Theism and Cognitive Science

Mark C. Baker

Rutgers University

1. Introduction

Cognitive Science is an interdisciplinary approach to the study of the mind that originated in the mid-1950s and draws on insights and techniques from psychology, linguistics, philosophy, computer science, and related disciplines. It can be defined more specifically as the study of the representations that are used by the mind, and the computational processes that operate over those representations. For example, Lepore and Pylyshyn (1999:vii) write: “The approach in cognitive science is … essentially computational; the capacity for intelligence is viewed in this discipline as arising from the processing of representations” (see also Wilson and Keil 1999:xxviii, Fodor 2000:4). In essence, it is dedicated to pursuing the view that the mind is (at least in part) a kind of computer.

Defined in this way, there is no intrinsic logical connection between Cognitive Science and Theism, the idea that there exists an all-powerful divine being that created the universe and is involved in how it unfolds. Most or all of what cognitive scientists have discovered could in principle be true (or false) regardless of whether Theism is true or false. Nor is there any prominent subdiscipline defined by the intersection of these two disciplines—no distinct field of “Theistic cognitive science.”

That being said, there are some important interactions between the two, both at the foundations of Cognitive Science and at its edges. Part of the interest in cognitive science for some researchers is that it provides a way of “naturalizing” our understanding of the mind. It gives a potential way of explaining what it is for something to be rational or intelligent without attributing to it any kind of immaterial soul. Then if it turns out to be unnecessary to posit souls for rational creatures such as human beings in order to account for their rationality, it may also become unnecessary to posit a God who creates and sustains these souls. Thus one kind of consideration that has sometimes been used to support Theism may be undermined by the successful pursuit of Cognitive Science.

However, the relationship between Cognitive Science and Theism is arguably more complex and multidimensional than this. For example, one consequence of Cognitive Scientific thinking in the minds of its founders was a revival of interest in Cartesian rationalism—the idea that minds have innate structure and content. Now for Descartes himself, there was a conceptual link between rationalism and dualism—and from there to his theism. One might wonder, then, whether it is possible to accept rationalism without some dualism and theism as well. Moreover, Cognitive Science seems to have hit some walls when it comes to providing a complete account of the human mind. The door thus remains open that some crucial role is to be played by some kind of immaterial soul—and hence perhaps by an immaterial deity that gives and sustains that soul, and from which the soul inherits its rationality. If so, Cognitive Science research actually makes this argument for theism stronger, although within a narrower domain.

The remainder of this article explores these themes in more detail.

2. A Taste of Cognitive Scientific Thinking

Let us begin by taking a closer look at what cognitive science is. When we define it as "the study of the representations that are used by the mind, and the computational processes that operate over those representations," what does this amount to? What are these representations and why do we think that they play an important role in the human mind?

We can get a sense of this by way of some example from the study of language that have been discussed by leading linguist Noam Chomsky, one of the founders of cognitive science. Consider what competent native speakers of English know about the following sentences (Chomsky, 1982: 30):

(1) The men are too stubborn to talk to Bill.

(2) The men are too stubborn to talk to.

There is more to these normal-sounding English sentences than first meets the eye. Superficially, the only difference between them is that (1) ends with *Bill* and (2) does not. But there is also a significant difference in interpretation concerning who does the talking, the men or someone else. In sentence (1), it is clearly the men who do the talking; it can be roughly paraphrased as (3), where the subject of *talk to* is a pronoun that refers back to the men.

(3) The men are so stubborn that they (the men) will not talk to Bill.

Since (2) is superficially identical to (1) up to the last word, one might expect that, by analogy, its interpretation should also be identical to that of (1) up to the last word. In other words, (2) should mean something like (4):

(4) #The men are so stubborn that they (the men) will not talk to.

But (2) cannot mean this. First, we want to know what is the object of the talking action: who is being addressed? Without this, the sentence is logically incomplete, like the deviant sentence *The men talked to*. So (4) needs to be completed with an indication of the one(s) being talked to. There are two logical possibilities: it could be the men under discussion, or it could be someone else is. We could thus complete the interpretation of (2) as either (5) or (6).

