Kant and Cognitive Science[1]

Andrew Brook[2]

Introduction

Immanuel Kant (1724-1804) has a serious claim to be the single most influential figure in the pre-20th century history of cognitive research. His influence continues to be so deep-running that in many respects he is the intellectual grandfather of contemporary cognitive science. Consider the widely-held view that sensory input has to be worked up using concepts or concept-like states, or the conception of the mind as a system of functions that lies behind the view. Kant originated the first view and worked the second up into something more than a sketch for the first time (the basic idea can be found in Aristotle, Descartes and Hobbes). Both views were central to Kant's model of knowledge and mind and they came to contemporary cognitive science from him by a direct line of descent. In Section 2, we will explore these influences.

Some great thinkers of the past may now be merely cultural artefacts, intriguing and historically significant but long since superseded. In one of the most patronizing comments ever made about a philosopher, William James expressed just that attitude about Kant:

> Kant's mind is the rarest and most intricate of all possible antique bric-à-brac museums, and connoisseurs and dilettanti will always wish to visit it and see the wondrous and racy contents. The temper of the dear old man about his work is perfectly delectable. And yet he is really ... at bottom a mere curio, a 'specimen' [James 1907, p. 269].

Kant is more than a cultural artefact, however. If the Kantian cast of much of contemporary cognitive science is striking, what cognitive science has not assimilated from Kant's work is equally striking.

As well as the ideas just mentioned about the relation of concepts to sensory input and the functional nature of the mind, Kant also held that processes of synthesis, mental unity, and consciousness are central to cognition as we know it and he had some highly original views about self-consciousness. Until recently, these ideas have played no role in cognitive science (what might turn out to be related ideas are now beginning to appear in some quarters). Far from Kant's work being superseded by work in the past half-century on cognition, much of what Kant has to offer has not even been assimilated by it, to its detriment. What cognitive science has not taken over from Kant's work on cognition will be the topic of Sections 3 and 4.

https://doi.org/10.36311/2318-0501/2014.v2n02.4115

1. Biography and Writings

Kant was the last great thinker of the German Enlightenment. As was true of most Enlightenment thinkers, he took the human individual and his or her experience of self and world to be the fundamental unit of analysis. Kant was born in 1724 and lived a very long life, dying just before his eightieth birthday in 1804. Though one-quarter Scottish (it is said that 'Kant' is a Germanization of 'Candt'), he lived his whole life in Königsberg in what was then East Prussia. (The area is now called Kaliningrad and is an autonomous region of Russia located just below Lithuania.) He was a devoutly religious man, though hostile to many forms of conventional religious observance, and came from a very humble background. By the time of his death, he had been rector of the University of Königsberg and was virtually the official philosopher of the German-speaking world.

Until middle age, he was a prominent rationalist in the tradition of Leibniz and Wolff. Then recollection of David Hume (probably Hume's *Enquiry)*, "interrupted my dogmatic slumbers", as he put it (*Prolegomena to Any Future Metaphysics*, Prol, AA 04: 260).[3] He called the new approach that ensued Critical Philosophy. One of its fundamental questions was, what must we be like to have the experiences that we have? The view of the mind that Kant developed in the course of answering this question framed all subsequent cognitive research until the advent of connectionism and dynamic systems theory.

Philosophy of mind and knowledge were by no means the only areas in which Kant made seminal contributions. He founded physical geometry. (Since it is said that he never travelled more the fifty kilometres from Königsberg in his whole life, field work was clearly not important.) Kant had a wide circle of sea-faring friends whose company he preferred to the company of academics; apparently Kant learned his physical geography from his many, many discussions with them.) His work on political philosophy grounds modern liberal democratic theory. And his deontological approach to the justification of ethical beliefs put ethics on a new footing, one that remains influential to this day. (Deontology is the approach of deducing ethical propositions from more general, factual propositions about the nature of the person, or the requirements of rationality, or some other factor that far transcends the specific domain of ethical thought.) He taught mechanics, theoretical physics, algebra, calculus, trigonometry and history, in addition to metaphysics, ethics, and physical geometry, an almost unimaginable range of topics for anyone now.

Kant's most famous work is the *Critique of Pure Reason*, which discusses perception, science, and the mind, among other things. He was already 57 when it was published in 1781 (his Humean awakening came relatively late in life). In addition to this work, he wrote two further *Critiques,* the *Critique of Practical Reasoning* (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. The three *Critiques* are only a tiny portion of his corpus. He wrote books on natural science, cosmology, history, politics, geography, logic—the list is long. For our purposes, the two most important books are the *Critique of Pure Reason* just mentioned and a small book that he worked up from lecture notes and published only when he became too old to lecture any longer, *Anthropology from a Pragmatic Point of View* (1798).

By the time the Critical Philosophy reached full maturity in the *Critique of Pure Reason* (hereafter the first *Critique),* Kant aimed to do two principal things with it:

- Justify our conviction that mathematics and especially physics are a unified body of necessary and universal truth.
- Insulate morality and religion from the corrosive effects of this very same science.

