Applying Global Workspace Theory to the Frame Problem

Murray Shanahan

Dept. Electrical & Electronic
Engineering,
Imperial College,
Exhibition Road,
London SW7 2BT,
England.
m.shanahan@imperial.ac.uk

Bernard Baars

The Neurosciences Institute,
10640 John Jay Hopkins Drive,
San Diego,
California 92121,
U.S.A.
baars@nsi.edu

June 2004

DRAFT

Abstract

The subject of this article is the frame problem, as conceived by certain cognitive scientists and philosophers of mind. The challenge is to explain the capacity of so-called informationally unencapsulated cognitive processes to deal effectively with information from potentially any cognitive domain without the burden of having to explicitly sift the relevant from the irrelevant. The paper advocates a global workspace architecture, with its ability to manage massively parallel resources in the context of a serial thread of computation, as an answer to this challenge. Analogical reasoning is given particular attention, since it exemplifies informational unencapsulation in its most extreme form. Because global workspace theory also purports to account for the distinction between conscious and unconscious information processing, the paper advances the tentative conclusion that consciousness goes hand-in-hand with a solution to the frame problem in the biological brain.

1. Introduction

The frame problem was originally couched as a difficulty within classical Artificial Intelligence: How can we build a program capable of inferring the effects of an action without reasoning explicitly about all its obvious non-effects? But many philosophers saw the frame problem as symptomatic of a wider difficulty, namely how to account for cognitive processes capable of drawing on information from arbitrary domains of knowledge or expertise. So-called "informationally unencapsulated" processes of this sort, exemplified by analogical reasoning, are especially troublesome for theories of mind that rely on some sort of modular organisation to render them computationally feasible.

However, one thing is clear. If the frame problem is a genuine puzzle, the human brain incorporates a solution to it. In global workspace theory, we find clues to how this solution might work. Global workspace theory posits a functional role for consciousness, which is to facilitate information exchange among multiple, special-purpose, unconscious brain processes (Baars, 1988). These compete for access to a global workspace, which allows selected information to be broadcast back to the whole system. Such an architecture accommodates high-speed, domain-specific processes (or "modules") while facilitating just the sort of crossing of domain boundaries required to address the philosophers' frame problem.

The paper is organised as follows. In Sections 2 and 3, the philosophers' conception of the frame problem is presented. Section 4 challenges the premise that informationally unencapsulated cognitive processes are, in principle, computationally infeasible. In Section 5, global workspace theory is outlined. The arguments in favour of the theory are reviewed, thalamo-cortical interaction is proposed as a possible mechanism for realising a global workspace in the biological brain, and the global workspace architecture is commended as a model of combined serial and parallel information flow capable of overcoming the frame problem.

Section 6 concerns analogical reasoning, the epitome of informational unencapsulation, and demonstrates that the most successful of contemporary computational models of analogical reasoning are strikingly compatible with global workspace theory. The concluding discussion addresses a variety of topics including modularity, working memory, conscious information processing, and the relationship between parallel and serial computation in a generic account of cognitive function.

2. The Frame Problem

The frame problem, in its original form, was to address the following question (McCarthy & Hayes, 1969). How is it possible to write a collection of axioms in mathematical logic that captures the effects of actions, without being obliged to include an overwhelming number of axioms that describe the trivial non-effects of those actions? In everyday discourse, we can describe the effect of, say, painting an object simply by detailing how its colour changes. There is no need to state explicitly that the object's shape remains the same, that the neighbour's cat remains asleep, that the Sun continues to shine, and so on. This is all taken for granted by common sense.

In mathematical logic, by contrast, nothing is taken for granted. Not only is it necessary to make explicit the changes an action brings about, it's also necessary to make explicit the things that do *not* change, and for most actions this will be a great deal. This was particularly troubling for AI researchers in the 1970s and 1980s who wanted to write programs that worked by carrying out inference with logical formulae. The burden of having to supply a set of explicit formulae describing every non-effect of every action seemed overwhelming, both representationally and computationally.

In the intervening decades, the frame problem in this narrow guise has receded. Researchers in logic-based AI – who have become something of a minority anyway – more-or-less solved the problem in the 1990s using a combination of judicious representation techniques and non-monotonic forms of inference (Sandewall, 1994; Shanahan, 1997; Reiter, 2001).[1] Yet the frame problem lives on in the minds of cognitive scientists. This is

[1] For an overview see (Shanahan, 2003).

largely due to the wider interpretation given to it by certain philosophers of mind, especially Fodor (1983; 1987; 2000).

The credit (or blame) for first bringing the frame problem to the attention of philosophers of mind and cognitive scientists goes to Dennett, who characterised it as the challenge of understanding how "a cognitive creature … with many beliefs about the world" can update those beliefs when it performs an action so that they remain "roughly faithful to the world" (1978, p.125). In *The Modularity of Mind*, Fodor invited the reader to consider a hypothetical robot, and poses a related question: "How … does the machine's program determine which beliefs the robot ought to re-evaluate given that it has embarked upon some or other course of action?" (Fodor, 1983, p.114).

Dennett (1984) highlights the issue with a memorable example. He asks us to consider the challenge facing the designers of an imaginary robot whose task is to retrieve an object resting on a wagon in a nearby room. But the room also contains a bomb, which is timed to explode soon. The first version of the robot successfully works out that it must pull the wagon out of the room. Unfortunately, the bomb is on the wagon. And although the robot knows the bomb is on the wagon, it fails to notice that pulling the wagon out brings the bomb along too.

So the designers produce a second version of the robot. This model works out all the *consequences* of its actions before doing anything. But the new robot gets blown up too, because it spends too long in the room working out what will happen when it moves the wagon.

> It had just finished deducing that pulling the wagon out of the room would not change to color of the room's walls, and was embarking on a proof of the further implication that pulling the wagon out would cause its wheels to turn more revolutions than there were wheels on the wagon – when the bomb exploded.[2]

So the robot builders come up with a third design. This robot is programmed to tell the difference between *relevant* and *irrelevant* implications. When working out the consequences of its actions, it considers

[2] Dennett (1984), p. 129.

only the relevant ones. But to the surprise of its designers, this version of the robot fares no better. Like its predecessor, it sits in the room "thinking" rather than acting.