(5) The men are so stubborn that they (the men) will not talk to someone else.

(6) The men are so stubborn that they (the men) will not talk to them (the men).

But neither of these is a good paraphrase of (2) either. There is something correct about (6): (2) is about hypothetical situations in which the relevant men are being talked to. But in this case it is not the men who are doing the talking; rather the situation envisioned is one in which other people are talking to the men. What (2) really means is something like (7).

(7) The men are so stubborn that people in general will not talk to them (the men).

What we learn here is that the way that the subject of *to talk to* is interpreted in these sentences depends on whether there is an object like *Bill* in the sentence or not. It seems unlikely that this can be explained simply by superficial analogy (following a slogan like "interpret similar sentences similarly") or by mere pragmatics (following a slogan like "interpret a sentence in the most plausible/useful way"). Rather Chomsky concludes that sentences like (1) and (2) have representations in the minds of English speakers in which some kind of unpronounced pronouns are filled in as the subjects and objects of 'to talk to', and these pronouns are marked as to what they can and cannot refer to. In other words, (1) is mentally represented as (8), and (2) as (9).

(8) The men_1 are too stubborn $\textit{pronoun}_1$ to talk to Bill_2.

(9) The men_1 are too stubborn $\textit{pronoun}_3$ to talk to $\textit{pronoun}_1$.

This example gives a taste of what cognitive scientists mean by mental representations, and what kind of explanatory work they might do. Speakers know unambiguously what they mean by a sentence, and hearers are often able to deduce what speakers mean. But (1) and (2) are not entirely explicit as to what the sentences mean. Hence speakers and hearers must have another representation of these sentences in their minds, one that does make explicit and unambiguous what the sentences mean, including who the talker is and who is being talked to.

Therefore there exist mental representations which are systematically related to things in the physical world (in this case, tokens of English sentences) but not identical to them.

In fact, speakers and hearers must control at least *two* distinct representations of these sentences: one that expresses what the sentences mean—something like (8) and (9)—and one that expresses how the sentences are pronounced—something more like the original (1) and (2). These two representations are somewhat different, because the extra pronoun-like elements are present in the meaning but not in the pronunciation. Cognitive scientists draw two conclusions from this. First, they infer that the mind contains *many* different kinds of representations, with different ones used for different purposes. Representations are thus a big deal when it comes to understanding the mind. Second, they infer that there must be processes of computation that take one kind of representation, suitable for one sort of purpose, and turn it into another kind of representation, suitable for another purpose. For example, part of speaking English is taking a representation like (8) and transforming it into a representation like (1) by systematically applying "rules" like the following:

> "Delete a pronoun when it is the subject of an infinitival verb if it is understood as referring to the subject of the larger sentence."

Conversely, hearers take a representation of what they hear like (1) and enrich it by adding in pronouns in the right places. More generally, if there are different kinds of representations in the mind, then it also stands to reason that there are computations defined over these representations, including computations that transform one into another. This then illustrates more fully the definition of cognitive science.

The connection with computer science now becomes evident. The various representations of what an English sentence means and how it should be said are just the kinds of "data structures" that a good programmer would build into a computer system that was designed to do natural language applications. Indeed, computer scientists have gained much experience and sophistication in understanding different ways of representing data, what each is good for, and how to transform one type into another efficiently. It has also become clear that any sort of robot or computer system that is designed to interact with the world effectively will have in its databanks some kinds of representations of the state of the world, and will be able to compute how those states are likely to change depending on what the robot does. So computer scientists have practical experience with the kinds of representational and computational issues that arise as psychologists and linguists seek to understand what people and other animals do mentally, making dialog across these disciplines fruitful.

