impenetrable prose ever written. Kant completely rewrote the main body of both chapters for the second edition (though not the introductions, interestingly).

TD is where Kant attempts to realize the first and most important part of the first project, the Paralogisms chapter is devoted to the part of the second project concerned with the mind. In the first edition, he seems to have achieved a stable position on self-consciousness only as late as this chapter, which comes well into the second half of *CPR*. Even his famous term for consciousness of self, 'I think', occurs for the first time in the first edition only in the introduction to the chapter on the Paralogisms.

Model of the Mind Arising Out of the First Project

Kant's exploration of the necessary conditions of experience in the first project led him to a number of substantive claims about the mind. The most famous is his claim that representation requires concepts as well as percepts – rule-guided acts of cognition as well as deliverances of the senses. As he put it in one of his most famous sayings, "Concepts without intuitions are empty, intuitions without concepts are blind" (A51=B75). In more contemporary terms, the

functions crucial for knowledge-generating activity are processing of sensory inputs and application of concepts to sensory inputs. Cognition requires concepts and percepts. As we might say now, to discriminate, we need information; but for information to be of any use to us, we must organize the information.

He also urged that the functions that organize sensory and conceptual raw materials into experiences are different abilities to synthesize. Kant postulated that there are three kinds of synthesis (A98-110; the three have a more diffuse presence in the second edition than in the first, though all of them are still there). Synthesis of Apprehension in Intuition locates the raw materials of experience temporally (and presumably also spatially, though Kant does not say so). Synthesis of Reproduction in the Imagination associates spatio-temporally structured items with other spatio-temporally structured items. And Synthesis of Recognition in a Concept recognizes items using concepts, the Categories in particular. This threefold doctrine of synthesis is one of the cornerstones of Kant's model of the mind.

In fact, Kant held that to organize information as we do, we require two kinds of Synthesis of Recognition in Concepts. The first ties the raw material of sensible experience together into objects (A105). This is now called binding. Put in contemporary terms, initially colours, lines, shapes, textures, etc., are represented separately. For an object to be represented, the contents of these representations have to be integrated.

The second kind of synthesis ties the contents of these individual representations and the representations themselves together so as to produce what might be called a *global representation*, a notion that will prove to be central to his story about self-presenting representations. A global representation connects individual representations and their contents to

one another in such a way that to be conscious of anything thus tied together is to be conscious of other things thus tied, too, and of the group of them as a single group. Kant thought that the capacity to form global representations is essential to both the kind of cognition that we have and the kind of consciousness that we have.

Kant's model of the mind is a model of cognitive function, not underlying mechanisms. This is an effective way to approach the mind, as cognitive science has shown, but Kant had a special reason for adopting it. One of his most deeply held general convictions was that we know nothing of anything as it is. We know things only as they appear to us – including the mind, even our own mind. However, various things he said seem to imply that we *do* have knowledge of the mind – that it must apply concepts, synthesize, and so on. He never addressed the tension squarely but a natural way out for him would have been to distinguish the mind's functions from its composition, what makes it up, and then maintain that what we can know are its functions and what we lack all knowledge of are its composition and makeup. This would be merely a radical version of the functionalist idea that function does not dictate form – a given function could be implemented by systems having very different forms (multiple realizability).⁴

Claims about Consciousness Arising Out of the Second Project

The mind enters the second project somewhat indirectly, too, even though it was here that Kant made many of his most penetrating observations about consciousness, especially consciousness of self.

As we saw, Kant's second project was to insulate 'God, freedom and immortality' from all argument and evidence, fearing that, if either were relevant to the question of their existence

at all, it would provide reason to doubt their existence. Immortality is the topic of interest to us here. His rationalist predecessors, Descartes and Leibniz for example, thought that they could prove that the mind is substantial, is simple (without parts), and persists in a special way. These conclusions would at minimum leave immortality an open possibility. (Thomas Reid took this approach, too, even though he wasn't in other respects a rationalist.⁵) Since the conclusions would appear to many people to follow from things Kant himself had said about the unity of consciousness in the first part of *CPR* (specifically in TD), if his strategy for insulating the possibility of immortality was to work, he also had to show that his earlier claims contain no such implications.

Kant's official topic in the chapter is faulty rationalist arguments and conclusions claiming to yield knowledge of the nature of the mind or soul as it is in itself. However, as he saw it, introspection provides strong *prima facie* support for the rationalist conclusions about what we can know about the mind. In introspection, we appear to ourselves to be substantial, simple and persisting, just as rational psychology held ('rational psychology' was Kant's name for these views). If so, he had to show that introspection reveals nothing of the sort. It was in the course of his deflationary attack on introspection that Kant made many of his most acute observations about consciousness of self.

Kant held surprisingly strong and not entirely consistent views on introspection and empirical methods generally as a basis for knowledge about the mind. Sometimes, he maintained that any empirical study of the mind is hopeless. Of course, he is notorious for harbouring a similar scepticism about chemistry (in his defence, it should be said that there was nothing resembling a single unified theory of chemical reactions in his time). The empirical method most

directly in his gun-sights here was introspection. The key text is in *The Metaphysical Foundations of Natural Science* (1786). After stating his view of chemistry, he went on, “the empirical doctrine of the soul ... must remain even further removed than chemistry from the rank of what may be called a natural science proper” (Ak. IV:471). The contents of introspection, in his terms inner sense, cannot be studied scientifically for at least five reasons.

