Kant and Cognitive Science

Andrew Brook
Interdisciplinary Studies
Carleton University
Ottawa, ON., Canada K1S 5B6
abrook@ccs.carleton.ca

Abstract

Some of Kant's ideas about the mind have had a huge influence on cognitive science, in particular his view that sensory input has to be worked up using concepts or concept-like states and his conception of the mind as a system of cognitive functions. We explore these influences in the first part of the paper. Other ideas of Kant's about the mind have not been assimilated into cognitive science, including important ideas about processes of synthesis, mental unity, and consciousness and self-consciousness. They are the topic of the second part of the paper.

Sometimes directly, sometimes indirectly, Kant's views on the mind have had a huge influence on cognitive science. Indeed, Kant is virtually the intellectual god-father of contemporary cognitive science, so deep-running is his influence. 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 cognitive functions that lies behind the view. Both notions arose with Kant, Indeed, both are central to his model of knowledge and mind, and they entered current epistemology and cognitive science by a process of direct descent from him. In the first part of the paper, we will explore these influences.

If the Kantian cast of much 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 played no role in most work in cognitive science, though what might turn out to be related ideas are beginning to appear in some quarters. What cognitive science did not take over from Kant's work on cognition and knowledge will be the topic of the second part of the paper. I am also of the view, of course, that cognitive science *should* have taken over these ideas, but will not be able to argue this point in any detail here.

What Cognitive Science Has Taken Over From Kant

Of the ideas that cognitive scientists and the philosophers associated with them have taken over 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 so well known that it is not necessary to say anything more about it. As Kant put it, "Concepts without intuitions are empty, intuitions without concepts are blind". In contemporary jargon, discriminations need information, but we must be able to discriminate patterns of various kinds in information before it is of any use to us.

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 toe-hold on the 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).

We might note, parenthetically, that Flanagan's choice of this name for transcendental argumentation is a bit curious, the intention to honour Kant notwithstanding, because 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.

Third, even Kant's general conception of the mind and what we can and cannot capture in our models of it have 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, my saying this may seem a bit strange, so let us examine the matter. 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, recent eliminativist pretenders 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, 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 to the representational model of the mind, indeed to a rather radical one if my reading of him is correct, 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);[1] for this, the mind must synthesize its various objects of representation into what I will call the *global object* of a *global representation*. (iii) Synthesis into either individual or global objects requires the application of concepts. -- These are the central elements of the model. What all three of these tenets describe are either functions or conditions such as unified awareness that are required for some or all functions to operate. Kant even called them functions (A68=B93, A94 and elsewhere). In general, like functionalism, Kant's approach to the mind centred on how it works, as opposed, for example, to how such a system might be built (even when such structure is abstractly not physically characterized). 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 (and, many theorists think, the environment and even the subject's history); that explanation of mental functioning is a special sort of explanation (focusing 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 non-psychological explanation); that this autonomy stems from such explanation being holistic in certain ways; and perhaps others. Kant paid no attention to such specifics. Nor did he consider how to characterize functional states in terms of their relation to other functions, context, the person's behaviour, how the functions fit in the history of the person who has them, etc. If Kant's theory is a precursor of contemporary functionalism, then, his functionalism was of a rather general sort and in that respect rather different from the contemporary versions. Nevertheless, I think it is fair to view it as a precursor of functionalism.

The thought that Kant was a functionalist *avant le mot* is no longer new. Sellars (1970) was perhaps the first to read Kant as a functionalist or proto-functionalist; 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, whatever these forms might be like. 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 (or we can never know that we know anything) 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 could 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 *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 *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 relation of concepts and percepts;

-- his main method, the method of transcendental argument; and,

-- his general picture of the mind as a system of concept-using functions for manipulating representations.

Indeed, some workers in cognitive science have even begun to explore the implications of more specific aspects of Kant's model of the mind for their work. (Martindale's 1987 work on Kantian mental machines is one example.)

Let us turn now to ideas of Kant's that have played a smaller role in cognitive science so far.

---

[1] References consisting of a number preceded by 'A' or 'B' are to Kant (1781/1787), and are given in the text. 'A' refers to the pagination of the first (A) edition of 1781 and 'B' to the second (B) edition of 1787, the only editions that Kant prepared himself.

What Cognitive Science Has Not Taken Over From Kant

Before we begin our investigation, 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 made with consciousness. This obtains, most would agree, whether it is consciousness of 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. As yet, cognitive science has certainly not paid much attention as yet to consciousness of objects or consciousness of self, certainly not to the features of consciousness that most interested Kant, but it has paid equally little attention to the aspects of the mind's synthesizing powers and its unity that most interested Kant.

