# Kant and Cognitive Science

Andrew Brook
Professor of Philosophy and Cognitive Science
Carleton University
Ottawa, Ontario, Canada

Indirectly through 19 th century intermediaries, Immanuel Kant (1720-1804) has had an enormous influence on cognitive research, indeed could be viewed as the intellectual godfather of cognitive science. In general structure, Kant's model of the mind was the dominant model in the 19th century empirical psychology that flowed from his work (Herbart, Helmholtz, Wundt) and then again, after a hiatus during which behaviourism reigned supreme (roughly 1910 to 1965), toward the end of the 20th century, especially in cognitive science. The models of the mind of thinkers otherwise as different as Sigmund Freud and Jerry Fodor are both broadly Kantian, for example.

Kant held that cognition requires application of concepts as well as sensory input, and that cognition proceeds by way of synthesis, indeed by way of three kinds of synthesis. He advanced a functionalist model of the mind almost 200 years before functionalism was officially articulated. And he developed a distinctive form of inference to the best explanation to study the mind. All these things are now orthodoxy in cognitive science. But he made discoveries that have not be taken up by cognitive science, too. He maintained that consciousness and specifically unified consciousness is central to cognition and he had some highly original things to say about consciousness of self. In sum, though Kant has had a huge influence on cognitive science, not only has his work not been superseded by cognitive science, some of his most distinctive contributions have not even been assimilated by it.

## Life and Influence: A Sketch

Kant was the last great thinker of the German Enlightenment (which among other things means that he focussed on the human individual rather than state, society or culture)

and one of the most influential philosophers of all time. He made important contributions to a diverse range of subfields within philosophy and also did important work outside, as we will see. We will focus in this article on his work on the mind.

Kant's life was as unremarkable as his work was epoch-making. He lived his whole life in Königsberg (now Kaliningrad), just below Lithuania. His father was a saddle maker. He was deeply engaged with religion but hostile to both the official Lutheranism of his time and the stricter pietistic offshoot in which he was raised. Though he started his academic career late (he tutored privately for many years), he became an important figure in the University of Königsberg and late in life served terms at different times as rector of the University, though he hated the work. (One story about the extent of his hostility to official religion is well-known: as Rector, he had to lead the church parade of students and instructors to the main church in Königsberg each Sunday. He refused to enter the church. Instead, he would step out on the church steps and go home.) By the time of his death, he was virtually the official philosopher of the German-speaking world. Though viewed as a quintessentially German philosopher, he is said to have been one-quarter Scottish. Some hold that 'Kant' is a Germanization of the Scottish name 'Candt' (Scottish philosophers are particularly fond of this idea).

Kant's most important work on the mind is found in the *Critique of Pure Reason (CPR)* of 1781/7 (two editions). He was already 57 when he wrote the first edition, yet he went on to write two further *Critiques,* the *Critique of Practical Reasoning* (1788) on moral reasoning, 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. For our purposes, *CPR* and a small book worked up from lecture notes late in life, *Anthropology from a Pragmatic Point of View* (1798) are his most important works.

After Kant and drawing on his work, German philosophy went off in two directions. One we have already seen: empirical (broadly, scientific) work on the mind. The other was in the direction of what came to be known as German romanticism (Hegel, Schelling, Schopenhauer). The romantics were interested in units much larger than individual minds (all that exists through all of history in the case of Hegel, a foundation of all that exists, Will, in the case of Schopenhauer). Hegel and Schelling and their followers were interested in particular in broad

human groupings: nations, cultures, and so on.

## Main philosophical views

Until middle age, Kant was a conventional rationalist. Then memories of reading David Hume "interrupted my dogmatic slumbers", as he put it *(Prolegomena to Any Future Metaphysics,* Ak IV:260). He called the new approach Critical Philosophy. Philosophy of mind was by no means the only area in which Kant made seminal contributions. He founded physical geometry. His work on social ethics grounds modern liberal democratic theory. His deontological approach to the justification of ethical beliefs put ethics on a new footing, one that remains influential to this day. In the lecture hall, he taught metaphysics, ethics, physical geometry, logic, mechanics, theoretical physics, algebra, calculus, trigonometry, and history, an almost unimaginable range of topics for anyone now.