First, having only one universal dimension and one that they are only *represented* to have at that, namely, distribution in time, the contents of inner sense cannot be quantified; thus no mathematical model of them is possible. Second, “the manifold of internal observation is separated only by mere thought”. That is to say, only the introspective observer distinguishes the items one from another; there are no real distinctions among the items themselves. Third, these items “cannot be kept separate” in a way that would allow us to connect them again “at will”, by which Kant presumably means, according to the dictates of our developing theory. Fourth, “another thinking subject [does not] submit to our investigations in such a way as to be conformable to our purposes” – the only thinking subject whose inner sense one can investigate is oneself. Finally and most damningly, “even the observation itself alters and distorts the state of the object observed” (1786, Ak. IV:471). Indeed, introspection can be bad for the health: it is a road to “mental illness” (‘Illuminism and Terrorism’, 1798, Ak. VII:133; see 161).

At other times, Kant links 'self-observation' to observation of others and calls them both sources of anthropology (Ak. VII:142-3). It is not clear why he didn't respect what he called anthropology more highly as an empirical study of the mind in his critical moments, given that he himself did it.⁶

Whatever, no kind of empirical psychology could ever yield necessary truths about the mind. In the light of this limitation, how *should* we study the mind? Kant's answer was: by the transcendental method using transcendental arguments. If we cannot observe the connections among the denizens of inner sense to any scientific purpose, we can study what the mind *must* be like and what capacities and structures (in Kant's jargon, faculties) it *must* have if it is to represent things as it does. With this method we can find universally true, that is to say, 'transcendental' psychological propositions. We have already seen what some of them are: minds must be able to synthesize and integrate, for example.⁷

Whatever the merits of Kant's attack on the rationalist picture of the mind and its introspective running mates, in the course of mounting it he made some penetrating observations about consciousness of self and the knowledge of self that it yields or does not yield. He urged that,

1. There are two quite different kinds of self-consciousness, consciousness of one's states and consciousness of oneself as the subject of these states.
2. The cognitive and semantic machinery used to obtain consciousness of self as subject is quite unusual. In it, we "denote" but do not "represent" ourselves (A382). Put otherwise, we designate ourselves without noting "any quality whatsoever" in ourselves (A355).
3. When one is conscious of oneself as subject, one is conscious of oneself in a way that does not provide consciousness of features of oneself, a way in which "nothing manifold is given." (B135)
4. One is conscious even of oneself only as one appears to oneself, not as one is.

As he put the latter point,

inner sense ... represents to consciousness even our own selves only as we appear to ourselves, not as we are in ourselves. For we intuit ourselves only as we are inwardly *affected* [by ourselves]. [B153]

And he underpinned all these views with an interesting view of the representational base of consciousness, which he built out of an equally interesting notion of self-presenting representations. Notice that, as with the first project, here too the mind enters Kant's analysis only indirectly, in the service of other ends. Here too the result is a sketchy, incomplete treatment of the aspects of mind thus drawn in.

2. Representational Base of Consciousness

By 'the representational base of consciousness', I mean whatever it is about representation and representing that subserves consciousness. Whether consciousness simply *is* representational or not, it at least *requires* representation of some kind. Kant's view of the representational base of consciousness is mainly an off-shoot of his first project, specifically, of his work on the question about what the mind must be like to have experiences. The core of it is quite straightforward: the representational base of consciousness of oneself and/or one's psychological states is not a special experience of either but having any experience of anything whatsoever. Though he never developed the view in any single discussion or in any detail at all, here is what he seems to have had in mind. Consider the sentence:

I am puzzled by what Kant says about apperception in A107.

Kant's thought seems to have been that my representation of the page, the sentences, and so on, is all the representation that I need to be conscious not just of the page, the sentences, etc., but also of the act of seeing them, and of *who* is seeing them, namely, me. A single representation can do all three jobs. Kant once put it this way: consciousness of representation and self is given "not indeed in, but with ... intuitions" (B161). Let us call an act of representing that can make one conscious of its object, itself, and oneself as its subject the *representational base* of consciousness of these three items. Here is how Kant's story goes.

Two Kinds of Consciousness of Self

The first thing we need to see is that Kant sharply distinguished consciousness of one's own psychological states from consciousness of oneself as subject of those states. Kant's term for the former was 'empirical self-consciousness'. His leading term for the latter was 'transcendental apperception' (TA). (Kant used the term 'TA' in two very different ways, as the name for a faculty of synthesis and as the name for what he also referred to as the 'I think', namely, one's consciousness of oneself as subject. It is the latter usage that is in play here.) In a passage from the *Anthropology*, Kant distinguishes the two kinds of consciousness of self very clearly:

the "I" of reflection contains no manifold and is always the same in every judgment ...