In both cases, some recent work is beginning to take up issues more like the ones of central interest to Kant. With respect to consciousness, one thinks of the work of Baars (1988) and two of the people Dennett (1991, xi) numbers among his 'princes of consciousness', Marcel and Bisiach (1988). With respect to synthesis and unity, there are signs of change in, for example, Fodor, 1983, and recent work on the subclass of metacognitive processes that have as their field of application information-processing, problem-solving and behaviour control functions that involve the whole mind or some substantial portion of it. (The contrast is with metacognitive processes that apply just to a single sub-system, memory for example.)[2] All this recent work notwithstanding, I think that there are still things to be learned from Kant.

With respect to synthesis, Kant claimed that the mind must synthesize multitudinous representations into a single, integrated representation of a world. With respect to unity, he urged that to perform this crucial synthesis, the mind must have a certain kind of unity. Before I attempt to define what is behind these notions more formally, it might help to get a clearer picture of what has and what has not been assimilated from Kant's work if we start with 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 down earlier. If, however, I take a careful look at what I scrawled down 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; the 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 non-conscious 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 scientists so far has focused on the first level, the transformation of the meaningless scrawl into a recognizable word and processing of that kind. Where cognitive science is Kantian, it is Kant's ideas about processing at this level that shape it: 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 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 awareness of self. Looked at in the light of this distinction between the two quite difference levels of activity involved in cognition, the contrast between what cognitive science has taken over from Kant and what it has not does not look so strange.

The ideas about synthesis, mental unity, consciousness, and self-consciousness that cognitive science has not taken over from Kant have a common feature. They concern the mind as a whole or some substantial part of it and are about functions that can draw on information in a great many sub-systems of the mind, that are isotropic in Fodor's (1983, p. 107) sense. Another way to put the point is to say that they are all relatively holistic features of the mind. But that does not seem to have been what has led to their neglect by cognitive scientists. Indeed, many cognitivists themselves pay attention to properties of the mind as whole. Here I am thinking, for example, of the work on production systems such as Newell's (1973, 1990) Soar and Anderson's (1983) ACT*, or Minsky's (1985) society of mind, as well as the work on consciousness and metacognition mentioned earlier. In the last few years, philosophers have joined in, first in the form of the work of Patricia Churchland and others (1986, 1991) on how connectionist models might be able to account for large-scale integration of data, and more recently in Dennett's (1991) multiple drafts model of awareness. The feature that seems to have led to the relative neglect of these ideas of Kant's is that they all concern activities of forming multiple representations into a what we might call a global representation.

We can define a global representation as follows:

> A global representation $=df.$ a representation that has a number of particular representations and/or their objects or contents as its single global object.

We then define 'single global object':

> A single global object $=df.$ an intentional object that represents a number of intentional objects and/or the representations that represent them, such that to be aware of any of these objects and/or their representations is also to be aware of other objects and/or representations that make it up and of the collection of them as a

[2] The work on the kinds of metacognition I have in mind is so diverse and fraught with unclarities of definition and domain that it would not be easy to cite it adequately in a few references.

single group.[3]

As a very simple example, each of us right now is aware of the words I am uttering, thoughts suggested by those words, thoughts I raised earlier in my talk, the heat of the room, wishes and concerns about the remainder of the day, and so on, not individually but, first, all at the same time and, second, as the complex object of a single consciousness.

Kant thought that the capacity to form global representations is absolutely essential to the kind of cognition we have, and he thought that both the two kinds of synthesis and a certain kind of unity are required to form them. Synthesis first.

As Kant saw it, synthesis comes in two flavours. First we must tie various kind of sensuous information together into single objects of unified representations. Then we must tie these individual representations together into global representations. The first kind of synthesis has reappeared in contemporary research as the notion of binding (in the psychological, not the linguistic sense) and been the object of considerable attention; but the second kind has hardly received any attention at all. Here is an example. Colours, lines, shapes, textures, etc., are represented in widely dispersed areas of the brain. The process of tying these dispersed representations together into a representation of an object is what is now called binding. Indeed, Treisman and her co-workers (1980) have developed a version of binding theory that even parallels Kant's in important ways. They hold that three stages of visual processing are involved. First the content of feature modules are applied, the result is located on a map of locations, and then the result of both processes is recognized via a recognition network and object files. These three processes parallel Kant's three stages of apprehension, reproduction, and recognition in concepts very closely.

But binding is only one of the two forms of activity to which Kant gave the name synthesis. The second kind of synthesis is the activity of tying multiple representations together into a global representation. Some of the work on production systems and on metacognition touches on this phenomenon; so, less directly, does Baars' (1988) work on consciousness and some of the work of the semantic holism theorists such as Davidson and Dennett. Here, I think, we may still be able to learn from Kant; these contributions touch only part of the problem.