The reason for (2.) was that for Kant there was not the slightest doubt that moral responsibility and God exist but, as he also thought, the universal causal determinism and mechanism of science would undermine both morality and religion if scientific evidence and argument are relevant to them. If so, morality and religion can survive modern science only if the truth of the propositions of science is irrelevant to the truth of the propositions of morality and religion. This Kant attempted to show, primarily by arguing as follows: 1. science is about how things appear to us, morality and religion are about how things are; 2. we can never know that the mechanism and determinism that we find in things as they appear to us reflects their real nature; therefore, 3. we are free to form our views about how things really are on the basis of factors other than scientific evidence, the latter being merely evidence concerning how things appear to us.

It was the pursuit of the first aim, the aim of putting mathematics and science on a secure footing, that led Kant to his views about the mind. This came about in the following way. Kant approached the foundations of mathematics and science by asking: What are the necessary conditions of experience? Now such conditions could be found in two places—in what our experience and the objects of our experience must be like, and in what *we* must be like to have such experience. It was in the former that Kant uncovered the foundations for mathematics and physics that he sought, in particular in the conditions of our experience having objects at all, but Kant went after the conditions of our *experience* having object via the latter: What must *we* be like to experience objects? Though this question is, as Kant once put it, strictly speaking inessential to his main task (Axvii), it led him to his discoveries about the mind.

From the point of view of contemporary cognitive science, two things about Kant's approach are interesting. The first is this. Like contemporary cognitive science but radically unlike other philosophies of his own time, Kant was blithely unconcerned about the great questions about knowledge of the external world, scepticism, solipsism, etc. His target is human knowledge, that is to say, objectively valid perception and belief, and he was a successor to Descartes, Berkeley and Hume. However, his concerns are strikingly different from theirs. Unlike the tradition but like contemporary cognitive science, he simply took it for granted that we have knowledge: *a priori* knowledge about conceptual structures and perceptual knowledge of the world of space and time. What interested him is how these various types of knowledge hang together. In any case, as he argued, our access to ourselves is neither better nor worse than our access to the spatiotemporal world, so the contrast at the centre of, for example, Descartes' account of knowledge between the access we have to ourselves and the access we have to things elsewhere in space does not obtain. (We will return to this point below.) Kant had concerns that go beyond those of contemporary cognitive science, of course, in particular his

negative interest in showing that knowledge has limits in order to make room for faith (Bxxx). Nevertheless, his positive interest in knowledge is strikingly like the interest of contemporary cognitive science.

This brings us to the second point about Kant's approach. If Kant's interest in knowledge was like contemporary cognitive scientists', his methods were utterly different in one crucial respect: unlike virtually all cognitive researchers now, Kant held that an empirical science of psychology is impossible. As he argues in a famous passage in *The Metaphysical Foundations of Natural Science*, "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" (MAN, 04: 471). (Kant's notorious remark about chemistry was made before it had been reduced to a single quantified theory.) First, mental states have only one universal dimension, namely distribution in time, so their contents cannot be quantified; this make a mathematical model of them impossible. Second, there is no objective basis for deciding where one mental state stops and the next ones starts. 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, each person can study the mental states of only one person, namely, him- or herself. Finally, "the observation itself alters and distorts the state of the object observed". (Little did Kant know how big an issue that would become!)

How then can we study the mind? For Kant, the answer was: by *a priori* reasoning—we study what the mind *must* be like and what capacities it *must* have to represent things as it does. He called this the transcendental method; as we will see, it came to have a huge influence on the research programme of cognitive science, its nonempirical roots notwithstanding.

Everything I have just said about Kant's hostility to the idea of an empirical science of psychology is true but it may also be misleading. Kant did not mean by 'psychology' what we mean by it. For him, psychology is the study of what we can be aware of in ourselves via introspection. What we would now call psychology, largely the study of behaviour and the causal context of behaviour, he gave the name 'anthropology'. To see the contrast, we need to return to a work mentioned earlier, *Anthropology from a Pragmatic Point of View*. In this unjustly-neglected work, Kant tells us that anthropology is the study of human beings from the point of view of their behaviour, especially behaviour toward one another, and of the things revealed in behaviour. Anthropology in this sense contrasts with what Kant understood as empirical psychology, namely the introspective observations of our own mental states. Kant's rejection of introspection and turn to behaviour have a very contemporary feel to them. (For more on Kant on introspection, see Brook 2004.)

The *Anthropology* is important for other reasons, too. In particular, it illuminates many things in Kant's picture of cognition. To make sense of behaviour, character, etc., Kant urges early in the work, we must know something of the powers and faculties of the human mind: how it gains knowledge and controls behaviour. Thus, before we can study character, etc., we must first study the mind. In fact, this study of the mind (Anthropological Didactic, he calls it) ends up being three-quarters of the book. In it, Kant discusses many topics more clearly than anywhere else. In one amusing passage, Kant indicates that he was, if anything, even more

hostile to the use of introspection to understand the mind than I have indicated. Introspection, he tells us, can be a road to "mental illness" (Anth, AA 07:161). Strangely enough, Kant never seems to have asked whether anthropology in his sense could be a science.

2. Where Kant Influenced Cognitive Science

Kant's influence on cognitive science was via a direct line of descent through 19th century cognitive researchers such as Herbart and Helmholtz, to turn-of-the-century thinkers such as Freud and James (even though James ridiculed Kant, his model of the mind is still quite Kantian) and on to contemporary researchers such as Fodor and the classical symbolic cognition model of classic cognitive science, 1970's-style.