Inner experience, on the other hand, contains the matter of consciousness and a manifold of empirical inner intuition.... [1798, Ak. VII:141-2, emphases in the original]

Here is another passage from the *Anthropology*:

§24. Inner sense is not pure apperception, consciousness of what we are doing; for this belongs to the power of thinking. It is, rather, consciousness of what we undergo as we are affected by the play of our own thoughts. This consciousness rests on inner intuition, and so on the relation of ideas (as they are either simultaneous or successive). [1798, Ak. VII:161]

Kant makes the same distinction in *CPR*:

the I that I think is distinct from the I that it, itself, intuits ...; I am given to myself beyond that which is given in intuition, and yet know myself, like other phenomena, only as I appear to myself, not as I am [B155]

This distinction is reflected in his doctrine of the representational base in the following way. The ‘consciousness of what we undergo as we are affected by the play of our own thoughts’ (and, presumably, perceptions, emotions, memories, and so on) is the consciousness of representations that having the representations gives us. The content of this kind of consciousness varies from representation to representation. The consciousness of oneself as the subject of those thoughts (and ... and ...) is also given to us by having those representations but this consciousness ‘contains no manifold and is always the same in every judgment’.

Representational Base of Consciousness of One’s Psychological States and of Oneself.

The two kinds of consciousness of self may appear to have very different sources. The source of empirical self-consciousness (particular representations) is said to be what Kant called *inner sense*. He did not work out his notion of inner sense at all well but seems to have had

something like perception of one's own psychological states in mind. Here are just a few of the problems. Kant insists that all representational states 'belong to' (presumably he means, 'are presented by') inner sense, including those representing the objects of outer sense (i.e., spatially located objects):

Whatever the origins of our representations, whether they are due to the influence of outer things, or are produced through inner causes, whether they arise *a priori*, or being appearances have an empirical origin, they must all, as modifications of the mind, belong to inner sense. [A98-9]

However, he also says that the body (including one's own) is the object of outer sense; the object of inner sense is the soul. Is outer sense part of inner sense, then? Sometimes he talks like it is, sometimes not. He comes close to denying that we can be conscious of the denizens of inner sense – they do not represent inner objects and have no manifold of their own. Yet he also says that we can be conscious of them – representations can themselves be objects of representations, indeed, representations can make us conscious of themselves. In its role as a form of or means to consciousness of self, apperception ought to be part of inner sense. Yet Kant regularly contrasted apperception, a means to consciousness of oneself and one's acts of thinking, with inner sense as a means to consciousness of – what? Presumably, particular representations: perceptions, imaginings, memories, etc. So do we have two means to consciousness of the same, or some of the same, particular representations, or of different representations?

Whatever exactly Kant meant or should have meant by the term 'inner sense', he said that one way in which we become conscious of the representational denizens of inner sense is quite different from the way in which we become conscious of the objects of those representations and

he said it a number of times. Notice the phrase in the passage quoted in the previous section from §24 of the *Anthropology*: “consciousness of what we are doing” – *doing* (1798, Ak. VII:161).

The way in which one becomes conscious of an *act of representing* is not by receiving what he called intuitions, i.e., sensible contents, but by doing it: “synthesis ..., as an act, ... is conscious to itself, even without sensibility” (B153); “this representation is an act of *spontaneity*, that is, it cannot be regarded as belonging to sensibility” (B132). ‘Sensibility’ and ‘intuition’ are closely related terms referring to the objects or contents of representations of the world, what can be sensed.) Thus Kant is saying that we do *not* become conscious of our own representational acts in the way in which we become conscious of the objects of those acts, by receiving a raw manifold of inputs and then working them up in acts of synthesis. We become conscious of our own representational acts by *performing* them.

So what about consciousness of oneself? What is the source of consciousness of oneself as subject?

Man, ... who knows the rest of nature solely through the senses, knows himself also through pure apperception; and this, indeed, in acts and inner determinations which he cannot regard as impressions of the senses. [A546=B574]

More specifically,

the mind could never think its identity in the manifoldness of its representations... if it did not have before its eyes the identity of its act, whereby it subordinates all [the manifold] ... to a transcendental unity.... [A108]

The acts in question are acts of apperception. Kant is saying that performing acts of apperception is the basis of consciousness of oneself as subject. I am conscious of myself as the single

common subject of a certain group of experiences by being conscious of “the identity of the consciousness in ... conjoined ... representations” (B133). Moreover, we can be conscious of ourselves as subject *merely* by performing acts of representing. No further representation of the act or of oneself is needed.

That the mind is active and knows itself as active was of fundamental importance to Kant. When one is conscious of oneself by doing cognitive and perceptual acts, one is conscious of oneself as spontaneous, rational, self-legislating, free – as the doer of deeds, not just as a passive receptacle for the contents of representations: “I exist as an intelligence which is conscious solely of its power of combination” (B158-159), of “the activity of the self” (B68) (Sellars, 1970-1; Pippin, 1987).

From Individual Representation to Global Representation

So far we have focussed on individual representations. For Kant, however, the acts of representation that serve as the representational base of consciousness of oneself as subject are usually much ‘bigger’ than that, i.e., contain multiple objects and often multiple representations tied together into what Kant called ‘general experience’.

When we speak of different experiences, we can refer only to the various perceptions, all of which belong to one and the same general experience. This thoroughgoing synthetic unity of perceptions is the form of experience; it is nothing less than the synthetic unity of appearances in accordance with concepts. [A110]

Here are some other expressions of what appears to be the same thought. Our experience is “one experience”; “all possible appearances ... stand alongside one another in one experience” (A108).