> "Do something!" they yelled at it. "I am," it retorted. "I'm busily ignoring some thousands of implications I have determined to be irrelevant. Just as soon as I find an irrelevant implication, I put it on the list of those I must ignore, and …" the bomb went off.[3]

In Fodor's words, this robot suffers from "Hamlet's problem: when to stop thinking", and the frame problem is "Hamlet's problem viewed from an engineer's perspective" (Fodor, 1987, p.140). Significantly, the problem of "when to stop thinking" arises not only when anticipating the consequences of an action. It is a difficulty that besets any *informationally unencapsulated* inference process, that is to say any process for which no *a priori* limit exists to the information that might be pertinent to it. Exemplary among such processes is *analogical reasoning*, "which depends precisely upon the transfer of information among cognitive domains previously assumed to be irrelevant" (Fodor, 1983, p.105).

3. The Computational Theory of Mind

The concern of this paper is the frame problem in the wide sense intended by Fodor.[4] Is the frame problem, in this sense, a real problem, and if so, how is it solved in the human brain? The first question to be addressed is the following. Granted that certain mental processes, such as analogical reasoning, are indeed informationally unencapsulated, why is this thought to be such a serious problem?

The significance of the frame problem is inherited from the Computational Theory of Mind (CTM). According to CTM, cognitive processes are computations. Specifically, they are truth-preserving

[3] Dennett (1984), p. 130.

[4] Controversy over the use of the term "frame problem" has led to some acrimonious debate between AI researchers and philosophers (Pylyshyn, 1987). The present paper takes the view that the philosophers have pinpointed an interesting and well-defined question relating to informationally unencapsulated cognitive processes, even if this issue bears only a slight resemblance to the original AI researchers' frame problem.

computations over representations that have a combinatorial syntax. The Computational Theory of Mind is stronger than the claim that the input-output function of every human brain is equivalent to some Turing machine. As Fodor puts it, CTM entails that "the mind is interestingly like a Turing machine" (Fodor, 2000, p. 30).

The frame problem is viewed as a problem for CTM because informationally unencapsulated processes are thought to be computationally infeasible.[5] Here's Fodor again.

> The totality of one's epistemic commitments is *vastly* too large a space to have to search if all one's trying to do is figure out whether, since there are clouds, it would be wise to carry an umbrella. Indeed, the totality of one's epistemic commitments is vastly too large a space to have to search *whatever* it is that one is trying to figure out.[6]

This sort of view is pervasive, and by no means confined to Fodor. Here is Carruthers on the same theme.

> The computational processes that realize human cognition will need to be tractable ones, of course; for they need to operate in real time with limited computational resources … And any processor that had to access the full set of the agent's background beliefs (or even a significant sub-set thereof) would be faced with an unmanageable combinatorial explosion.[7]

Famously, such considerations lead Fodor to propose a distinction between the mind's *peripheral processes*, which are supposedly informationally encapsulated and therefore computationally feasible, and its *central processes*, which are informationally unencapsulated and therefore computationally infeasible. The mind's peripheral processes – which do only "moderately" interesting things, like parsing and early vision – can be

[5] The term computationally "infeasible" rather than "intractable" is used here because intractability has a mathematically precise meaning in the context of computational complexity, which we will come to shortly. Philosophers writing on the frame problem are typically informal in their use of such terms.

[6] Fodor (2000), p. 31.

[7] Carruthers (2003), p. 505.

understood as a system of *modules*, and are amenable to study by cognitive scientists. On the other hand, the mind's central processes – which do all the really interesting and important things, like belief revision and analogical reasoning – because they are computationally infeasible, are altogether beyond understanding for contemporary cognitive science (Fodor, 1983; 2000).

> … it probably isn't true, at least in the general case, that cognitive processes are computations. … [so] it's a mystery, not just a problem, what model of the mind cognitive science ought to try next.[8]

Fodor's prognosis is dire. But his argument rests on several questionable assumptions. The first is that the Computational Theory of Mind is the only viable research methodology in cognitive science today, "the only game in town" as he put it.[9] This assumption has been vigorously challenged, and just as robustly defended.[10] But the target of the present paper is the assumption that informationally unencapsulated processes are computationally infeasible, a claim that can be detached from CTM. As Haselager and van Rappard (1988) point out, if we accept the need to account for the "systematicity" of thought (Fodor & Pylyshyn, 1988), then the frame problem is equally an issue for a connectionist theory of mind. The following section deals specifically with Fodor's argument, so its notion of a "cognitive process" is inherited from CTM. But the solution ultimately offered here, based on global workspace architecture, admits a wider variety of forms of computation.

4. Complexity and Informational Encapsulation

There is no doubt, of course, that some tasks *are* computationally intractable, in a sense that has been made mathematically precise (Garey & Johnson, 1979). To sharpen the discussion, it's worth reviewing the basic computer science. Consider a function *F*. Suppose it can be proved that an

[8] Fodor (2000), p. 23.

[9] Fodor (1975), p. 406.

[10] Highlights of the debate include the connectionist attack by Smolensky (1988), the response from Fodor & Pylyshyn (1988), and more recently the dynamical systems challenge (van Gelder, 1997).

algorithm exists that, for any input string x of length n, can compute $F(x)$ in less than or equal to $T(n)$ steps. So T sets an upper bound on how long the computation will take, in the general case. The rate of growth of T can succinctly be characterised by the dominant term in T, dropping other terms and coefficients. For example, if $T(n) = 5n^2 + 3n + 7$, then the growth of T is characterised by n^2, and we say that the task of computing F is $O(n^2)$, pronounced "order n^2". If a computational task is $O(n)$, then it is easy, from a computational point-of-view. If a task is $O(n^2)$, then it is not so easy. Even so, any task that is $\mathrm{O}(n^k)$ for some constant k is said to *tractable.* The class of all such problems is denoted P.