Of the ideas about cognition that came down to cognitive science from Kant, probably the best-known is the doctrine that representation, much of it at any rate, requires concepts as well as percepts—rule-guided acts of cognition as well as deliverances of the senses. This doctrine has become as orthodox in cognitive science as it was central to Kant. Its origins in Kant are well-known; as he put it, "Concepts without intuitions are empty, intuitions without concepts are blind" (A51=B74). The idea, put in more contemporary terms, is that to discriminate anything from anything, we need information on which to base the discrimination; but for information to be of any use to us, we must also bring capacities to discriminate to it.

Second, Kant's central methodological innovation, the method of transcendental argument as he called it, has become a major, perhaps the major, method of cognitive science. One way to describe the role of transcendental arguments is to say that they attempt to infer the conditions necessary for some phenomenon to occur. Other ways include: they are used to infer the constraints on any such phenomenon occurring, and, they are used to infer what must be true of a system which could contain that phenomenon. This method is important in cognitive science because it provides a toehold on which to climb from observable behaviour to unobservable psychological antecedents. Transcendental arguments are a way of identifying constraints on what the unobservable antecedents could be like. So closely linked is this method to Kant that Flanagan, for example, evens calls it the method of transcendental deduction, Kant's term for his most important form of analysis (1984, p. 180).[4]

Here is a simple example: in an early experiment aimed to tease out how memory works, subjects were asked to memorize a short list of random items (letters of the alphabet or whatever). They were then asked if a given letter occurred in the list. Researchers then reasoned as follows. If it takes subjects less time to find the target letter when it is near the beginning of the list than near the end, memory must work by serial search starting at the beginning of the list. If it takes subject the same length of time no matter where the target item occurs, they must be doing an exhaustive search. If, on the other hand, reaction times vary but unsystematically, then it is likely that subjects are searching only till they find the target item but entering the list at randomly varying points. And so on. We do not need to describe the whole range of possibilities to see the Kantian point: what researchers *actually observed* was reaction times. They then *inferred*

something about underlying mechanisms as causes of the observed behaviour. This is precisely the method of transcendental argument as Kant described it.

Third, even Kant's general conception of the mind and what we can and cannot capture in our models of it has been taken over by cognitive science and philosophers associated with it, at least in a general way. In the light of what cognitive science has not taken up in Kant's model of the mind, this may seem a bit strange, so let us explore the matter further. In cognitive science at the moment, functionalism, specifically the functionalist version of the representational model of the mind, is virtually the official philosophical view of the mind, eliminativist antagonists such as P. M. Churchland (1984) and P. S. Churchland (1986) notwithstanding. The basic idea behind functionalism is this. The way to model the mind is to model what it does and can do, that is to say, to model its functions (in the words of one slogan, 'the mind is what the brain does'). In representational models, the basic function of a mind is to shape and transform representations. Kant too held a representational model of the mind and he too viewed the mind as a system of functions for applying concepts to percepts.

The three tenets of Kant's model of the mind are as follows. (I) Most or all representation is representation of objects; such objects are the result of acts of synthesis. (ii) For representations of objects to be anything to anyone, they must "belong with others to one consciousness" (A116); for this, the mind must synthesize its various objects of representation into what I will call the *global object* of a *global representation* (these terms are defined in Section 3). (iii) Synthesis into either individual or global objects requires the application of concepts. —These are the central elements of the model. All three tenets describe either functions or conditions on functions operating (unified consciousness, for example). Kant even called them functions (A68=B93, A94 and elsewhere). In general, like functionalism, Kant's approach to the mind is centred on how it works, as opposed, for example, to how such a system might be physically constituted. He even shared functionalism's lack of enthusiasm for introspection.

Functionalism now comes in many flavours—that mental content can only be specified by its relationship to other mental content (plus, perhaps, the environment and the subject's history); that explanation of mental functioning is a special sort of explanation (focussing on reasons for action); that explanation of mental functioning must be conducted in the language of psychology; that this vocabulary and the style of explanation conducted by using it have 'autonomy' (cannot be reduced to nonpsychological explanation); that this autonomy stems from such explanation being holistic in certain ways; and perhaps others. Kant had no notion of such variation, of course. Kant's functionalism was of a rather general sort. Nevertheless, I think it is fair to view his model of the mind as a precursor of functionalism.

The thought that Kant was a functionalist *avant le mot* is not new. Sellars (1970) was perhaps the first to read Kant as a functionalist or protofunctionalist; more recently Dennett (1978), Patricia Kitcher (1984), Meerbote (1989), Powell (1990) and others have joined him. (Sellars, 1968, also offers an early version of functional classification, and in a Kantian context.) It is less often noticed that Kant was committed to a vital negative doctrine of functionalism, too, the dictum that function does not determine form.