We have “one and the same general experience” of “all ... the various perceptions” (A110), “a connected whole of human knowledge” (A121). Kant’s term, ‘general experience’, being a bit on the bland side, let us call what he is introducing here a *global representation*.

Transcendental apperception (hereafter TA) now enters. It is the ability to tie ‘all appearances’ together into ‘one experience’.

This transcendental unity of apperception forms out of all possible appearances, which can stand alongside one another in one experience, a connection of all these representations according to laws. [A108]

It performs a “synthesis of all appearances according to concepts”, “whereby it subordinates all synthesis of apprehension ... to a transcendental unity” (A108). This, he thought, requires unified consciousness. Unified consciousness is required for another reason, too.

The introduction of unified consciousness opens up an important new opportunity. Kant can now explore the necessary conditions of conscious content being unified in this way. To make a long story short, Kant now argues that conscious content could have the unity that it does only if the contents themselves are tied together causally. With this, his deduction of the relational categories is complete and his first project, the defence of the necessity of physics, is well under way.

If we now make *global* representations the representational base of consciousness, the story would go this way. When I am conscious of many objects and/or representations of them as the single object of a single global representation, the latter representation is all the representation I need to be conscious not just of the global object but also of the global

representation itself and of myself as the common subject of all the constituent representations.

To reintroduce a relevant passage just quoted,

the mind could never think its identity in the manifoldness of its representations... if it did not have before its eyes the identity of its act, whereby it subordinates all [the manifold] ... to a transcendental unity.... [A108]

I am conscious of myself as the single common subject of a certain group of experiences by being conscious of “the identity of the consciousness in ... conjoined ... representations” (B133).

Alas, it has to be admitted that none of this is nearly as clear in the original texts as my reconstruction would suggest.

Unity of Consciousness

At heart of the notion of a global representation is the unity of consciousness. Though by no means everything in a global representation need be consciously accessible to us (Kant is widely misunderstood on this point), Kant thought that we have unified consciousness of a good deal of what our current global representation represents and that having the representation is the base for being or becoming conscious of the representation itself as a single, unified representation and of oneself as its subject. Indeed, two kinds of unity are required:

1. The consciousness that this subject has of represented objects and/or representations must be unified.
2. The global representation must have a single common subject (A350).

Kant said little about what a 'single common subject' is like, so we won't say anything more about it. He never said what he meant by 'unified consciousness', either, but he did use the notion often enough for us to be able to see what he had in mind.

Kant refers to the unity of consciousness both as the unity of consciousness (A103) and as the unity of apperception (A105, A108). The notion plays a central role in both projects. In the first project, Kant argues that we could not have unified consciousness of a range of items unless we could tie the items themselves together causally. (This argument is not very successful, though it is better than some neo-Kantian moves that have spun off from it [Brook 2004].)

Unity of consciousness is also a central topic of the second project. The first edition attack on the second paralogism (A352) focuses on the unity of consciousness at a given time among other things) and what can (or rather, cannot) be inferred from it about the nature of the mind, the first part of it anyway (how the self appears to itself also makes an appearance) The attack on the third paralogism focuses on what can(not) be inferred from unified consciousness over time. These are all from the first edition of *CPR*. In the second edition, Kant makes further remarks about it, quite unlike anything in the first edition, for example, "this unity ... is not the category of unity" (B131).

By 'unity of consciousness', Kant seems to have had something like the following in mind: I am conscious not only of single experiences of single objects but also of experiences that have many normal objects as their single, integrated object. The same is true of actions; I can do and be conscious of doing a number of actions at the same time. In addition to such synchronic unity or unit at a time, many global representations, as we called them, display diachronic unity or unity across time: current representation is combined with retained earlier representation. In

fact, any representation that we acquire in a series of temporal steps, such as hearing a sentence, will have diachronic unity (A104; A352). Thus, diachronic unity is often part of synthesis of recognition.

Kant himself did not explicate his notion of unified consciousness but here is one plausible articulation of the notion that we find at work in his writings.

The unity of consciousness =*df.* consciousness of a number of representations and/or objects of representation in such a way that to be conscious of any of them is also to be conscious of others of them and of at least some of them as a group.

As this definition makes clear, consciousness being unified is more than just being one act of consciousness. The act of consciousness in question is not just singular, it is unified.

In Kant's view, moreover, to have unified global representations, we must unify them. We must synthesize them using unifying acts of synthesis, acts of transcendental apperception (in one of his uses of this term). It takes a unified consciousness to perform unified acts of synthesis. The unity of consciousness and Kant's views on it are complicated issues (Brook and Raymont, *plato.stanford.edu*). What matters for present purposes is that it is representations providing unified consciousness of their objects that are the representational base of consciousness of one's psychological states and of oneself as their single, unified subject. The notion of a self-presenting representation scattered through Kant's work is a rich one indeed!