While I have illustrated using examples from the study of language, similar reasons to believe in representations and computations are found in many other areas. Another relatively well-developed area of cognitive science is the study of vision. Here too representation and computation must be essential, as shown for example by ground-breaking work by David Marr (1982) and others. We know what kind of visual input we receive—two two-dimensional patterns of light striking our retinas—and we know what we perceive—a single three-dimensional world of objects arrayed in space. Since these are quite different, there must be (at least) two distinct representations in use, with the three-dimensional world calculated from the two-dimensional patterns of color in systematic ways. Indeed, we can figure out quite a bit about the details of our mental representations of visual space by a variety of techniques, such as showing people carefully constructed artificial stimuli and seeing which ones trick them into perceiving a particular object in a particular way. Other areas of cognition that have been open to investigation in these terms include: people's mental representations of number and quantity, of spatial relations and mental maps, of human faces, of kinship relations and other social networks, and of the belief states of other people around them—among others (see Pinker 1997 and Wilson and Keil 1999 for overviews). In addition, a large subbranch of the field has studied the similar topics in infants and children, to see how the representations and computations of the mind change as the person grows and learns. This then is the domain of cognitive science.

3. The computational theory of mind and naturalizing intelligence

Cognitive science has been successful, then, across a fairly broad range of topics by pursuing its twin notions of representation and computation, which seem to be ubiquitous in the mind when studied in this way. It therefore becomes natural and tempting to conjecture that this will add up to a complete theory of the mind. One might imagine, then, that all the mind is is a large repository of mental representations of different kinds, together with ways of computing one from another. This conjecture is known as the computational theory of the mind: the idea that the mind can be fully understood as a complex computational device (Fodor, 1975, Fodor, 1981, Dennett, 1981: ch.7, Dennett, 1991); see Pinker 2002:31-41 for an informal overview).

In proposing the computational theory of mind, Cognitive scientists certainly do not mean that the human mind works the same way that a digital computer does at the physical level. That would clearly be false: the computations of the mind presumably run on cells called neurons bathed in a chemical soup, whereas those of the computers we know run on tiny *nor*-gates engraved on silicon chips. But part of the core idea of cognitive science is that these significant physical differences might not matter as much as you would think. Alan Turing (1937) formulated a precise mathematical definition of what computation is, and it is completely independent of the physical details of the "machine" that performs the computation. For example, we can imagine many different adding machines that will receive the inputs 7 and 5 and produce the output 12. Any such machine is performing the same computation, although its method of doing so could be quite different. Given this, it is better to express the computational theory of mind not as “the mind is a computer” (invoking the images we have of computers as physical objects) but rather as “the mind is a computational device”—meaning that it performs computations in the mathematical sense defined by Turing.

This brings us to the first serious point of contact between cognitive science and theism and associated religious notions. To the extent that the mind can be understood as something that performs computations of a specific sort, operating over a finite set of well-defined symbols, to that extent the mind is demystified and rendered understandable.

This would be a huge step, because apart from Cognitive Science our experience of having a rational mind is remarkably mysterious (cf. Pinker 1997:4, 560; Pinker 2002: 31). We all think, and we realize that this can be a conscious activity that requires effort. But what exactly we are doing when we are thinking is almost completely opaque to us. I stare for a while at a chess position or a linguistics problem or my financial statements, and suddenly I get an idea of what to do. I might test that idea and either accept it or reject it; if I reject it, I might try to get another idea. But many of my best, most rational ideas seem to pop into my head from nowhere, and I do not know what steps I went through to get them. Sometimes in more complex tasks I have a sense of going through a process, with conscious steps, but even that is mysterious, because each individual step feels like a discontinuous leap from one idea to the next. And combined with our lack of awareness of what is actually involved in thinking is the fact that most things in the world do not think in this sense: rocks do not, trees do not, worms probably do not, and even dogs do so only in a sense that seems very limited compared to ours. Putting this together, rational thought seems to be something quite miraculous, which goes beyond the bounds of the familiar physical world. The temptation is thus quite strong to see it as something spiritual, as a divine spark in us—perhaps as something given to us by God, when we were made in his image. So it has been very plausible throughout history to see our power of conscious, rational, voluntary thought in theistic terms.