Now, there exists an important class of problems for which the best known algorithms are $O(2^n)$, even though polynomial-time algorithms exist for verifying the correctness of a given solution to the problem. This class of problems is denoted NP, and any task that falls into this class is said to be computationally *intractable*. The best known of these is the so-called SAT problem, which is the task of showing that a given expression of propositional logic (of a certain form) is satisfiable. A typical strategy for proving that a task belongs to NP is to show that it can be reduced to the SAT problem in polynomial time. Any such problem is said to be *NP-complete*.

NP-complete problems, though hard from a computational point-of-view, are not the hardest. Beyond NP, we have a class of problems for which it can be proved that no algorithm exists at all that is guaranteed to produce a solution. Such problems are said to be *undecidable*. Examples include the Halting Problem for Turing machines and theorem proving in first-order predicate calculus.

It is generally taken to be a bad thing if a computer programmer becomes entangled with an NP-complete problem, and worse still if they end up wrestling with an undecidable one. Despite this, computer scientists (especially AI researchers) routinely confront both NP-complete and undecidable problems. As far as intractability is concerned, if n is guaranteed to be small, an exponential algorithm is not so worrying. Moreover, it should be remembered that these complexity results are worst-case analyses. Individual instances of an NP-complete problem are frequently soluble in reasonable time, even for large n, especially if clever heuristics are used. Additionally, various practical techniques, including *resource bounded* computation (Russell & Wefald, 1991) and *anytime algorithms* (Zilberstein & Russell, 1993), can be used to mitigate the effects of an unfavourable complexity result in a working system.[11]

Now let's return to the supposition that informationally unencapsulated cognitive processes are computationally infeasible. Does this claim stand up to proper scrutiny? In particular, are such processes intractable in the theoretically precise sense? First, let's reconsider the fact that, according to the computational theory of mind, cognitive processes are "truth-preserving computations over representations that have a combinatorial syntax". In other words, cognitive processes prove theorems in some formal logic. Moreover, for CTM to be capable of explaining the systematicity of thought (Fodor & Pylyshyn, 1988), the logic in question must be at least first order.[12] Since it is known that theorem proving is undecidable for first-order predicate calculus, it's natural to ask whether this is the reason informationally unencapsulated cognitive processes are alleged to be infeasible. But this cannot the basis of the allegation, because the same observation applies to supposedly untroublesome informationally encapsulated cognitive processes. If these also carry out first-order logical inference, they too are undecidable, regardless of the problem size.

Therefore the real concern over computational feasibility is not an accompaniment to the presumption that first-order logical inference is taking place.[13] The real worry, in the context of the frame problem, seems to be that the set of sentences having a potential bearing on an informationally unencapsulated cognitive process is simply "too big". This suggests that the issue isn't really tractability, in the proper sense of the term, for tractability is to do with the rate at which computation time grows

[11] The inspiration behind these techniques is the concept of *bounded rationality* introduced in (Simon, 1957).

[12] According to (Fodor & Pylyshyn, 1988), systematicity is a property of relational representations. Therefore propositional calculus, though decidable, would be inadequate.

[13] Exactly why supporters of CTM are not worried by this issue is hard to fathom. Perhaps they are impressed by the armoury of techniques developed by AI researchers to contain the problem. If so, they never seem to mention the fact.

with the size of the problem. Rather, the difficulty seems to be the upper bound on n, not the upper bound on $T(n)$, where the n in question is the number of potentially relevant background sentences.

But why should a large such n be a problem? Recall Fodor's claim that "the totality of one's epistemic commitments is vastly too large a space to have to search". Perhaps the perceived difficulty with a large n is the supposed cost of carrying out an exhaustive search of n items of data. Yet plenty of algorithms exist for search problems that have very favourable computational properties. For example, the task of searching a balanced ordered binary tree for a given item of data is $O(\log_2 n)$. Other data structures for search have even better complexity results. This is why Internet search engines can find every website that mentions a given combination of keywords in a fraction of a second, even though they have access to several billion web pages.

Similarly, take Carruthers' assertion that a process that has to "access the full set of the agent's background beliefs … would be faced with an unmanageable combinatorial explosion". It should be clear by now that the truth of this claim depends entirely on the nature of the process in question. As long as this is not spelled out, the purported infeasibility of informationally encapsulated cognitive processes remains in question. If isolating relevant information from a mass of background beliefs is a matter of searching a very large data structure, then there is no combinatorial problem.

To see how it might not be merely a matter of searching a large data structure, let's introduce a distinction between *explicit* belief that is "near the surface" and easily accessible to cognition and the mass of *implicit* potential belief buried deep in the logical consequences of the set of explicit beliefs. Perhaps explicit beliefs can indeed be thought of as stored in a large data structure, and as therefore amenable to efficient search. But it seems reasonable to suppose there could be an implicit belief relevant to an ongoing problem while none of the explicit beliefs entailing it had anything about them to indicate this. How would could an informationally unencapsulated process find this implicit belief without having to visit all the logical consequences of every explicit belief?

Of course, it would be ridiculous to suppose that just because someone believed *P* they also believed all the logical consequences of *P*, and it's equally ridiculous to suppose human cognition will always successfully exploit all the relevant implications of explicit belief (Cherniak, 1986).[14] Human beings are fallible, and this is one among many reasons why. But this observation is a red herring. The frame problem is an issue even allowing for the limits of human cognition (Fodor, 2000, Ch. 2). The question is how human beings *ever* manage to select relevant information from a large mass of candidate beliefs in arbitrary domains and bring it to bear on a novel problem. The issue of how far into the consequences of explicit belief a cognitive process can feasibly penetrate is orthogonal. So it cannot be a cornerstone of the computational infeasibility thesis.