If having a unified self-presenting global representation is the representational base of consciousness, is having one also sufficient for consciousness? We need to break this question down into bite-sized chunks. Consciousness of the world around us? Kant never said. Nowadays we would require at least that the representation be cognitively active before we would grant that

it provides consciousness of the world, but Kant did not investigate the matter, part of his general neglect of action and the body. Consciousness of the representation itself? Again, Kant said little that bites directly on the question. However, he did have something to say about the third issue, consciousness of oneself as subject. What he said is both interesting in its own right and relevant to the second question. Kant required only that we *be able* to attach 'I think' to our representations (B130), not that we must always do so. That is to say, we must *be able* to become conscious of our representations, that we are having them, and so on, but we need not actually be conscious of them or of having them. Note that the passage just discussed is from the second edition, which makes more use of consciousness of self than the first edition, where he did not require even this much. So the answer would seem to be, no, Kant did not hold that unified consciousness of the world is sufficient for consciousness of oneself as subject.

If so, what else might be required? Here Kant gives almost no guidance. From the fact that he mentions it only once and in a footnote (B156fn.), a footnote moreover irrelevant to the current issue, it would seem that direction of attention, many theorists' favourite candidate at the moment, would not have been Kant's candidate. However, he never addressed the question so we have no way of knowing what his candidate would have been.

To close off our reconstruction of the picture of representation that seems to make up the substructure of Kant's thought, we should note that he also had the makings of a story about how global representations are implemented in a cognitive system. When the system synthesizes representations and/or represented objects with other representations and/or represented objects by bringing the various items together under concepts, specifically, under the relational categorial concepts, then a unified global representation results.

3. Putting the Self-Presenting Global Representation to Work

We now turn to the interesting question of what work Kant's notion of the self-presenting global representation and the notion of the representational base built on it can do. First, what work did he get out of them? At least this much work.

The job of the Paralogisms chapter is to insulate the possibility of immortality from all possible attack by argument or evidence. He does so by urging that we know nothing of our nature, therefore nothing about whether we might be immortal or not. Therefore, immortality remains a possibility.

As part of mounting this argument, Kant has to show that our consciousness of ourselves yields no knowledge of our nature. Here he has to resist arguments of two kinds. First, if consciousness is to be unified, it would appear that the separate bits cannot be anything like separate parts of a machine. Distribute the words of a verse among a number of items, defined how you like. Nothing will be aware of the whole verse (A352). It would seem to follow that the mind must be simple (i.e., not made of parts). But this would entail that it is not material. If this argument goes through, we would know a lot about the mind, including things directly relevant to the possibility of immortality. So Kant has to resist the argument. Second, we just appear to ourselves to be substantial and simple and to persist in a special way. So Kant has to explain this appearance away.

Moreover, at least by the time of the second edition of *CPR*, Kant had come to see that it would be quite implausible to maintain that when one is conscious of oneself, one is conscious

only of appearances, highly doctored representations of oneself – that even when one is conscious of oneself as the subject of one’s experience, agent of one’s acts, by having these experiences and doing those acts, one has no consciousness of one’s actual self. In the second edition, he reflects this sensitivity as early as B68; at B153, he goes so far as to say that an apparent contradiction is involved.

Kant’s doctrine of global representations as the representational base of consciousness of self contains the basis for everything he had to say on both issues, though it takes some reconstructive surgery on our part to see the connections.

No Manifold in Consciousness of Self

First, Kant held that when one is conscious of oneself as subject, one is not, or need not be, conscious of any properties of oneself, certainly not any properties that are contingent or changing. One has the same consciousness of self no matter what else one is conscious of – thinking, perceiving, laughing, being miserable or whatever. Kant expressed the thought this way,

through the ‘I’, as simple representation, nothing manifold is given. [B135]

And this,

the I that I think is distinct from the I that it ... intuits ...; I am given to myself beyond that which is given in intuition. [B155]

This idea is similar to Shoemaker's notion of 'self-reference without identification', the notion that reference to self as subject involves or need involve not identification of oneself as anything (Brook 2001). Since, on Kant's view, it is not just identifying properties but any properties of oneself whatsoever that one does not need to know in order to refer to oneself as oneself, 'non-ascriptive reference to self' might capture what is special about this form of consciousness of self better than Shoemaker's term.

Consciousness of Self is not Knowledge of Self

The notion that in consciousness of self as subject, no manifold is given, is or can be interpreted as the idea that in it, no properties of the self are represented. (We will see major support for this interpretation in a moment.) It immediately gives Kant what he needs to be able to allow that when one is conscious of oneself as subject, one is conscious of oneself, not just of an appearance of oneself while continuing to maintain that one has no knowledge of oneself as one is. If consciousness of self ascribes nothing to the self, it is or could be a "bare ... consciousness of self [as one is]" that yields no knowledge of self, indeed, that is "very far from being a knowledge of the self" (B158).

But how is ascriptionless reference to self possible? Here Kant offered some remarkably prescient insights about reference to self, insights that next appeared only in Wittgenstein (1934/5) or perhaps even Castañeda (1966) and Shoemaker (1968). Moreover, while neither Castañeda nor Shoemaker had an explanation for them, they flow directly out of Kant's idea of

the representational base of consciousness of self. The ideas first, then their explanation by reference to the representational base.

Referential Machinery of Consciousness of Self

Here are some of the things that Kant said about reference to self as subject. We have already seen him say that it is a consciousness of self in which “nothing manifold is given.” (B135) In the kind of reference in which we gain this consciousness of self, he says that we “denote” but do not “represent” ourselves (A382). We designate ourselves “only transcendently” without noting “any quality whatsoever” in ourselves (A355).