What cognitive science seems to offer, then, is a credible alternative to this theistic view of the mind. It reveals that we have certain representations in our minds, and that we perform systematic calculations using those representations. And so do computers. We know in great detail how to endow machines with similar representations, and how to build them so that they perform similar calculations in a systematic way. Indeed, we can program them so that they do well various things that we have traditionally taken to be signs of distinctively human intelligence, at least over limited domains: they can play good chess, prove mathematical theorems, diagnose diseases, and pick investment strategies. Nevertheless, there is nothing mysterious or divine about computers or what they are doing when they do these things. On the contrary, we know that they are purely material objects, "mindlessly" obeying the laws of physics, because we built them that way. Thinking along these lines, it becomes quite imaginable that humans could be purely material objects too, and still be (or appear to be) rational and intelligent. As a result, Pinker 2002:31 declares that "new ideas from four frontiers of knowledge—the sciences of mind, brain, genes, and evolution" are "exorcising the ghost [soul] in the machine [body]."

Cognitive science even gives a plausible way of thinking about why thinking seems mysterious to us when we are doing it. After all, the behavior of computers also seems mysterious to the casual observer looking at them from the outside. I punch a number into a calculator and press the square root sign. Instantly a long string of numbers appears, mysteriously, apparently out of nowhere. In fact, the machine went quickly through many perfectly simple and unmysterious intermediate calculations along the way; I just didn't see them. The machine only reported to me its final answer, not all the steps it went through to arrive at it. There is every reason to think that the same is true of the human mind. We are evidently always doing all sorts of normal and simple computations that we are not consciously aware of. For example, in the English examples discussed above, we unconsciously delete pronouns in certain places from sentences we are thinking of uttering when they refer to other phrases in that sentence. The computation is simple enough to specify, and we can infer that we must be doing it, given how what we mean matches up with what we say. But we have no awareness of performing the computation. Just like the calculator, my mind only reports to me (the conscious part of my mind) certain key steps, such as the final answer, not every intermediate step. (And a good thing too, since tracking all the intermediate stages in the calculation would surely be very tedious.) Hence the mind, like the calculator, *looks* mysterious, but only until one figures out how to get inside it, according to the cognitive scientist.

In the final analysis, we can describe this in two ways: we can say that well-programmed computers really *are* intelligent, or that they only give the illusion of being intelligent because when we look at the details we see that they are not using anything worth calling *intelligence* after all. But whichever choice we make, it is not implausible to say the same thing about us: either our intelligence is a purely physical phenomenon, or we only have the illusion of intelligence ourselves—an illusion that will dissipate once we know in more detail exactly what we are doing. It then becomes clearer than ever before that there may be no need to attribute an immaterial soul to human beings. And if not, then there is no analogy to be drawn from the spirit of a person that controls his or her body to a great Spirit that controls the universe. Nor is there need to posit a nonmaterial being that creates and sustains such souls in us. Cognitive Science thus gives us that much less reason to be theists.

4. The Neutrality of Cognitive Science

However, there are also other, less obvious connections between cognitive science and theism that can be explored. Some of these are not antagonistic at all, but rather neutral, or even positive and synergistic.

First, as mentioned above, the conceptual power of Turing's definition of computation, which lies at the heart of the computational theory of mind, is that it is entirely neutral about the nature of the "machine" that performs the computations. As a result, the success of the computational theory of mind might open up the serious possibility that the mind is a purely physical machine, consisting only of physical particles obeying familiar laws of physics, but it in no way entails that the mind must be physical. Most of the successes of cognitive science to date have been at the level of figuring out what *functions* a human being (or animal) must be computing, and perhaps what *algorithms* (general computing strategies) they are using to compute those functions. They do not reach down anywhere close to what computer scientists would call the "machine code", the grittiest level of programming, where decisions are made about exactly how a high-level computational function will be implemented using the specific hardware available in this particular machine.

The whole point of looking at a topic in terms of functions, algorithms, and computations is to be able to say something substantive while abstracting away from these details. Since it is at that level that most of cognitive science operates, cognitive science is, by definition, neutral as to whether the "machine" that performs the calculations is made up of neurons, or nor-gates, or ectoplasm, or fairy dust, or some combination—or indeed out of nothing at all. Any conclusions about which of these hypotheses is true or the most likely does not come strictly speaking from cognitive science, but from cognitive science combined with other physical or metaphysical assumptions—such as a judgment about which hypothesis is simpler overall (Ockham's Razor), or a background bias in favor of materialism. And of course all these considerations were already in play before the advent of Cognitive Science. Cognitive Science itself is officially neutral about these matters.