Kant's aim in the part of the critical philosophy of interest to us, *CPR,* was to do two principal things:

> Justify our conviction that physics, like mathematics, is a body of necessary and universal truth.
>
> Insulate religion, including belief in immortality, and the free will necessary for morality from the corrosive effects of this very same science.

Kant had not the slightest doubt that 'God, freedom and immortality' *(CPR,* Bxxx) exist but feared that, if science were relevant to the question of their existence at all, it would provide reasons to doubt their existence. Fortunately, as he saw it, science is quite irrelevant to these questions.

In course of his attempt to put knowledge in general and physics in particular on a secure foundation, Kant asked the following question: What must the mind be like for our knowledge to be as it is? 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. But this also tells us a lot about what our minds must be like. The model of the mind that resulted became the framework model adopted by most subsequent cognitive researchers.

## Model of the mind

Interestingly, Kant held that psychology (by which he meant the introspective study of the mind) could never be a science. After saying that chemistry would never be science, he once 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" (*Metaphysical Foundations of Natural Science,* Ak. IV:471). So how should we study the mind? By thinking though what the mind *must* be like and what capacities it *must* have to represent things as it does. This is his famous *transcendental method.*

> 1. To understand the mind, examine the necessary conditions of its experience being as it is.

The arguments that result from this examination are called transcendental arguments. Translated into contemporary terms, this method is simply inference to the best explanation: postulate unobservable mental mechanisms to explain observed behaviour. This was the approach adopted when introspection fell out of favour about 100 years ago. Its nonempirical roots in Kant notwithstanding, is now by far the most important method used by cognitive scientists.

Kant's application of this method 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" *(CPR,* A51=B75). In more contemporary terms,

> 2. The functions crucial for mental, 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. Next,

> 3. The functions that organize sensory and conceptual raw materials into experiences, are forms of the ability to synthesize.

Kant postulated three kinds of synthesis. 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 requires two kinds of Synthesis of Recognition in Concepts. The first ties the raw material of sensible experience together into objects. To put his point in terms of contemporary binding theory, 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 and integrated into a representation of a single object.

The second kind of synthesis ties these individual representations together into what might be called *global representations*. A global representation connects individual representations to one another in such a way that to be aware of any of the representations thus tied together is to be aware of some of the others, 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. As we will see, the two kinds of synthesis in a concept have had very different fates in contemporary cognitive science.

In Kant's model, cognitive functions play the central role. This is a plausible approach to the mind, as we know, but Kant had a particular reason for adopting it. One of his most deeply held general convictions was that we know nothing of anything as it is. All we know of things is as they appear to us. Since this was a completely general view, he had to say the same about the mind. But he seems to allow that we do know things about the mind – that it must apply concepts, synthesize, and so on. He never addressed the problem squarely but a natural way out for him would have been to distinguish the mind's functions from its structure and composition and then maintain that what we lack knowledge of are the latter things. This would be to adopt the functionalist view that, in the mind, function does not dictate form – a given function could be implemented by systems having very different forms. (This idea is at the heart of contemporary functionalism.) Kant's view that we know *nothing* of the structure and composition of the mind would then merely be a radical version of this idea. At any rate, in his model, cognitive functions are central.

4. The mind is complex set of abilities (functions).

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

have seen, and their belief that we can model cognitive function without knowing anything very much about underlying structure.

The four ideas just laid out are the core of Kant's model. They are all at the foundation of contemporary cognitive science (not all parts of his doctrine of synthesis, however, the point to which we have to return).