About the relation of function to form, functionalists maintain two things: (I) mental functioning could be realized in principle in objects of many different forms; and, (ii) we know too little about the form or structure of the mind at present to say anything useful at this level in any case, except that mental functions will never be straightforwardly mapped onto any forms that may be associated with them. Kant accepted a variant of both these positions. Concerning (ii), Kant maintained not just that we know little about the 'substrate' (A350) that underlies mental functioning but that we know nothing about it. This is his doctrine of the unknowability of the noumenal mind. If the noumenal mind is unknowable, however, (I) immediately follows; the mind as it is has to be able to take different forms. Otherwise, how it functions would tell us how it is. Indeed, function imposes so few constraints on form that, so far as we can infer from function, we cannot determine even something as basic as whether the mind is simple or complex (A353). In short, Kant not only accepted the notion that function does not dictate form, but accepted a very strong version of it.

Indeed, his doctrine of the unknowability of the noumenal mind is little more than a strong version of that idea, at least on some readings. And noumenalism is no mere personal fancy in his system. On the contrary, the doctrine was absolutely vital to him. The very possibility of free will and immortality hang on it, and our belief in freedom and immortality are two of the three great practical beliefs whose possible truth Kant wrote the whole first *Critique* to defend (Bxxx). (The third was belief in God's existence; the possibility of its truth depends on noumenalism, too, but noumenalism about the world, not the mind.) The first *Critique* has other goals too, of course—more positive, theory justifying goals. But noumenalism is vital to the work's practical goals.

In short, three of Kant's most central insights have been embraced by cognitive science:

- his epistemological insight into the interdependence of concepts and percepts in experience;
- his main method, the method of transcendental argument; and,
- his general picture of the mind as a system of concept-using functions who task is to manipulate representations.

Indeed, some workers in cognitive science have even explored the implications of more specific aspects of Kant's model of the mind for their work, Martindale (1987) for example.

Let us turn now to ideas of Kant's that have not been taken up by cognitive science so far.

3. Ideas of Kant's That Have Not been Taken Up by Cognitive Science

If some of Kant's ideas have been taken over by cognitive science, came down indeed via a direct line of descent, others have not been taken up by cognitive science at all, not from Kant and not from anywhere else either. Before we begin our investigation of the latter, it would be helpful to say a word about the general nature of what cognitive science has and has not taken over from Kant. There are some systematic differences between the two groups of ideas.

Begin with the commonly accepted point that cognitive science has made better progress with mental content (information bearing states of certain kinds) and the processing of content (cognition) than it has with consciousness. This obtains, most would agree, whether it is consciousness of external objects of which we are speaking or consciousness of self. It would be natural to expect that what has been taken over from Kant and what has not would split along the same fault line. That would be only partly true. Until very recently, cognitive science has certainly not paid much attention to consciousness, neither of objects nor of self, but it has paid equally little attention to other aspects of the mind that Kant emphasized, one kind of synthesizing power that we have and the mind's unity in particular. Nor has the explosion of writing on consciousness since about 1985 changed things with respect to either topic. What makes this absence so peculiar is that no feature of cognition and consciousness is more obvious than that we tie the various elements of our experience together, and that what results is a single, unified representations of the world.

Kant held that cognition of the sort that we have requires two kinds of synthesis. The first ties the raw material of sensible experience together into objects. The second ties these individual representations together into what I will call *global representations* (to be introduced shortly).

The first kind of synthesis is to be found in contemporary research in the form of the notion of binding (in the psychological, not the linguistic sense). It has been the object of considerable attention. However, the second has hardly received any attention at all. Here is one standard picture of binding. Colours, lines, shapes, textures, etc., are represented in widely dispersed areas of the brain. These dispersed representations have to be brought into relation to one another if they are to become parts of a single representation of an object.

Interestingly, one influential current model of binding even parallels Kant's in important ways, namely, the model developed by Treisman and her colleagues (1980). Though they do not indicate any awareness of the parallel, like Kant they hold that three stages of visual processing are involved. First the content of feature modules are applied to the input of the senses, next the result of this application of stored features to sensible input is located on a map of locations, and then the result of both processes is recognized via a recognition network and object files. Kant too had a three stage model of synthesis of objects and Treisman's stages parallel his stages of apprehension, reproduction, and recognition in concepts very closely.

Binding is only one of the two forms of activity to which Kant gave the name synthesis, however. The other is the activity of tying multiple representations together into a global representation. What is special about a global representation is the unity that it displays. It is a single representation, and it is single by virtue of connecting the representations that are its parts to one another in such a way that to be aware of any part of the representation is to be aware of other parts of it, too, and of the collection of them as a single group. Let us try to capture the unities we are discussing more formally.

The unities all begin with the activity of forming multiple representations into what we might call a global representation. We can define a global representation as follows:

> A global representation =*df.* a representation that a number of objects or contents of representation, and usually a number of ways of representing, as its single global object.

We can then define 'single global object':

> A single global object =*df.* an intentional object that represents a number of intentional objects such that to be aware of any of these objects and/or their representation is also to be aware of other objects and/or their representation and of the collection of them as a single group.[5]

As a very simple example, each person reading this chapter is aware of the words I have written, the page of the book, the surrounding room, various bodily sensations, thoughts about what I have written ('why is he taking me through this silly exercise?'), and so on. And each of us is aware of these various things not individually but all together, as the single complex object of a single representation.

Kant thought that the capacity to form global representations is absolutely essential to both the kind of cognition that we have and the kind of consciousness that we have. One of the interesting aspects of Kant's work on synthesis is that he tried to unite the two kinds of synthesis he distinguished in a single theory, something that no other theorist to my knowledge has done.