The idea behind these claims is a remarkably penetrating one; remember, the study of reference and semantics generally is usually thought to have begun only with Frege. Kant is anticipating two important theses about reference to self that next saw the light of day 200 years later.

1. In certain kinds of consciousness of self, one can be conscious of something as oneself without identifying it (or anything) as oneself via properties that one has ascribed to it (Shoemaker’s self-reference without identification) ⁸

And,

2. In such cases, first-person indexicals (I, me, my, mine) cannot be analysed out in favour of anything else, in particular anything descriptionlike (the essential indexical) (Perry 1979).

Was Kant actually aware of (1) and/or (2) or had he just stumbled across something that later philosophers recognized as significant?

One standard argument for (1) goes as follows:

My use of the word 'I' as the subject of [statements such as 'I feel pain' or 'I see a canary'] is not due to my having identified as myself something [otherwise recognized] of which I know, or believe, or wish to say, that the predicate of my statement applies to it. [Shoemaker 1968, pp. 558]

A standard argument for (2), that certain indexicals are essential, goes as follows. To know that *I* wrote a certain book a few years ago, it is not enough to know that someone over six feet tall wrote that book, or that someone who teaches philosophy at a particular university wrote that book, or ... or ... or ... , for I could know all these things without knowing that it was *me* who has these properties (and I could know that it *was* me who wrote that book and not know that any of these things are properties of me). As Shoemaker puts it,

no matter how detailed a token-reflexive-free description of a person is, ... it cannot possibly entail that I am that person. [1968, pp. 560]

Kant unquestionably articulated the argument for (1):

In attaching 'I' to our thoughts, we designate the subject only transcendently ... without noting in it any quality whatsoever – in fact, without knowing anything of it either directly or by inference. [A355]

This transcendental designation, i.e., referring to oneself using 'I' without 'noting any quality' in oneself, has some unusual features. One can refer to oneself in a variety of ways, of course: as the person in the mirror, as the person born on such and such a date in such and such a place, as

the first person to do X, and so on, but one way of referring to oneself is special: it does not require identifying or indeed any ascription to oneself. So Kant tells us.

The question is more complicated with respect to (2). We cannot go into the complexities here (see Brook 2001). We will just note three passages in which Kant may be referring to the essential indexical or something like it.

The subject of the categories cannot by thinking the categories [i.e. applying them to objects] acquire a concept of itself as an object of the categories. For in order to think them, its pure self-consciousness, which is what was to be explained, must itself be presupposed. [B422]

The phrase ‘its pure self-consciousness’ seems to refer to consciousness of oneself as subject. If so, the passage may be saying that judgments about oneself, i.e., ascriptions of properties to oneself, ‘presuppose ... pure self-consciousness’, i.e., consciousness of oneself via an act of ascription-free transcendental designation.

Put the above beside this, “it is ... very evident that I cannot know as an object that which I must presuppose to know any object” (A402), and this,

Through this I or he or it (the thing) which thinks, nothing further is represented than a transcendental subject of the thoughts = X. It is known only through the thoughts which are its predicates, and of it, apart from them, we cannot have any concept whatsoever, but can only revolve in a perpetual circle, since any judgment upon it has always already made use of its representation. [A346=B404]

The last clause is the key one: “any judgment upon it has always already made use of its representation”. Kant seems to be saying that to know that anything is true of me, I must first know that it is me of whom it is true. This is something very like the essential indexical claim.

Now we make the connection to the consciousness without knowledge issue. If reference to self takes place without ‘noting any properties’ of oneself, the consciousness of self that results will not provide any knowledge of self. But what right does Kant have to help himself to his notion of transcendental designation?

Transcendental Designation and the Representational Base of Consciousness

The answer is: Every right. Even though he did not explicitly make the connection himself, that reference to self as subject is ‘transcendental’, i.e., non-ascriptive, flows directly from the idea that the representational base of such reference is having global representations. A global representation will have no neighbour, to use a phrase of Wittgenstein’s. That is to say, while we can distinguish the subject from all objects (A342=B400), we cannot compare it to, contrast it with, one subject *rather than another*. There will no other simultaneous global representation of which one is conscious in the same way (namely, by having it, though Kant never says this explicitly) that presents a subject from which to distinguish the subject presented in one’s current global representation.

If not, reference to self as subject will not, to use a phrase of Bennett’s, be experience-dividing – “i.e. [statements expressing it have] no direct implications of the form ‘I shall experience C rather than D’” (Bennett, 1974, p. 80). In a statement such as,

I am looking at the words on the screen in front of me, the verb expression or the object expression may divide experience but the subject expression will not. If so, awareness of self as subject can proceed and, to the extent that ascription requires dividing experience, must proceed without any ascription to self.⁹ From this it follows that consciousness of self as subject will have all the special features delineated earlier: reference that does not ascribe, manifoldlessness, and a consciousness of self that yields no knowledge of self. In this account, Kant goes further than any theorist past or present.¹⁰

Kant's Notions and Current Issues

Can Kant's notion of the representational base or the rich notion of a self-presenting global representation that underlies it help us with any contemporary issues? Perhaps. Start with the question, in virtue of what are certain states conscious states? Kant did not address this question but he seems to have assumed that when having a global representation is not just the base of being conscious of it but is actually providing such consciousness, then the global representation will simply *be* a conscious state – nothing else needed. (Thus, Kant would probably have been very dubious about the idea that *qualia*, felt qualities, are somehow different from and outstrip the representational properties of representations.)