A global representation displays a certain kind of unity. It is *one* representation, and it is one by virtue of connecting a number of other representations to one another in a certain way. This form of unity is found very widely in the mind. In addition to being a feature of global representations and therefore of representational experience, it is also a feature of *recognition* of representations -- we recognize them in *single* acts of recognition -- and of consciousness -- we are aware of a number of things in *single* acts of consciousness. This kind of unity also seems to be *distinctive* to processes of combining, recognizing, and being conscious of representations. We could define it as follows:

> The unity of representation, recognition, consciousness =*df.* (i) a single act of representation, recognition, consciousness, in which (ii) a number of representations and/or objects of representation are combined in such a way that to represent, recognize or be aware of any of these representations is also to be aware of at least some of the other representations combined with it, and 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, act of recognition or act of consciousness. All three are not just singular, but also unified: one act combining a multiplicity of items into one object.

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. Even though no other theorist has ever done this, to my knowledge, I do not have the space to consider how Kant tried to do so here. Likewise, I cannot consider the extent to which the work on metacognition and consciousness that I have mentioned has regained some of the ground that Kant won two hundred years ago with his notions of the global representation and its various unities. There is a point on which I would like to say a few words, however. If Kant's insights into the various unities central to cognition have been relatively neglected in cognitive science heretofore, part of the blame, it seems to me, lies with philosophers of mind. Here is why.

About the unities of cognition, we can ask two questions: Are these unities synchronic, diachronic, or both?, and, What is their relationship to what philosophers call personal identity? (What philosophers call personal identity is being one person, usually over time. Clinical psychologists use the term for something quite different.) The question about synchronic and diachronic unity first.

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 representations *at the same time*. It is equally clear that this was the form of mental unity of greatest interest to Kant (of the many passages that indicate this, see especially A100, A103, A108, and A352). 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: in her work, she always interprets Kant's talk about mental unity to be about diachronic unity exclusively. 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, too. My point is simply that it is not the only form of mental unity.

Now the relationship of unity to identity, being one person. Many philosophers, including virtually all commentators on Kant, have taken it as obvious that mental unity requires mental identity, that a number of representations can be unified into one global representation only if they are all the representations of a single mind. Since cognitive scientists have generally not been much interested

[3] The notions of a global representation and global object are considered further in Brook (1994), as indeed are many of the ideas discussed in this paper.

in what it is to be one mind, I suspect that the way philosophers have linked unity and identity may be have helped to turn them away from questions of mental unity, too. If so, that is a shame, because mental unity is central to cognitive systems of our kind, and in any case, the linkage between unity and identity is far looser than many philosophers think.

Again, there is both a synchronic and a diachronic question here. Synchronically, the linkage may be closer; if a number of representations are combined in one global representation, it is plausible to think that they will always be the representations of a single mind. (If the claimed linkage is found here, it is strange that philosophers have typically ignored the synchronic forms of both phenomenon.) When we turn to diachronic unity and identity, however, the linkage is anything but close. There seems to be no reason in principle why a mind could not combine earlier representations had by someone else with his or her current representations, so long as he or she had the right kind of memory access to those earlier representations. (I will not go into the question of what the right kind of access might be like. It is complicated.) 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 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, it is possible to have mental unity across time without personal identity.

Conclusion

A great deal more could be said about Kant's views on the mind and what they might still have to offer to contemporary cognitive researchers. In particular, we have said nothing about his extremely interesting views on consciousness of self. Contrary to the received view, Kant did not consider consciousness of self to be essential to cognition. Quite the contrary, he entertained the idea of cognitive system not aware of themselves a number of times. Nevertheless, he certainly thought that self-consciousness is a prominent feature of the cognitive systems we are, and he had some interesting things to say about it. For example, he distinguished between awareness of one's psychological states and awareness of oneself as the subject of those states, something not often done, and he had some deep insights into the latter. He urged, for example, that in this form of awareness of self, one refers to the object (oneself) "without noting in it any quality whatsoever" (A355), a notion that seems to anticipate Shoemaker's (1968) self-reference without identification. Unfortunately, I have no room to look at these ideas and will have to leave them for another occasion. I hope that what I have said about Kant on synthesis, unity, and consciousness has been enough to suggest that we may still have things to learn from him.

References

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

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

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

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

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

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

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown and Co.

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

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

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).

Kitcher, P. (1984). Kant's real self. In Allen Wood, ed., *Self and Nature in Kant's Philosophy* (pp. 111-45). Ithaca, NY and London: Cornell Univ. Press.

Kitcher, P. (1990). *Kant's Transcendental Psychology*. New York: Oxford University Press.

Marcel, A. and Bisiach, E., eds. (1988). *Consciousness in Contemporary Science*. New York: Oxford University Press.

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

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

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

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

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

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

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

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

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

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