The two kinds of synthesis can be viewed as operating on two different levels. Here is an example. As a result of having bad handwriting, I am all too often in the position of not being able to recognize a word I wrote earlier. If, however, I take a careful look at what I scrawled and then go and do something else for a while, I will eventually recognize what I wrote. The word 'marginalized' was a recent example. (If the brain is a neural network, that is about what one would expect; a neural network needs time to settle on a solution.) All of this happens without any apparent recourse to complex reasoning and quite outside of consciousness. However, at the end of this process of nonconscious interpretation, a second level of activity commences; I form a representation of the word, recognize it, and set out to do whatever I choose to do with it.

Much of the work of cognitive science so far has focussed on the first level, the transformation of the meaningless scrawl into a recognizable word and similar kinds of processing. Where cognitive science is Kantian, it is Kant's ideas about processing at this level that it displays: ideas about the synthesis (or binding as it is now called) of diverse sensory information into representations of single objects, ideas about the functional nature of minds able to do such synthesizing, and so on. Where cognitive science has not assimilated ideas of Kant's, on the other hand, it is generally ideas that he had about what is going on at the second level, ideas about broader and more complex processes of synthesis, about the unity of minds able to perform these more complicated kinds of synthesis, and about the consciousness involved in recognition of representations and in consciousness of self. Looked at in the light of this distinction between the two quite different levels of cognitive processing, the contrast between what cognitive science has taken over from Kant and what it has not begins to look quite interesting.

Earlier I said that cognitive science has neglected more than consciousness in Kant and that is true. However, it *has* neglected what he had to say about consciousness. Indeed, the unity found in global representations is also the feature of consciousness that most interested Kant. Global representations need not be conscious, certainly not conscious of themselves and perhaps not even conscious of objects. This is an important point and one on which Kant is widely misunderstood. It is equally important, however, that when representations are conscious, they display the same unity as global representations generally.

In fact, this unity is a feature not just of global representations and of consciousness. It is also a feature of *recognition* of representations—we recognize them as single representations. We can pull these points together in a single definition:

The unity of representation, recognition, consciousness =*df.*

- a single act of representation, recognition, consciousness, in which,
- a number of objects of representation and, often, the representing of them are combined in such a way that to represent, recognize or be aware of any of these items is also to be aware of at least some of the other items, as the object of a single representation.

As this definition makes clear, the kind of unity in question is more than just being one representation, one object of recognition, one object of consciousness. All three are not just singular but also combine a multiplicity of representational items into one representation. This latter is what their unity consists in.

Unity on the side of representations also requires unity on the side of the thing doing the representing, the mind. Though Kant dealt with this topic quite briefly, he left us at least the outline of a theory that unified global representations are the result of unifying acts of synthesis, and/or recognition and/or consciousness, and that to perform such unifying acts, the actor must be a single, unified mind. None of these claims about the unities crucial to cognition has played any significant role in contemporary cognitive science.

The ideas about synthesis, consciousness, and the mental unity underlying them that cognitive science have not taken over from Kant have a common feature. They all concern the mind as a whole and are about functions that can draw on information in a great many subsystems of the mind, functions that Fodor (1983, p. 107) calls isotropic. Another way to put the point is to say that they are all relatively holistic features of the mind. Now, some cognitive scientists have paid attention to properties of the mind as whole. Here I am thinking of the work on production systems such as Newell's (1973, 1990) Soar and Anderson's (1983) ACT, and Minsky's (1985) society of mind. There is also relevant work in metacognition theory and in philosophy, for example the work of Patricia Churchland and others (1986, 1991) on connectionist models and large scale integration of data and Dennett's (1991) multiple drafts model of consciousness. However, none of this work takes up the unities that interested Kant.

That Kant's insights into the various unities central to cognition have been neglected in cognitive science heretofore is due in part, I think, to the way recent philosophers of mind have dealt with issues closely related to mental unity.

It seems obvious, *prima facie,* that the most interesting and cognitively central unities are synchronic: the representing or recognizing or being aware of a number of things *at the same time*. Synchronic unity was the form of mental unity that most interested Kant (of the many passages that indicate this, A100, A103, A108, and A352 are especially relevant). Yet when contemporary philosophers of mind talk about mental unity at all, they almost always take up only unity *across* time—even when they are discussing Kant! Kitcher (1990) is a good example: she always interprets Kant's talk about mental unity to be about diachronic unity. Of course, diachronic unity, the representing or recognizing or being aware of earlier representations and combining them with current ones, is vital to many cognitive activities. But it is not the only or even the most important form of mental unity.

The way many philosophers have linked unity to personal identity exacerbates the problem. (When philosophers use the term 'personal identity', they mean 'being one person', usually over time. What they mean by the term is thus very different from what clinical psychologists mean.) Philosophers, including most commentators on Kant, tend to tie them closely together, taking it to be obvious that mental unity requires personal identity, that a number of representations can be unified into one global representation only if they are all the representations of a single person or mind. Since cognitive scientists have generally not been much interested in personal identity, I suspect that the philosophers' way of linking unity and identity may be have helped to turn them away from questions of unity. Whatever, the neglect is a shame; mental unity is central to our kind of mind.