Perhaps Fodor, Carruthers, and other like-minded cognitive scientists have been led astray by Dennett's caricature of a ploddingly stupid robot trying not to get blown up. Recall that Dennett's imaginary robot explicitly considers each irrelevant implication of its actions before it decides to ignore it. The robot's supposed difficulty is that it doesn't know how to stop thinking. But the design of Dennett's robot is absurd. It carries out a serial computation that exhaustively works through a long list of alternatives one-by-one before it terminates. In the following section, a rather different computational model is brought to bear on the frame problem.

5. Global Workspace Theory

The discussion of the previous section suggests that a convincing case for the computational infeasibility of informationally unencapsulated cognitive processes has not been made. Proponents of the infeasibility thesis are insufficiently rigorous in their treatment of algorithmic complexity and are unsuccessful in demonstrating that computational problems follow from the nature of the cognitive processes in question. So it is legitimate to regard the existence of such processes as a problem rather than a mystery. Yet it is an injustice to these authors to deny the intuitive force of their challenge. The task remains to explain how an informationally unencapsulated process

[14] See (Gabaix & Laibson, 2003) for a discussion of this issue in the context of economics.

might be realised in a biological brain. Since the human brain plainly does not use data structures like ordered binary trees, how does a cognitive process like analogical reasoning manage to select among all the information available to it?

However, as it stands this question is somewhat ill-posed. It betrays the assumption that it is the responsibility of the cognitive process itself to make the selection of relevant information. According to the proposal of the present section, a better question would be the following. How is it that all and only the relevant information is *made available to* a cognitive process like analogical reasoning? The answer offered here relies on the idea of distributing the responsibility for selecting relevant information among multiple parallel processes. The proposed means for achieving this is a *global workspace architecture* (Baars, 1988).

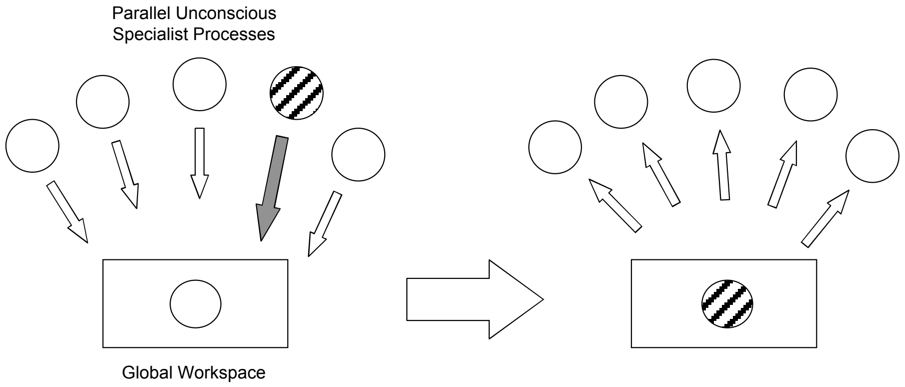

Figure 1: The Global Workspace Architecture

The essence of the global workspace architecture is a model of parallel information flow. Multiple parallel *specialist processes* compete and co-operate for access to a *global workspace* (Figure 1, left). A specialist process can be responsible for some aspect of perception, long-term planning, problem solving, language understanding, language production, action selection, or indeed any posited cognitive process. If granted access to the global workspace, the information a process has to offer is broadcast back to the entire set of specialists (Figure 1, right). The means by which access is granted to the global workspace corresponds to an attention mechanism.

According to the global workspace hypothesis, the mammalian brain is organised in this way, and using this architectural blueprint it is possible to distinguish between conscious and unconscious information processing. Unconscious information processing is carried out by the parallel specialist processes. Only information that is broadcast via the global workspace is consciously processed.

Possible neural mechanisms capable of supporting a global workspace are the subject of ongoing research. A naïve interpretation of the architectural proposal in Figure 1 suggests an analogy with the computer scientist's notion of data transmitted along a bus and stored in a buffer. But this makes little sense at the neurological level.[15] Rather, what is sought is a mechanism by which the activity of small selected portions of cortex – those that have gained access to the global workspace – can systematically influence the activity of the rest of cortex.[16] A likely means of achieving this is through thalamo-cortical interaction (Llinás, *et al.*, 1998; Edelman & Tononi, 2000, pp. 148–149).[17]

The thalamo-cortical interaction hypothesis is consistent with the recently reported discovery of a pan-cortical pattern of aperiodic alternation between coherent EEG activity and decoherent EEG activity (Rodriguez, *et al.*, 1999; Freeman & Rogers, 2003).[18] This phenomenon is suggestive of repetitive switching between episodes of competition for access to the global workspace (marked by decoherent electrical activity), and episodes of passive "listening in" by the rest of cortex to the neural population thus granted access (marked by coherent electrical activity). The frequency of this oscillatory pattern (in the alpha range of 7-12 Hz) is compatible with

[15] Although concepts inherited from computer science are not always applicable to the biological brain, they remain valid building blocks from a purely engineering standpoint. To build a robot that doesn't suffer from the frame problem, the whole spectrum of implementation options for a global workspace architecture remains open.

[16] The distinction between global *access* and global *broadcast* breaks down at this level, and these terms are used interchangeably. Dehaene, *et al.* (2003) use the evocative term "ignition" for the sudden eruption of widespread brain activity that is the proposed signature of conscious information flow.

[17] Dehaene, *et al* . (2003) present computer model that implements a global neuronal workspace through thalamo-cortical interaction, and use it to simulate the attentional blink.

[18] See also the review by Varela, *et al.* (2001).

widely accepted figures for the minimum time required for a stimulus to become conscious (approximately 250 ms) (Libet, 1996).[19]

It would make sense, in terms of global workspace theory, if an episode of decoherent activity came to an end when positive feedback enabled one or more small neural sub-populations to emerge as dominant.[20] Using lateral inhibition to suppress other cortical regions, a dominant population could start to drive the rest of cortex via the thalamus, allowing its message to be heard by all. The result would be an episode of coherent activity. But this would destabilise as soon as the influence of the driver population(s) started to fade and the newly broadcast pattern of activation enabled previously suppressed populations to compete for access again.