Of course, this might change in the foreseeable future. The brain sciences are an area of intense current research. Those who study brains are trying to work their way upward toward the mind, while cognitive scientists who study the mind are trying to work their way down to the brain, both trying to crack the "machine code" in which our hardware computes identifiable functions and algorithms. In recognition of these efforts, we have since about 1980 the name of a new field, cognitive neuroscience, which aspires to bridge the gap between strict brain science and traditional higher-level cognitive science (Albright and Neville, 1999).

But so far it is debatable whether this emerging field has realized its aspirations, and the problems that have come to light seem more striking than the solutions. I believe that so far it is not known how a single piece of abstract propositional knowledge known by a human being is actually represented in the physical structures of the brain. Indeed, it is not even known how abstract knowledge is represented in the much simpler brain of the ant, as discussed forcefully by C.R. Gallistel (Gallistel and King, 2009). It can be shown beyond reasonable doubt that an ant foraging for food, even though it wanders in a complicated path, always keeps track of the exact distance and direction back to its hive. As a result, whenever its mission is completed it can immediately and accurately go straight home by the shortest path. If an experimenter picks the ant up and moves it to a new spot, it still heads straight to where its home ought to be (not thrown off by the change of terrain), and it acts confused when it arrives there to find no hive. This behavior shows that the mind of the ant stores two numerical values, which it continually updates: the distance to home, and the compass bearing to home. However, existing theories of neural structure cannot account for this. These theories are built around "realistic" idealizations of what we see happening in brains, and that is complex patterns of electrical activity, which are continuous in nature and always in flux. Gallistel and King show that theories that try to be realistic to the brain science in this way are exactly the wrong sorts of theories for representing numerical values — like the distance to home. They conclude (p. 287) that "If the computational theory of mind, the core assumption of cognitive science, is correct, then brains must possess a mechanism for carrying information forward in time in a computationally accessible format. … [But] as of this writing, neuroscience knows of no plausible mechanism for [doing this]." Until problems like this have been solved, the theoretical neutrality of cognitive science to questions of how computation is realized in bodies and/or souls remains a serious matter.

5. Cognitive science, innateness, and theism

Another possible point of connection between cognitive science and theism comes from the association between cognitive science and rationalism (see, for example, Chomsky 1965:ch.1, Chomsky 1966). Here is how the matter arises.

Cognitive scientific research shows that people's minds include many sophisticated representations of language, meaning, space, and relationships, and that they perform complex computations with those representations, most of which we are unaware of. The question then arises, how did we acquire all this? How did we learn what representations to use, and what computations to go through to arrive at them? It is all well and good to say that the mind works like a computer, but what plays the role of the programmer in the analogy? Who set up for us our representational systems and defined our computational algorithms? It is hard to imagine children figuring all this out for themselves from scratch, given how slow they are on the uptake in many other ways.

The bite of these questions might become clearer if we reflect a bit more on our mental representations of *The men are too stubborn to talk to*, as considered in section 2. How do we come to know that the ones understood as talking in this example are not the same as the men whose stubbornness is under discussion, as in the superficially parallel *The men are too stubborn to talk*? In fact, English speakers do not have to learn this directly; it can be inferred from other, more general facts about the language, including the following:

A. A clause must be logically complete.

B. A null object can only be understood as the same as the higher subject

C. The pronoun subject of a clause cannot refer to something else in the same clause.

Each of these premises holds for English sentences that do not involve the word *stubborn*. For example, premise A explains why we cannot say *The men talked to*; there must be some kind of noun phrase after *talked to*. Given this, the English speakers know to include an unpronounced noun phrase ("NP") in their representation of the end of the sentence:

(10) The men are too stubborn NP to talk to NP.