Moreover, the linkage between unity and identity is looser than these philosophers hold. Again, there is both a synchronic and a diachronic question. Synchronically, the link may be close; if a number of representations are combined in one global representation, it is plausible to think that that will be enough to make them the representations of a single mind. (The possibility of the link here makes it, if anything, even more strange that philosophers have typically ignored the synchronic forms of both unity and identity.) When we turn to diachronic unity and identity, however, the link is anything but close. There seems to be no reason in principle why a mind could not combine earlier representations had by another mind with his or her current representations. All it would take is the right kind of memory access to the earlier representations. (Of course, what the right kind of access *is like* might be tricky to specify.) Moreover, and this is what makes the standard treatment of Kant on the subject so surprising, Kant was well aware of this possibility. In a famous footnote to A363, he entertains the possibility of minds so structured that "one [mind] communicates representations together with the consciousness [memory] of them" to another one, and so on in a chain. Clearly, both for Kant and in fact, mental unity across time can be and should be distinguished from personal identity.

4. Kant's Claims about Consciousness of Self

Contrary to what is often said, Kant did not consider consciousness to be essential to all forms of unified cognition, not consciousness of self at any rate. To the contrary, he spoke of

cognitive systems that are not aware of themselves at all a number of times. (Whether simple consciousness of objects is required is a more complicated matter; on one view, consciousness of objects simply *is* a synthesized, unified global representation of them.) Nevertheless, Kant was well aware that consciousness of self is at least a prominent feature of cognitive systems as we find them in people and he had some interesting things to say about it.

Indeed, Kant made both positive and negative contributions to our picture of consciousness of self. His negative contributions are contained in his attack on rationalist pretensions to infer fundamental facts about minds from concepts and an appeal to consciousness of self alone. Descartes, Leibniz and maybe Reid were the prime targets; Kant called their reasoning paralogisms and his attack on it is found in the well-known chapter called Paralogisms of Pure Reason in the first *Critique.* The key inferences are that the mind or soul is simple and that it has some form of strict and absolute persistence. Kant's attack is devastating but it had no enduring influence because the ideas under attack had no enduring influence, thanks in part to Kant's attack. So we will not consider it further.

By contrast, some of Kant's positive contributions are not well known at all. They address six topics.

1. How many kinds of self-consciousness there are. Most theorists of self-consciousness, whether they are working in the terminology of self-consciousness as most philosophers do or the terminology of metarepresentation as is more common among cognitive psychologists, treat it as all being much alike. Kant did not. He distinguished between two kinds, consciousness of one's representational states and consciousness of oneself as the subject of these states. In the latter, I am aware of myself as myself, the common subject of my representations. In the former, I am aware of particular psychological states and activities, states and activities that are in fact mine, though I may or may not be aware of that. It seems obvious that there are these two quite different kinds of consciousness of self, indeed that it should be fundamentally important to distinguish them, yet few theorists have followed Kant in doing so.

2. The cognitive and semantic machinery used to obtain consciousness of self as subject. Here Kant made a major discovery: we use a very special kind of referential apparatus to become aware of ourselves as ourselves, as the subject of our representations. Kant says that when we refer to ourselves in this way, we 'denote' but do not 'represent' ourselves (A382) or we designate ourselves 'only transcendentally', without noting in ourselves 'any quality whatsoever' (A355). What Kant is isolating here anticipates Frege's and other work on indexicals and bears a striking resemblance to Shoemaker's (1968, [67]p. 558) notion of reference to self without identification. Compare Kant's last remark to this statement of Shoemaker's:

> 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 [1968, p. 558].

That is to say, I am aware of myself, as myself, without inferring this from any other feature of myself. If so, that the referent is myself is something I know independently of knowing anything else. If so, I must be able to refer to myself as myself independently of 'noting any quality' in myself, just as Kant said. Let us call this *nonascriptive reference to self.* Shoemaker attributes the core of the idea to Wittgenstein but it goes all the way back to Kant.

3. The representational base of consciousness of self as subject. Kant had an explanation for the peculiarities of this form of reference to self. He never laid this theory out fully in any one place, so it is easy to miss it, but it is there. The fundamental idea is this. To become aware of a representation of X, usually we do not need any representation other than the representation of X (we also need some general cognitive skills). In the same way, virtually any representation can make us aware of ourselves as its subject. It is this universality that opens the way to reference to self without ascription. (The details are complicated. I discuss them in Brook 1994, Ch. 4). No cognitive theorist has ever developed a general explanation of the peculiar semantics of reference to self that is better than Kant's. Most theorists do not even try.

4. How we appear to ourselves when we are aware of ourselves as subject. When we are aware of ourselves as subject, Kant thought that the way we appear to ourselves has important peculiarities. As we appear to ourselves here, "nothing manifold is given" to ourselves (B135). As well as explaining the peculiarities of self-reference without identification, the theory of the universality of the representational base of consciousness of self explains this peculiar lack of content, too.

5. The unity in our consciousness of ourselves as subject. When we are aware of ourselves as subject, we are aware of ourselves as the "single common subject" of a number of representations (A350). The representation itself is equally unified. These instances of unity strongly resist explication by the resources of any existing theory of mental contents, and may be a main source of the tenacity of the problem of the homunculus.