There is a growing body of other empirical evidence in support of global workspace theory (Dehaene & Naccache, 2001; Baars, 2002). In particular, the hypothesis that consciousness enables global activation has now been supported in dozens of brain imaging studies. The typical experimental paradigm is based on *contrastive analysis* (Baars, 1988), wherein closely matched conscious and unconscious conditions are compared in waking subjects, either by stimulus manipulation (binocular rivalry, masking, attentional tasks, etc.) or by overpractice of automatic habits.

In all cases tested so far, the conscious condition recruits very widespread cortical resources while the matched unconscious event typically activates local regions only (Baars, 2002; 2003). For example, in an fMRI study by Dehaene, *et al.* (2001) using visual backward masking, unconscious words activated visual word areas only, while the identical conscious words evoked widespread activation of parietal and prefrontal cortex as well. Kreiman, *et al.* (2003) have shown that medial-temporal areas are also activated by pictures, but only if they were conscious, using implanted electrodes in epileptic patients. These regions are specialized for memory and emotion.

[19] The interpretation of Libet's data is controversial (see the whole of *Consciousness and Cognition* 11(2)). Pockett (2002), for example, using Libet's original data, revises the stimulus-to-awareness time down to approximately 80 ms. But this is still compatible with the present interpretation of Freeman's results.

[20] As well as allowing the rest of cortex to "listen in", an episode of coherence might facilitate temporal binding among multiple populations (Engel, *et al.*, 1999).

In a complementary experimental paradigm, brain response to stimulation has been compared in conscious versus unconscious states. Unconscious states studied include sleep, general anesthesia, epileptic loss of consciousness and vegetative states. Sensory stimulation in all four unconscious states evokes only local cortical responses, but not the global recruitment characteristic of sensory input in conscious subjects (Baars, *et al.*, 2004). This general pattern of results has now been shown for vision, hearing, pain perception, touch, and sensorimotor tasks. It appears that conscious events recruit global activity in the cerebral cortex, as predicted by the theory.

An intuitively pleasing consequence of the model of information flow proposed in (Baars, 1988) is that the contents of the global workspace unfolds in a serial manner, yet it is the product of massively parallel processing. In accordance with subjective experience, a sequence of moment-to-moment snapshots of the contents of the global workspace would reveal a meaningful progression, and each state would typically be related to its predecessor in a coherent way. Yet the state-to-state relation itself is highly complex, and is certainly not obtainable by a single computational step in a serial von Neumann architecture. Rather, it is the outcome of a selection procedure that has at its disposal the results of numerous separate computations, each of which might have something of value to contribute to the ongoing procession of thoughts. To put it another way, the global workspace has limited capacity, but it enjoys vast access to the brain's resources (Baars, 1997).

From an engineering standpoint, the global workspace architecture has a pedigree dating back to the earliest days of artificial intelligence. Its origins lie in Selfridge's *pandemonium* architecture (1959), which inspired Newell's *blackboard* metaphor for problem solving (1962).

> Metaphorically we can think of a set of workers, all looking at the same blackboard: each is able to read everything that is on it, and judge when he has something worthwhile to add to it.[21]

[21] Newell (1962).

Newell's model was successfully applied to the problem of speech recognition in the 1970s (Erman & Lesser, 1975), and this led to the development of so-called *blackboard architectures* for AI systems in the 1980s (Hayes-Roth, 1985; Nii, 1986). These systems supplied the original inspiration for the global workspace model of conscious information processing presented in (Baars, 1988). Nowadays, the blackboard architecture is a standard piece of AI technology, and with the recent design of AI software agents based on Baars' global workspace theory, the influence of the blackboard architecture has come full circle (Franklin & Graesser, 1999; 2001).

The global workspace architecture is advantageous from a purely computer science perspective, but it's noteworthy that these advantages gel with empirical considerations. First, it is easily deployed on parallel hardware, which is an obvious requirement for a neurologically plausible cognitive architecture. Second, it ensures that the overall system degrades gracefully when individual processes break down, which is compatible with the brain's robustness. Third, it facilitates the addition of new processes without modification to the existing system, which renders it consistent with evolutionary constraints. Fourth, it accommodates heterogeneous forms of information processing and integrates their results.

The final point is critical here, since it is precisely the ability to manage a high bandwidth of information generated by a large number of distinctive, special-purpose parallel processes that enables a global workspace architecture to realise an informationally unencapsulated cognitive process without falling foul of complexity problems. Notably, the concurrent activation of multiple disparate brain processes is also the signature of *conscious* information processing according to the global workspace hypothesis. In other words, consciousness goes hand-in-hand with a solution to the frame problem in the biological brain.

6. Analogical Reasoning

Fodor says little about the computational model behind his claim that informationally unencapsulated cognitive processes are computational infeasible. Yet there are strong hints of a commitment to a centralised, serial process that somehow has all the requisite information at its disposal, and that has the responsibility of choosing what information to access and when to access it. Although parallel peripheral processes are part of the picture, they are *passive* sources of information that wait to be called upon before delivering their goods (Figure 2).[22] This centralised, serial model of computation is evoked by Dennett's caricature robot, by the idea of "Hamlet's problem", and by such phrases as "the totality of one's epistemic commitments is vastly too large a space to have to search" (Fodor) and "access to the full set of an agent's background beliefs gives rise to an unmanageable combinatorial explosion" (Carruthers).