What about the relationship of the various forms of consciousness? Earlier we saw that Kant allowed that one can have a global representation of, and therefore presumably be conscious of, the world without being conscious of the representation or oneself. All he demanded is, as we saw, that one be *able* to attach 'I think' to each such representation. If global representations are self-presenting, he might well have insisted that they provide all the

representation needed, that nothing representational could be missing. But he could happily have allowed, indeed the claim about ‘I think’ suggests that he would have allowed, that something else needed for consciousness could still be missing.

Certainly Kant’s view has implications for some contemporary doctrines about consciousness. In common with most views anchored in self-representing representations, his view entails that nothing higher-order is needed for consciousness. *Pace* Rosenthal (1991), each representation can present itself; no higher-order state taking the first representation as its object is needed.

Equally, Kant’s approach puts real pressure on the so-called transparency thesis. The transparency thesis is the thesis that we are not directly, noninferentially aware of our own representations. Being aware of representing redness is no more than being aware of the redness represented (Harman 1990, Dretske 1995). We are aware *via* representations but not *of* representations. On Kant’s view, having a representation can make us just as conscious of the representation as of what it represents. If Kant is right, the transparency thesis is wrong.¹¹

In short, Kant’s notion of self-presenting representations, the notion that representations present not only their objects but also themselves and their subject, the thing that has them, and the attendant theory of the representational base that he sketched and hinted at in various places are unusually rich and interesting, more rich and more interesting than most contemporary discussions of these and related issues.¹²

Endnotes

References

Works by Kant Cited

Kant, I. 1781/1787. *Critique of Pure Reason* (trans. P. Guyer and A. Woods). Cambridge and New York: Cambridge University Press, 1997. (I have consulted this translation and the Kemp Smith translation but decided translations for myself.)

References to *CPR* are in the standard pagination of the 1st (A) and 2nd (B) editions. A reference to only one edition means that the passage appeared only in that edition. All other references to Kant are in the pagination of the *Gesammelte Schriften*, ed.

Koniglichen Preussischen Academie der Wissenschaften, 29 Vols. Berlin: Walter de Gruyter et al., 1902- (in the format, Ak.XX:yy).

Kant, I. 1783. *Prolegomena to Any Future Metaphysics* (trans. P. Carus, rev. with intro. by James Ellington). Indianapolis, IN: Hackett Publishers, 1977 (Ak. IV).

Kant, I. 1786. *The Metaphysical Foundations of Natural Science* (trans. with intro. by James Ellington). Indianapolis, IN: Library of Liberal Arts, 1970. (Ak. IV).

Kant, I. 1798. *Anthropology from a Pragmatic Point of View* (trans. Mary Gregor). The Hague: Martinus Nijhoff, 1974 (Ak. VII).

Other Works Cited

Ameriks, K. 2000. *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*, 2nd edition. Oxford: Oxford University Press.

- Bennett, J. 1974. *Kant's Dialectic* Cambridge: Cambridge University Press.
- Brook, A. 1993. Kant's *A Priori* Methods for Recognizing Necessary Truths. In P. Hanson and B. Hunter, eds. *Return of the A Priori, Canadian Journal of Philosophy Supplementary Volume 18*, pp. 215-52 .
- Brook, A. 1994. *Kant and the Mind*. Cambridge and New York: Cambridge University Press.
- Brook, A. 2001. Kant on self-reference and self-awareness. In A. Brook and R. DeVidi, eds. *Self-Reference and Self-Awareness*. Amsterdam: John Benjamins.
- Brook, A. and P. Raymont. Forthcoming. *A Unified Theory of Consciousness*. Cambridge, MA: MIT Press.
- Brook, A. and P. Raymont. The unity of consciousness. *The Stanford Electronic Encyclopaedia of Philosophy*. URL: plato.stanford.edu.
- Castañeda, H.-N. 1966. 'He': A Study in the Logic of Self-Consciousness. *Ratio* 8, pp. 130-57.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Harman, G. 1990. The intrinsic quality of experience. In J. E. Tomberlin, ed. *Philosophical Perspectives 4: Action Theory and the Philosophy of Mind*. Atascadero, CA: Ridgeview Press.
- Meerbote, R. 1989. Kant's functionalism. In J. C. Smith, ed. *Historical Foundations of Cognitive Science*. Dordrecht, Holland: Reidel.
- Perry, J. 1979. The essential indexical. *Nous* 13, pp. 3-21.

Pippin, R. 1987. Kant on the spontaneity of mind. *Canadian Journal of Philosophy* 17, pp. 449-476.

Rosenthal, D. 1991. The independence of consciousness and sensory quality. *Philosophical Issues* 1, pp. 15-36.

Shoemaker, S. 1968. Self-reference and self-awareness. *Journal of Philosophy* 65 (20), pp. 555-567.