## Consciousness and Self-Consciousness

### *Unified Consciousness*

Now we move to some ideas of Kant's that have had less influence in contemporary cognitive science. The first is the unity of consciousness. Though by no means all global representations are conscious (Kant is widely misunderstood on this point), Kant thought we must be conscious of what some of them represent and to be thus conscious, our consciousness must have a distinctive kind of unity. In his view, briefly, to have global representations, we must synthesize (construct) them using unifying acts of synthesis (Kant's name for which was *transcendental apperception),* and it takes a unified consciousness to perform unified acts of synthesis.

Indeed, two kinds of unity are required:

1. The experiences must have a single common subject (A350);

and,

2. The consciousness that this subject has of represented objects and/or representations must be unified.

Kant said little about what a 'single common subject' is like and we will follow him in this. Likewise, he never said what he meant by 'unified consciousness' but he did use the notion often enough that one can see what he must have had in mind. One plausible articulation of the notion at work in his writings would go as follows,

> The unity of consciousness=*df*. (i) a single act of consciousness, which (ii) makes one conscious of a number of representations and/or objects of representation in such a way that to be conscious of any of this group is also to be conscious of others in the group

and of at least some of them as a group.

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

### *Consciousness of Self*

In addition to what he said about consciousness in general, Kant made some remarkably original discoveries about consciousness of self. These claims arose in the course of pursing his second objective, to insulate immortality (and God and free will) from the corrosive threat of science. His rationalist predecessors thought that they could prove that the mind is substantial, simple (without parts), and persists in a special way. This opened the door to a proof of immortality. Descartes, Leibniz and Reid all took this approach. However, if arguments can support immortality, then science can also undermine it. (Similarly for free will.)

For Kant, the best hope was to insulate all these matters from both argument and evidence. That way, the conclusions could be accepted on the basis of faith (and Kant did so accept them) without being at risk from science. Kant thought that introspection provides strong *prima facie* counter-evidence to his anti-intellectual conclusions about what we can know about the nature of the mind. In introspection, one does appear to be substantial, simple and persisting, just as rational psychologists says ('rational psychology' was Kant's name for these views). If so, it was incumbent upon him to show that introspection gives us nothing of the sort. In the course of this deflationary attack on introspection, Kant made a number of claims:

1. He distinguished two quite different kinds of self-awareness, awareness of one's states and awareness of oneself as the subject of these states.

2. He urged that the cognitive and semantic machinery used to obtain awareness of self as subject is quite unusual. In it, we 'denote' but do not 'represent' ourselves *(CPR*, A382). Put otherwise, we designate ourselves without noting 'any quality whatsoever' in ourselves *(CPR,* A355).

3. He argued that the representational base of awareness of self as subject is not a special experience of self but any experience of anything whatsoever.

4. When one is aware of oneself as subject, he claimed, one is aware of oneself in a way that is not awareness of features of oneself, a way in which "nothing manifold is given."

*(CPR,* B135).

The special form that consciousness of self takes allowed Kant to advance two other theses of great importance to him:

> 5. In inner sense, one is conscious even of oneself only as one appears to oneself, not as one is.

As he put it,

> inner sense... represents to consciousness ever 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)?

Since to insist that even when we are aware of ourselves as subject of our experience, we are aware only of an appearance of the self, not the self itself would have been implausible, Kant then put consciousness of self as subject in a special category:

> 6. When one is conscious of oneself as subject, the bare consciousness of self that we have, a consciousness that has no manifold (see 4. above), yields no knowledge of self.

Here we have closely paraphrased Kant's own words. We have a "bare ... consciousness of self [that is] very far from being a knowledge of the self" (B158).

### Where Kant Has and Has Not Influenced Cognitive Science

As we said, Kant's influence on cognitive science was via 19th century cognitive researchers, Herbart, Helmholtz, and Wundt in particular. Contemporary researchers are right to take these figures as the precursors of contemporary cognitive research but seldom realize that they all viewed themselves as Kantians. As we have seen, some of Kant's doctrines are built into the very foundations of cognitive science. Others have been virtually ignored.