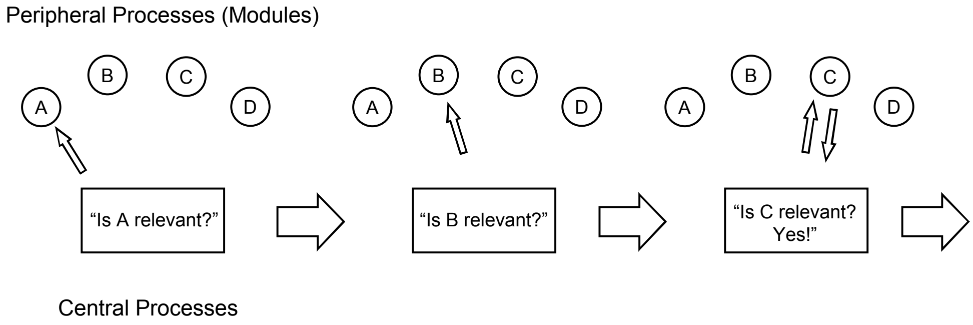

Figure 2: A Naïve Model of Information Flow

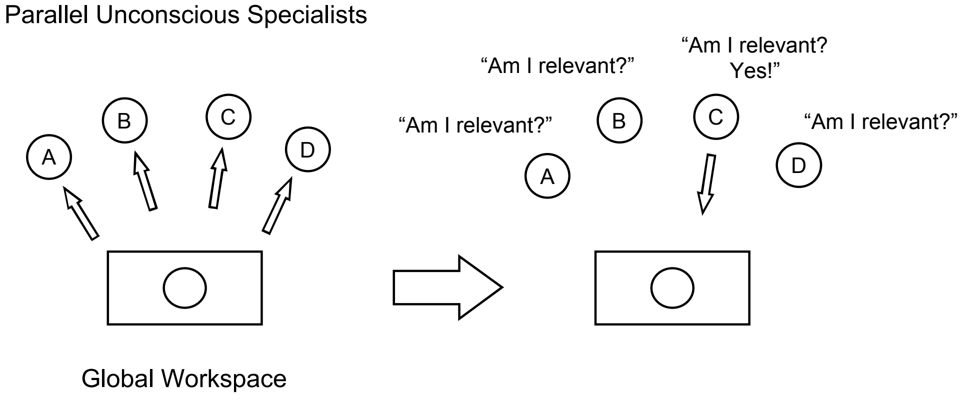

Figure 3: The Global Workspace Model of Information Flow

[22] Fodor's peripheral processes perform both input and output. The emphasis of the present discussion is on perception.

By contrast, global workspace theory posits multiple, parallel processes that all contribute *actively* to cognition (Figure 3). Consider the computational processes that might underlie the likening of a Rorschach inkblot to, say, an elephant. Fodor's argument hints at a centralised process that poses a series of questions one-at-a-time – is it a face? is it a butterfly? is it a vulva? and so on – until it finally arrives at the idea of an elephant. Instead, the global workspace model posits a specialist, parallel process that is always on the lookout for elephantine shapes.[23] This process is aroused by the presence of an inkblot that actually resembles an elephant, and it responds by announcing its findings. The urgency with which this process commends itself means that the information it has to offer makes its way into the global workspace, and is thereby broadcast back to all the other specialist processes.[24]

The Rorschach inkblot example is especially pertinent, since it is closely related to analogical reasoning. As Fodor emphasises, reasoning by analogy is the epitome of an informationally unencapsulated cognitive process because it "depends precisely upon the transfer of information among cognitive domains previously assumed to be irrelevant" (Fodor, 1983, p.□105). Moreover, analogical reasoning is arguably central to human cognitive prowess (Gentner, 2003). So we should be able to offer a *prima facie* case that the global workspace architecture relieves the computational burden allegedly brought on by the informational unencapsulation of analogical reasoning. The Rorschach blot example is a start.[25] But a more convincing case can be made in the context of a more fully realised theory

[23] This might seem to be taking specialisation too far. But with 100 billion neurons and many trillions of synapses, the human brain has plenty of scope for truly massive parallelism, and without resort to "grandmother cell" coding.

[24] The use of words like "aroused", "announced", and "urgency" here is purely stylistic, and doesn't mask the lack of an implementable concept. Competitive access to a global workspace can be realised in a working computational model in various ways. For example, each parallel process can be made responsible for assigning a value to its own relevance in a given situation, and a central winner-takes-all strategy can be used to select the process to be granted access. For working implementations of the global workspace model see (Franklin & Graesser, 1999; 2001).

[25] The Rorschach blot is an example of literal similarity rather than true analogy.

of analogical thinking. So what follows is a short review of recent research on the computational aspects of analogical reasoning.[26]

6.1. Computational Models of Analogical Reasoning

Most recent computer models of analogical reasoning take their cue from the *structure mapping* theory of analogy (Gentner, 1983). According to Gentner's theory, a central task in analogical reasoning is to find "maximally structurally consistent" mappings between two representations – a *target* located in working memory and a *base* located in long-term memory.[27] A *structurally consistent* mapping is one that conforms to certain constraints. In particular, no element in the target/base can be mapped on to more than one element in the base/target, and if two elements are put in correspondence then their sub-elements must also be put in correspondence. Finally, the theory includes a heuristic of *systematicity*, whereby deep and rich mappings are preferred to shallow and sparse ones.[28]

The Structure Mapping Engine (SME) is a computational realisation of Gentner's structure mapping theory (Falkenheimer, *et al.*, 1989). Its core is an algorithm for finding mappings between a pair of propositional representations that conform to the principles of the theory. It can find a mapping between two given representations in polynomial time, and it assimilates new information incrementally. ACME (Analogical Constraint Mapping Engine) is another computer implementation of the mapping process that adheres to the chief tenets of Gentner's structure mapping theory (Holyoak & Thagard, 1989). But unlike SME, which uses graph matching, ACME views the mapping process as a constraint satisfaction problem and employs connectionist techniques to solve it.

[26] A related topic in AI is *case-based reasoning* (Kolodner, 1993). However, case-based reasoning tends to operate within domain boundaries, while the nub of the present discussion is the ability to cross domain boundaries.

[27] The ensuing discussion takes place against the backdrop of a language-like, viewpoint-free representational medium (even when realised in a distributed connectionist network). But global workspace theory is equally applicable to an analogical, viewer-centred representational medium, as recommended in (Shanahan, 2004).

[28] The "systematicity" of an analogical mapping in Gentner's terminology has no direct connection with the "systematicity" of thought (Fodor & Pylyshyn, 1988).