Shoemaker, S. 1970. Persons and their pasts. *American Philosophical Quarterly* 7, pp. 269-285.

Sellars, W. 1970. "... this I or he or it (the thing) which thinks ...". *Proceedings of the American Philosophical Association* 44, pp. 5-31.

Wittgenstein, L. 1934-5. *Blue and Brown Books*. Oxford: Basil Blackwell Publishers.

1. He was already 57 when he wrote the first edition of the first *Critique*, yet he went on to write two further *Critiques*, the *Critique of Practical Reason* (1788) on moral reasoning and the *Critique of Judgment* (1790), a work devoted to a number of topics including reasoning about ends, the nature of judgment, and aesthetics, and books on natural science, cosmology, history, geography, logic, anthropology – the list is long.

2. To over-simplify and sticking to the official line on the topic, rationalists hold that we can discover truths just by thinking about things, without need of observation or experience.

3. All further references to Kant will be to this work unless otherwise noted. ‘A’ refers to passages in the first edition, ‘B’ to passages in the second. ‘=’ means that the passage occurs in both editions.

4. As Meerbote (1989) and many others have observed, Kant held a functionalist view of the mind almost 200 years before functionalism was officially articulated in the 1960s by Hilary Putnam and others. Kant even shared functionalists’ lack of enthusiasm for introspection, as we will see, and their belief that we can model cognitive function without knowing anything very much about underlying structure. Given his personal hostility to materialism about the mind (1783, Ak. IV, end of §46), he would have found the naturalizing tendencies of much contemporary functionalism repugnant.

However, because the unknowability of things as they are in themselves, one of his most deeply-held views, entails that one must be utterly neutral about what the underlying composition of the mind might be like, he would have had to allow that multiple realizability and even naturalism are open *intellectual* possibilities, however repugnant they might have been to him or dangerous to things of the deepest importance to him, namely, that we have free will and that personal immortality is possible (for a fine discussion of these issues, see Ameriks 2000, postscript).

5. The relationship of Kant to Reid is interesting. Among other things, they were near-contemporaries and some of the phrasing of views that Kant attacks is very similar to phrasing that Reid used. I explore the matter briefly in 1994, pp. 191-2.

6. Indeed, what Kant called anthropology is very like what we would now call behavioural psychology. He lectured on the subject every year for decades and finally pulled his lectures together into the little *Anthropology* (1798). Yet not only does he not exempt what he called anthropology from his strictures on the empirical doctrine of the soul, he does not even mention it when he is attacking empirical methods!

7. The core of a transcendental argument is what we now call inference to the best explanation: postulate unobservable mental mechanisms to explain observed behaviour. To be sure, Kant thought that he could get more out of his arguments than just ‘best explanations’. He had a tripartite doctrine of the *a priori*. He held that some features of the mind and its knowledge had *a priori* origins, i.e., must be in the mind prior to experience (because using them is necessary to have experience). That mind and knowledge has these features are *a priori* truths, i.e., ‘necessary’ and ‘universal’ (B3/4). And we can come to know these truths, or that they are *a priori* at any rate, only by using *a priori* methods, i.e., we cannot learn these things from experience (B3). Kant thought that transcendental arguments were *a priori* or yielded the *a priori* in all three ways (Brook 1993). Nonetheless, at the heart of this method is inference to the best explanation. The latter was just the approach taken up by researchers when introspection fell out of favour about 100 years ago. Its nonempirical roots in Kant notwithstanding, is now the major method used by experimental cognitive scientists.

8. (2) is often taken to be closely related to another putative peculiarity of consciousness of self, what Shoemaker calls immunity to error through misidentification with respect to the first person

(Shoemaker 1970, who claims to have found the core of the idea in Wittgenstein 1933-4, pp. 66-70).

We will not explore this issue here (see Brook 2001).

9. So awareness of self as subject also does not distinguish me from or identify me with anything of which I am aware as an object, anything in “the world”. Something of great interest for Kant’s second project follows: so far as anything my awareness of myself as subject could tell me, I could be *any object* or *any compilation of objects* or *any succession of objects* whatsoever. Not by accident are these exactly the topics of the first three Paralogisms. That is to say, Kant himself put the idea of transcendental designation to work to explain how one can appear to oneself to be substantial, simple and persisting without these appearances reflecting how one actually is. The reason that one appears in these ways is not that the self is some strange, indefinable being. It is because of the kind of referring that we do to become conscious of oneself as subject. One of the mistakes of rational psychology was to take the simplicity (lack of manifoldness) in the unified representation of self to be a representation of simplicity and unity in the self.

10. This is a short version of a rather long story. For the full story, again see Brook 2001.

11. There is some reason to think that Kant might be right here. Being conscious of touching something rather than seeing it, being conscious of seeing it dimly rather than clearly – that we make these and a thousand other discriminations, while perhaps not conclusive, certainly suggests that we are conscious, not just of what representations represent, but also of the representations themselves.

12. Brook and Raymont, forthcoming, attempt a contemporary account that takes full account of Kant. Thanks to Uriah Kriegel, Rick DeVidi, Julian Wuerth, and especially Paul Raymont for helpful comments and discussions, in Paul Raymont’s case many, many, many such discussions.