These ideas became the foundation of the model of the mind of contemporary cognitive science:

> 1. The transcendental method, the method of postulating unobservable mental mechanisms in order to infer to the best explanation of observed behaviour.
>
> 2. The claim that most representation requires concepts as well as percepts.
>
> 3. Part at least of the doctrine of synthesis (we will return to this topic immediately

below).

4. The functionalist conception of the mind and the doctrine that cognitive function does not dictate underlying form.

As we saw, Kant went further than contemporary functionalists in this regard: he held that we can know *nothing* about underlying structure.

Some ideas equally central to Kant have not played much role in cognitive science. Recall Kant's claim that experience requires two kinds of synthesis of recognition in a concept. In the form of binding, the phenomenon that he had in mind in the first kind of synthesis of recognition, recognition of individual objects using concepts, is widely studied. Indeed, one model, Anne Treisman's (1980) three-stage model, is similar to Kant's whole threefold doctrine of synthesis. According to Treisman, object recognition proceeds in three stages: feature detection, location of features on a map of locations, and integration and identification of objects under concepts. These compare closely to Kant's three-stage model of apprehension of features, association of features in something like clusters (Kant called this stage 'reproduction'), and recognition of these 'clusters' as objects falling under concepts *(CPR*, A98-A106). So far, we see that, once again, Kant's ideas have become the current orthodoxy.

But what of the second form of synthesis of recognition in a concept, namely, the tying of individual represented objects together in a global representation *(CPR,* A107-14)? Here we find an entirely different story. Cognitive science has paid very little attention to this kind of synthesis, part of its general neglect of properties of the mind as a whole.

The same was true until recently of Kant's doctrine of the unity of consciousness. However, this is changing. In cognitive science from its inception in the 1960s 'til about half way through the 1980s, consciousness in general and the unity of consciousness in particular received very little attention. That has now changed. In past twenty years, there have been hundreds of books and thousands of papers on both topics.

Finally, Kant's views on consciousness of self have played little role in cognitive science. Contrary to what is often said, Kant did not consider consciousness of self to be essential to all forms of unified cognition but, as we saw, he did make a number of penetrating discoveries about it. Some closely related ideas have reappeared in philosophy in the past few decades (Shoemaker 1968, 1970; Perry 1979) but none of this work has had much impact in cognitive science.

In short, the dominant model of the mind in contemporary cognitive science is Kantian, but some of his most distinctive contributions have not been taken up by it.

The topics that we have discussed do not exhaust Kant's ideas about cognition. He also had a complex model of representation in space and time. He held that the spatial and the temporal matrices are properties of the mind, not the world, properties that we impose on the raw materials of our experience. Coordinate with these views, he also made some statements about the raw materials of experience themselves. As we have seen, he had strong views on what we can and cannot know about the mind. None of these views have had much impact on contemporary cognitive science. Unlike other views of his that have not had much impact such as his views on consciousness of self, it is hard to see any enduring value in them, either, and many philosophers reject them. In this short paper, we have stuck to the views of his that do have enduring value. As we saw, many of them are now part of the current orthodoxy.

## References

*Relevant Books by Kant*

(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. (1781/1787) *Critique of Pure Reason* (trans. P. Guyer and A. Woods). Cambridge and New York: Cambridge University Press, 1997. (Quoted passages are in my own translations.)

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

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

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

*Some Important Works about Kant on the Mind*

Ameriks, K. (1983) *Kant's Theory of Mind*. Oxford: Oxford University Press.

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

Guyer, P. (1987) *Kant and the Claims of Knowledge*. Cambridge and New York: Cambridge University Press

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

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

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

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

*Other Works Referred to*

Brook, A. and DeVidi, R., eds. (2001) *Self-Reference and Self-Awareness*. Amsterdam: John Benjamins

Brook. A. and Raymont. P. (2004) *A Unified Theory of Consciousness*.

Perry, J. (1979) The essential indexical. In Brook and DeVidi 2001

Shoemaker, S. (1968) Self-reference and self-awareness. In Brook and DeVidi 2001

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

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