Both the SME and ACME models find mappings between two *given* representations, and neither addresses the question of *retrieval* (or *access*), that is to say how to find candidate base representations from a large pool of possibilities. Retrieval is the aspect of analogical reasoning responsible for seeding "the transfer of information among cognitive domains previously assumed to be irrelevant". So the retrieval process is inherently informationally unencapsulated, and is therefore the locus of the frame problem for analogical reasoning. To quote Thagard, *et al.* (1990), "the problem of analogical retrieval is how to get what you want from memory without getting more than you can use".[29]

The retrieval problem has been tackled in the context of ACME by a computational model called ARCS (Thagard, *et al.*, 1990), and in the context of SME by a computational model called MAC/FAC (Forbus, *et al.*, 1994). The ARCS model (Analogical Retrieval by Constraint Satisfaction) employs the same combination of constraint satisfaction and connectionism as the ACME system from which it is descended. The ARCS retrieval process is executed in two stages. The first stage finds candidate base representations using pre-coded notions of semantic similarity, while the second stage evaluates how well these candidates satisfy a number of generic constraints on analogical mapping related to those advocated by Gentner's theory. The MAC/FAC model is a similar two-stage retrieval process. In the MAC stage (Many Are Called), a crude but computationally cheap filter is used to isolate a manageably small number of potential base representations for mapping onto a given target. In the ensuing FAC stage (Few Are Chosen), SME is used to select the best among the candidate mappings recommended by the MAC stage.

Both the SME-MAC/FAC model and the ACME-ARCS model can account for a good deal of well established psychological data about analogical reasoning, as enumerated in the key papers already cited. However, they also have limitations. Of especial interest here, as we'll see later, are certain criticisms levelled at these models by Keane, *et al.* (1994)

[29] Thagard, *et al.* (1999), p. 261. As they go on to say, "this is the core problem of memory retrieval in general". It's also a good redescription of what, following Fodor, this article means by the frame problem itself.

and Hummel & Holyoak (1997). Both sets of authors draw attention to the unrealistic demands these models place on working memory capacity. Here is a summary from (Hummel & Holyoak, 1997).

> Both ACME and SME form explicit representations of very large numbers of possible local matches between elements of source and target analogs, most of which are then discarded as mapping proceeds. Such processes do not seem compatible with the generally accepted limits of working memory.

This, alongside other concerns, motivated the development of IAM (Keane, *et al.*, 1994) and LISA (Hummel & Holyoak, 1997). Both IAM (Incremental Analogy Machine) and LISA (Learning and Inference with Schemas and Analogies) carry out mapping serially and incrementally, processing small propositional structures one at a time. For this reason, they are able to emulate human analogical reasoning performance more closely than their predecessors, duplicating the difficulties humans have in forming "unnatural analogues" and human sensitivity to the order in which the parts of the target and base representations are processed.

The IAM model tackles mapping only. However, the LISA model tackles retrieval and mapping together, reflecting another of Hummel and Holyoak's desiderata, namely the need for a graceful integration of these two aspects of analogical reasoning. The result is a system in which retrieval is naturally cast as a parallel process while mapping is inherently serial. For this reason, the LISA model – which is arguably the most accurate and psychologically plausible of the current computational models of analogical reasoning[30] – is also the one that maps most cleanly and elegantly onto a global workspace architecture.

6.2. The Interplay of Parallel and Serial Computation

The thrust of the ongoing argument is that the global workspace architecture eases the computational burden on an informationally unencapsulated process through its inherent parallelism. Accordingly, this section asks whether it is possible to map a computational model of analogical reasoning

[30] See (Keane & Eysenck, 2000, p. 436), for example.

to the global workspace architecture in order to realise the potential for parallel execution. The main focus of attention is the retrieval process. As their respective authors emphasise, the ARCS, MAC/FAC, and LISA models of retrieval are each amenable to parallel implementation. But in the context of the LISA model, we shall also look at the interplay of parallel and serial computation.

We begin with ARCS. Following Thagard, *et al.* (1990), the time complexity analysis for a single run of ARCS must consider two components – the time required to form a network of mapping constraints across two potential representations and the time taken for the corresponding network to settle.[31] The former is purportedly $O(n^2m^4)$ for a serial machine, where n is the total number of candidate base representations and m is the number of propositions in the largest base or target representation. This is not a favourable statistic, but it is the worst-case analysis and the authors' experimental data is more encouraging. Moreover, it seems likely that a parallel implementation would reduce the worst-case statistic to better proportions, although this possibility is not discussed in (Thagard, *et al.*, 1990).

The explicitly claimed advantage of parallelism for ARCS hinges on the second component of the complexity analysis, namely the time required for the network to settle (to find a solution). The authors do not offer a worst-case analysis of this statistic, but they appeal to empirical results to justify the claim that parallel implementation would reduce it from $O(n)$ to constant time. Although parallelism is clearly helpful for this component of the computation, the dominant term in the overall analysis is still n^2m^4, and this is worrisome if n is very large. This remains true in spite of the authors' empirical evidence, which was gathered with a comparatively small set of potential base representations. Therefore, since a very large n is precisely what the frame problem is all about, there is only modest support for the ongoing argument from the ARCS direction.

So let's move on to the MAC/FAC model. Both the MAC and FAC stages employ a set of parallel *matchers* that compare the target to each

[31] Thagard, *et al*. (1990), p. 282. The authors also offer a space complexity analysis, which isn't relevant here.

potential base in long-term memory and a single *selector* that chooses among the outputs of the matchers. Since the matchers are completely independent, the time complexity of a maximally parallel implementation reduces to that of finding a single match hypothesis between a given target and base. According to Forbus, *et al.* (1994), this is $O(m)$ in the worst case on a parallel machine and should typically be better than $O(\log m)$, where m is the number of propositions in the larger of the base or target representation.[32]

This is a far more pleasing complexity result than was obtained for ARCS. It suggests that analogical reasoning, if implemented on massively parallel hardware, is computationally feasible, even for a very large pool of potential base representations. Moreover, in global workspace terms, the array of parallel matchers in MAC/FAC corresponds directly to specialist, unconscious processes, the selector process corresponds to the attention mechanism that gates access to the global workspace, and the working memory in which the final mapping is carried out corresponds to the global workspace itself.