Premise B now implies that the object NP must be refer to the subject of the sentence as a whole, further enriching the representation to give:

(11) The men_k are too stubborn NP to talk to NP_k.

Like hypothesis A, hypothesis B is valid for other sentences, such as (15), which means approximately the same thing as *It is easy for me to talk to the men*.

(12) The men_k are easy for me to talk to (NP_k)

Finally, we come to the question of how the first unpronounced noun phrase in (11) is interpreted. In the abstract there might be a preference for it refer to the men, that being the only immediately available antecedent for the pronoun. But premise C blocks that representation in this particular case. Premise C also explains simpler facts, such as the fact that the pronoun *she* in the following examples cannot refer to the same person as *Mary* or *her* (Lasnik, 1989).

(13) She talks to Mary/her. (Can't mean: She/Mary talks to herself)

Premise C also implies that the complete representation of (11) cannot be *The* men_k *are too stubborn* NP_k *to talk to* NP_k but must be *The* men_k *are too stubborn* NP_n *to talk to* NP_k with the talkers being someone other than the men.

This kind of reasoning takes us part of the way toward understanding how people arrive at the relatively complex representations and computations that they do—but only part of the way. The question arises now as to how English speakers learn that A-C are true for English. Certainly nobody taught my children A-C explicitly, and yet they know these quite specific and abstract principles, because they understand this range of sentences in the same way that I do.

Considerations like this lead Chomsky and other cognitive scientists to rationalist conclusions: they infer that principles like A-C are innate aspects of the human mind. People do not learn them per se, but they emerge spontaneously in the normally developing human mind. In other words, a relatively large amount of structure and information is preprogrammed into the mind, providing it with the computational infrastructure it needs to develop the relatively complex representations that it does in language, vision, and other areas of cognition. The traditional picture of a child's mind being a blank slate that experience writes on does not seem to get one very far if one want to account in any detail for the mental representations and computations that children come to use. Gallistel and King (2009:252) come to the same conclusion, considering quite different cognitive domains (navigation, foraging), remarking that there is an "ineluctable nativism inherent in *any* representational machinery." Indeed, our everyday experience of personal computers also points in this direction, in that we know what computers with blank hard drives are good for: virtually nothing. It is computers that are preloaded with sophisticated operating systems and an interesting variety of applications that can receive meaningful data do meaningful things with it (cf. Pinker 2002:34-35). There is no reason to think that the human mind is different from a PC in this respect.

The radical-sounding idea that the human mind comes equipped with significant innate structure for dealing with natural language data, including principles like A and C, makes a bold prediction that seems to be true. This is that all human languages, despite their superficial differences, will follow these same principles. In fact, it has been shown that languages as widely different as English, Mohawk, Thai, Swahili, Turkish, Mapudungun and many others respect these principles (when stated correctly), despite their very different histories, locations, and cultural contexts (Baker, 1996). In that respect, then, the rationalism that cognitive science depends on can be confirmed empirically.

But this line of analysis pushes the hard question one step back. We do not have to explain how children learn the right representations and computations from experience; we say that they do not learn them, but come with the relevant infrastructure already in place—not only the hardware, but also the firmware, operating system, and many of the "apps". But now we want to understand where all this preprogrammed computational infrastructure came from.

To Descartes and his fellow 17th century rationalists the answer would perhaps have been obvious. Descartes also believed in an immaterial soul, created by God. The innate structure which does not come from experience but is a precondition for our knowledge and experience could be part of that soul, which God makes in his image. This forms a coherent picture, with theism undergirding dualism, dualism undergirding rationalism, and rationalism undergirding workaday Cognitive Science. In short, one could say that God plays the role of software engineer that is arguably presupposed by the computational theory of the mind.

This seems to be as coherent an answer as any that we now have available. Two other research programs exist. One is to find ways of doing cognitive science that further reduce the preconditions for having a functioning mind, so that less innate structure is needed. That would involve deriving things like premises A-C from still more general principles, so less needs to be preprogrammed into the mind. That route might be attractive in theory, but it has been hard to do in detail. For example, there is nothing logically necessary about premise C such that it follows from the fundamental dynamics of communication. Artificial languages, for example, usually have no equivalent of C and get along fine without it.