To reveal the power of this notion of the unity of consciousness of self as subject, let us apply it to that old saw of anti-functionalism, the mind whose 'neurons' are the population of China. As an objection to functionalism, the story goes as follows: 'If what defines a mind is functional organization, then size is irrelevant. If so, then the people of China could be one mind. All that would be required is that they be hooked up to one another so as to exhibit the right functional organization. And that is absurd.' It is interesting to reflect on how Kant might have reacted to this claim.

Perhaps along the following lines. To find out whether the population of China could be a single mind, we need to determine two things. First, could the information realized in the relationships among the members of this population, some of it at least, be integrated so as to become a single object of a global representation? Second, could we imagine the spread-out entity composed of this population becoming aware (I) of such an object, (ii) of the global representation of this object, and (iii) of itself, and *as* itself, the common subject of the elements of this representation? I do not know how to answer these questions and I am not sure what

Kant's answer would have been but he does give us the right questions. His notions of the unity of experience, the unity of consciousness, and the consciousness of oneself as the common subject of one's representations clarify the issue of what the Chinese population *would have to be like* to be a mind.

6. When I am aware of myself as subject, of what am I aware? Here the question is: Am I aware of myself? That is to say, does nonascriptive reference to self give me epistemic access to the thing that I am, or is consciousness of self, so-called, just consciousness of another representation, in this case a representation of myself? Such a representation would presumably be as concept-laden and doctored by the mind as all other representations are. All cognitive theorists in Kant's time and most since have simply taken it for granted that nonascriptive reference to self gives us consciousness of ourselves, not just of another representation. Even among contemporary theorists, the Churchlands are among a rare few who think otherwise. Kant is the only pre-twentieth century theorist that I know of who rejected the idea; he would have been firmly on the side of the Churchlands. We "know even ourselves only through inner sense, and therefore as appearance ..." (A278; see B153-154). However, there is a twist to Kant's rejection. He certainly thought that we have no *knowledge* of ourselves as we are but he may have thought that we do have a 'bare consciousness' of ourselves as we are, a "consciousness of self [which is] very far from being a knowledge of the self" (B158). This twist is enough by itself to make his version of the 'no direct consciousness of self' thesis more subtle than, for example, the Churchlands' version.

As has been indicated, not one of Kant's six ideas about consciousness of self has been taken up by cognitive science. Even when parallel ideas have appeared in recent work, as with Shoemaker's reference to self without identification and the Churchlands' denial that we have direct, unmediated consciousness of self, the authors of the idea do not seem to know that Kant beat them to it over two hundred years ago. As a general theorist of the mind, Kant is no mere cultural artefact.

Final Comments

In this chapter we have explored ideas of Kant's that have been incorporated into contemporary cognitive science and ideas that have not. The latter concern his claims about synthesis, about the unity of representation, consciousness and mind, and about the peculiarities of consciousness of self as subject.

The topic that we have discussed do not exhaust Kant's ideas about cognition. In particular, he had a complex, sophisticated model of representation in space and time. I have not discussed it because, unlike his claims about synthesis, unity, consciousness, etc., his views on cognition in space and time have been almost universally rejected (Kitcher, 1990, and Falkenstein, 1996, are excellent treatments; for my thoughts on Falkenstein, see Brook 1998). There are also some important questions about Kant. One of the more intriguing is the one raised by Dascal at the end of his chapter in this volume: Why did Kant care so little about

language? It wasn't as though he wasn't exposed to sophisticated theorists about language, Herder for example. Since there is not much to be said by way of an answer to this question, I have not taken it up.

ABSTRACT: In this paper I argue that while the dominant model of the mind in cognitive science is deeply Kantian, some of Kant's most arresting ideas have not been assimilated into the contemporary picture. The Kantian elements in the contemporary picture are mainly these three: Representation requires both percepts and concepts, the study of cognition is based on inference to the best explanation (Kant called it transcendental argument), and the mind is a complex system of functions – minds are (part of) what brains do. Three other important ideas of Kant's have played little role: That unified consciousness is essential to our kind of cognition (beginning to change), that such unity is the result of concept-using synthesizing activities of the cognitive system, and that the knowledge that such a system has of itself has some highly specific and unusual features.

KEYWORDS: Kant - cognitive science - mind - unified consciousness - synthesis - self-knowledge

References

Anderson, J. *The Architecture of Cognition*. Cambridge, MA: Harvard Univ. Press, 1983.

Baars, B. *A Cognitive Theory of Consciousness.* Cambridge: Cambridge University Press, 1988.

Brook, A. *Kant and the Mind.* New York: Cambridge University Press, 1994.

———— 'Critical Notice of L. Falkenstein, Kant's Intuitionism: A Commentary on the Transcendental Aesthetic', Canadian Journal of Philosophy (1998).

————'Kant's View of the Mind and Consciousness of Self'. Stanford Electronic Encyclopedia of Philosophy (2004) http://plato.stanford.edu/entries/kant-mind/

Churchland, P.S. *Neurophilosophy*. Cambridge, MA: MIT Press/Bradford Books, 1986.

Churchland, P.S. and Sejnowski, T. J. *The Computational Brain.* Cambridge, MA: MIT Press/ Bradford Books, 1991.