The second alternative research program is that of Evolutionary Psychology, which has grown in influence within Cognitive Science in the last two decades. Evolutionary Psychology embraces innateness, but replaces God in the rationalist picture with evolution (Pinker, 1997, Carruthers, 1992). Evolutionary psychologists say that children do not have to learn the computational infrastructure for dealing with sentences like *The men are too stubborn to talk to* over their individual lifespans, but the human species developed it over evolutionary time, as we evolved. On this view, the innate structure of the human mind is part of our biological heritage, specified somehow by our DNA, which in turn has been shaped by natural selection.

This may be the best nontheistic game in town, but like cognitive neuroscience, the questions it faces look at least as impressive as its answers. Recall that we do not know how abstract propositional knowledge is represented in neurons. A fortiori, then, we do not know how it can be represented in the DNA that guides the construction of neurons. Also potentially puzzling is the discovery that the human genome contains relatively few genes. Hence we might wonder whether we have enough of them to specify (for example) premise C, along with the rest of our innate mental structure, in addition to our entire physical blueprint. For aspects of intelligence that are distinctively human, such as language, this issue is compounded by the fact that there is remarkably little genetic difference between a human being and a chimpanzee. Are there enough novel genes, then, and enough evolutionary time, to account for the impressive mental differences between a human mind and a chimpanzee mind in this way? Finally, we must ask how plausible it is that specific features of human cognition, such as premise C (or something that implies it), were actually favored by natural selection. Would organisms that had C as part of their mental architecture have had an advantage in surviving and reproducing compared to similar creatures that did not? It is hard to see how. It is easy enough to say that language as a whole has proved useful as a means of communication in ways that enhance fitness, but it is challenging to apply this conceptual framework to the more specific features of human language. One can suppose that C plays a role in making sentences less ambiguous, and that aids communication. But it also makes language less flexible, and that could be a hindrance to communication. Moreover, human languages tolerate massive amounts of ambiguity in other respects, so if there is evolutionary pressure against ambiguity, why hasn't it been reduced more than it has? Given unanswered questions like this, it is an open question whether the rationalism that a fully developed cognitive science of the human mind seems to depend on can in fact be fully naturalized within an atheistic evolutionary view.

6. Theism and the Incompleteness of Cognitive Science

A final consideration is whether the computational theory of mind really will amount to a complete theory of the mind. Only if it does can cognitive science claim to have shown that there is no need to invoke influences outside of normal physical forces in the area of the mental. If on the contrary cognitive science ends up providing explanations for some aspects of the human mind but not others, it might actually enhance the argument that something of potential religious significance is going on in those aspects of the mind that it does not properly apply to, since a plausible alternative will have been eliminated.

In fact, cognitive science cannot claim to have finished the job. Generally speaking, it works best for relatively narrow input-output functions, where the mind interfaces with the rest of the world, such as perception, language, and motor planning. For matters arguably at the center of the mind, such as rational thought and planning, its successes are much more limited. Moreover, insider appraisals of how likely it is to reach completion vary widely. For example, Steven Pinker comes close to proclaiming a total victory in his 1997 book *How the Mind Works.* Nevertheless, even Pinker's work ends with a list of some important unsolved and perhaps unsolvable mysteries, drawing on work by Chomsky (1975) and philosopher Colin McGinn (1993). And Jerry Fodor strenuously disagrees with Pinker's overall optimism in his 2001 response, *The Mind Doesn't Work that Way: The scope and limits of computational psychology*.