Dennett, D. *Brainstorms*. Montgomery, VT: Bradford Books, 1978.

————*Consciousness Explained*. Boston: Little, Brown and Co., 1991.

Falkenstein, L. *Kant's Intuitionism: A Commentary on the Transcendental Aesthetic*. Toronto: University of Toronto Press, 1995.

Flanagan, O. J. *The Science of the Mind*. Cambridge, MA: MIT Press/Bradford Books, 1984.

Fodor, J. *Modularity of Mind*. Cambridge, MA: MIT Press/Bradford Books, 1983.

James, W. *Pragmatism.* Cambridge, MA: Harvard University Press, 1907.

Kant, I. (1781/1787). *Critique of Pure Reason*, trans. N. Kemp Smith in 1927 as *Immanuel Kant's Critique of Pure Reason*. London: Macmillan Co. Ltd., 1963.

——— (1783). *Prolegomena to Any Future Metaphysics that will be able to come Forward as a Science.* Trans. P. Carus, rev. with intro. by L. W. Beck. Indianapolis: Library of the Liberal Arts, 1970.

——— (1786). *The Metaphysical Foundations of Natural Science*. Trans. with introd. by James Ellington. Indianapolis: Library of Liberal Arts, 1970.

——— (1798) *Anthropology from a Pragmatic Point of View.* Trans. Mary Gregor. The Hague: Martinus Nijhoff, 1974.

Kitcher, P. Kant's real self: In Allen Wood (ed.). *Self and Nature in Kant's Philosophy*. Ithaca/NY/ London: Cornell Univ. Press, 1984, p. 111-145.

———— *Kant's Transcendental Psychology*. New York: Oxford University Press, 1990.

Martindale, C. 'Can we construct Kantian mental machines?'. The Journal of Mind and Behaviour 8 (1987), p. 261-68.

Meerbote, R. Kant's functionalism: In J. C. Smith (ed.). *Historical Foundations of Cognitive Science*. Dordrecht: Reidel, 1989.

Minsky, M. *The Society of Mind.* New York: Simon and Schuster, 1985.

Newell, A. Production systems: models of control structures: In W. G. Chase (ed.). *Visual Information Processing.* New York: Academic Press, 1973.

———— *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press, 1990.

Powell, C. T. *Kant's Theory of Self-Consciousness*. Oxford: Oxford University Press, 1990.

Sellars, W. *Science and Metaphysics*. New York: Humanities Press, 1968.

————'... this I or he or it (the thing) which thinks ....'. Proceedings of the American Philosophical Association 44 (1970)

Shoemaker, S. 'Self-reference and self-awareness'. Journal of Philosophy 65 no. 20 (1968), p. 555-567

Strawson, P. F. *The Bounds of Sense.* London: Methuen, 1966.

Treisman, A. and Gelade, G. 'A feature-integration theory of attention'. Cognitive Psychology 12 (1980), p. 97-136.

Notes

1 First published in Andrew Brook (ed.). *The Prehistory of Cognitive Science*. Basingstoke, UK: Palgrave Macmillan, 2007, pp. 117-36.

2 Andrew Brook is Chancellor's Professor of Philosophy and Cognitive Science at Carleton University in Ottawa (Canada). He is the founder and former Director of the Institute of Cognitive Science there and former President of the Canadian Philosophical Association. He is the current President of the Canadian Psychoanalytic Society. He has published about 120 books and papers.

3. Except for references to the *Critique of Pure Reason,* references to Kant's work in the text follow the practice of using the volume and page number of the twenty-nine volume German edition begun in 1902 by the *Preussischen Academie der Wissenschaften* and still not completed. References to the *Critique of Pure Reason* is in the pagination of the first two editions, usually called the 'A' and the 'B' editions. (These were the only two editions that Kant prepared himself.) Translations are from Norman Kemp Smith's 1927 translation, *Immanuel Kant's Critique of Pure Reason.* If a reference is to only one edition, the passage does not appear in the other one.

4. Flanagan's choice of this name for transcendental argumentation is curious, the intention to honour Kant notwithstanding. Kant himself used the term 'transcendental deduction' for something quite different, namely the kind of analysis used to *deduce* that use of certain concepts is necessary for representations to come to have objects. Kant used transcendental arguments in the course of this deduction, of course, but they are still different things. Nevertheless, Flanagan is quite right to pick Kant out as the originator of the method of transcendental argumentation. Given that Kant urged that empirical psychology (= introspective psychology) is impossible and never pursued the question of whether an empirical science of anthropology in his sense is possible, it might seem paradoxical that the main method behind his nonempirical, purely *a priori* work on cognition should become the methodological basis for current empirical theory-building about cognition but that is what happened.

5. The notions of a global representation and global object and the rest of the ideas discussed in this chapter are considered further in Brook (1994).

6. Among commentators on Kant, Strawson (1966) came closest to seeing that Kant spotted something similar to what Shoemaker later labelled 'reference to self without identification'. Strawson's name for the phenomenon, 'criterionless self-ascription' (p. 165), obscures more than it reveals, however. What is in question is not ascription but reference.

7. I thank Rob Stainton for helpful comments on a draft of this chapter.

Recebido / Received: 16.5. 2014
Aprovado / Approved: 28.8. 2014