Significantly, Fodor's central concerns do not depend on controversial experimental results, which might be corrected or overturned. Instead they go to the very heart of what computation is, according to Turing's definition. Fodor reminds us that, powerful and general as the notion of computation is, it has certain intrinsic limitations built into it. He presses the point that the computational theory of mind works well for *deductive* forms of reasoning, but it can have no account for *abductive* forms of reasoning—forms of thought that involve inference to the best overall explanation, when there is no way of knowing in advance what facts are relevant. According to Turing, the computations that a computer can perform on an input must depend only on the local, formal, syntactic properties of that input, not on what the input is taken to mean. *By definition*, the most sophisticated computers simply transform strings of 1s and 0s into new strings of 1s and 0s in systematic ways, without any consideration of what those 1s and 0s refer to. Given this, it should be no surprise that well-programmed computers can reason deductively, telling us what conclusions follow because of the *form* of the premises, but they cannot reason abductively, telling us what conclusions follow because of the *meaning* of the premises. But it is this second kind of reasoning that is more prominent and important in human cognition. Usually the conclusions that one can draw based on deductive logic alone are quite limited and modest—things like if all men are mortal and Socrates is man, then Socrates is mortal. The really interesting conclusions come from integrating, interpreting, and applying real world knowledge over a broad domain, and this computers cannot do reliably in a finite amount of time. Fodor concludes that the theory of computation gives us an excellent account of one aspect of rationality, but it cannot give us an account of another aspect of rationality, almost by definition. Fodor (2001:99-101) thus identifies the question of how abductive thought is possible as one of the great mysteries (along with consciousness) that hovers over contemporary cognitive science, without any prospects of a solution on the horizon. He concludes that "What our cognitive science has done so far is mostly to throw some light on how much dark there is." (See also Heil 1981 for an even more general critique.)

Conclusion

Following Fodor, then, I conclude that cognitive science has shed considerable light on some aspects of the human mind, but it only highlights what is still mysterious in other parts of the human mind. Moreover, the parts that remain unexplained are precisely those parts that have traditionally been taken to be the most religiously significant by theists, such as our reason and our free will. Cognitive science thus does not foreclose on the possibility that we have those abilities primarily because we share something significant in common with a rational and free God who in some fashion creates and sustains us.

References

Albright, T. and Neville, H. (1999) In *The MIT Encyclopedia of the Cognitive Sciences*(Eds, Wilson, R. and Keil, F.) MIT Press, Cambridge, Mass., pp. li-lxxii.

Baker, M. (1996) *The polysynthesis parameter,* Oxford University Press, New York.

Carruthers, P. (1992) *Human knowledge and human nature,* Oxford University Press, Oxford.

Chomsky, N. (1965) *Aspects of the theory of syntax,* MIT Press, Cambridge, Mass.

Chomsky, N. (1966) *Cartesian linguistics,* Harper and Row, New York.

Chomsky, N. (1975) *Reflections on language,* Pantheon, New York.

Chomsky, N. (1982) *Some concepts and consequences of the theory of government and binding,* MIT Press, Cambridge, Mass.

Dennett, D. (1981) *Brainstorms,* MIT Press, Cambridge, Mass.

Dennett, D. (1991) *Consciousness explained,* Little-Brown, Boston.

Fodor, J. (1975) *The Language of Thought,* Crowell, New York.

Fodor, J. (1981) *Representations,* MIT Press, Cambridge, Mass.

Fodor, J. (2000) *The mind doesn't work that way,* MIT Press, Cambridge, Mass.

Gallistel, C. R. and King, A. P. (2009) *Memory and the computational brain: why cognitive science will transform neuroscience,* Wiley-Blackwell, Malden, Mass.

Heil, J. (1981) *Mind,* **90,** 321-342.

Lasnik, H. (1989) *Essays on anaphora,* Kluwer, Dordrecht.

Lepore, E. and Pylyshyn, Z. (Eds.) (1999) *What is Cognitive Science?,* Blackwell, Oxford.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information,* W.H. Freeman, San Francisco.

McGinn, C. (1993) *Problems in philosophy: the limits of inquiry,* Blackwell, Cambridge.

Pinker, S. (1997) *How the mind works,* W.W. Norton, New York.

Pinker, S. (2002) *The blank slate: the modern denial of human nature,* Viking, New York.

Turing, A. (1937) *Proceedings of the London Mathematical Society,* **42,** 230-265.

Wilson, R. and Keil, F. (1999) *The MIT Encyclopedia of the Cognitive Sciences,* MIT Press, Cambridge, Mass.

Word count: 6697 words, plus references