The fact that analogical reasoning can be rendered computationally feasible through implementation on a form of parallel hardware that is compatible with global workspace theory is very encouraging. However, other models of parallel computation would fit MAC/FAC just as well as the global workspace architecture. The aim here is to demonstrate a more intimate link between the demands of informationally unencapsulated cognitive processes in general and specific features of the global workspace architecture. In this respect, the LISA model turns out to be more pertinent than MAC/FAC.

Although they don't offer a formal complexity analysis, Hummel and Holyoak (1997) state that: "During mapping, driver propositions are activated serially … At the same time, recipient propositions respond in parallel to the semantic patterns generated by the role-filler bindings of an active driver proposition." (In Hummel and Holyoak's terminology, "driver propositions" belong to the representation resident in working memory,

[32] Forbus, *et al*. (1994), p. 160.

while "recipient propositions" belong to representations held in long-term memory.) So LISA effectively carries out retrieval in parallel and mapping serially. Consequently, as with MAC/FAC, retrieval time will not grow with the size of the pool of potential base representations, given sufficient parallelism in the underlying hardware.

Additionally, the combination of parallel and serial processing makes LISA more compatible with global workspace theory than any of its rivals. The currently active proposition in LISA's working memory corresponds to the current contents of the global workspace. The parallel activation of propositions in LISA's long-term memory corresponds to the broadcast of the contents of the global workspace. And the distributed propositions in LISA's long-term memory correspond to parallel, unconscious processes in global workspace theory. But most importantly, the necessarily *serial* presentation of propositions in LISA's limited capacity working memory matches the necessarily *serial* presentation of coherent material in the limited capacity global workspace.[33] It is surely no accident that the features of the LISA model that make it more psychologically plausible than its competitors are the very same features that also align it so closely with global workspace theory.

7. Discussion

Let's review the argument so far. We set out by undermining the in-principle claim that informationally unencapsulated cognitive processes are computationally infeasible. It turned out that the case put forward by Fodor and others is too weak to sustain such a conclusion. The way the biological brain handles such processes is thereby demoted from an out-and-out mystery to a scientific challenge. The global workspace architecture, with its blend of parallel and serial computation, was then proposed as an answer to this challenge.

In place of the naïve, serial, exhaustive search that seems to be behind the intuitions of those authors who see the frame problem a serious obstacle,

[33] Also, the mechanism of temporal binding deployed in LISA is compatible with a thalamocortically realised global workspace along the lines proposed by Engel, *et al.* (1999), as remarked in an earlier footnote.

the global workspace architecture supports the idea of a single, serial thread of computation that can also draw on the resources of a massively parallel set of distributed processes. A review of current computational models of analogical reasoning – often taken to be the epitome of informational encapsulation – demonstrated a close fit between global workspace theory and the most psychologically plausible of these models, namely that of Hummel and Holyoak (1997).

The compatibility between Hummel and Holyoak's model of analogical reasoning and the global workspace architecture lends mutual support to both theories. Of particular interest is the prospect of accounting for the fact that, while some aspects of analogical reasoning are undoubtedly carried out unconsciously, others plainly are not. Observing the 20th Century taboo on discussions of consciousness, most analogical reasoning researchers have chosen to ignore this issue. Contemporary accounts of analogical reasoning tend to speak in neutral terms of a target representation in "working memory" being mapped to some base representations retrieved from long-term memory.[34] But working memory is conventionally characterised as a temporary repository of rehearsible material (Baddeley, 1986). Since there is surely no such thing as unconscious rehearsal, this concept of working memory is irrevocably tied to that of consciousness (Baars & Franklin, 2003), and by implication so is analogical reasoning.[35]

Thankfully, there is no longer any need to shy away from this fact. Global workspace theory can offer an intellectually respectable account of the conscious aspects of analogical reasoning. Moreover, global workspace theory rests on the hypothesis that conscious information processing is cognitively efficacious because it integrates the functionality of numerous separate and independent brain processes. This suggests a deep explanation of the psychological plausibility of models of analogical reasoning that combine serial and parallel computation. Such models fit human behaviour because the human brain is built to take advantage of the cognitive efficacy

[34] See Waltz, *et al*. (2000) for a detailed psychological account of the role of working memory in analogical reasoning.

[35] As acknowledged by Courtney, *et al*. (1999), "by definition, working memory includes those processes that enable us to hold in our 'mind's eye' the contents of our conscious awareness". Hummel and Holyoak (1997) also refer to working memory as "the current contents of 'awareness'".

of conscious information processing realised through a global workspace architecture with a distinctive serial/parallel mix.

The discussion has for some time concentrated on analogical reasoning, which is of course only one among many informationally unencapsulated cognitive processes. However, another advantage of bringing a computational model of a specific cognitive process within the wider compass of global workspace theory is that it embeds that model in a generic account of cognitive function. Instead of seeing analogical reasoning as an isolated process, it comes to be seen as interleaved with many other cognitive activities – such as perception, action selection, planning, belief revision, language production, language comprehension, and so on – each of which can exploit the brain's diversity of massively parallel specialist processes.[36]

This brings us to our final topic of discussion, namely the implications of the present proposal for the so-called *modular* theories of the human mind that many contemporary cognitive scientists subscribe to in one form or another (Fodor, 1983; Tooby & Cosmides, 1992; Karmiloff-Smith, 1992; Sperber, 1994; Mithen, 1996; Pinker, 1997; Fodor, 2000; Carruthers, 2002). Although modular theories come in various guises, nearly all share two central tenets. First, that some proportion of human mental structure comprises a set of special-purpose modules with a narrow remit. Second, that the flexibility and creativity of human thought demands an additional mechanism capable of transcending rigid modular boundaries.

The second of these tenets is the flash-point for the present debate, and authors differ widely in the mechanisms they propose. Fodor (1983; 2000) assigns cross-modular capabilities to the mind's "central processes", which he believes to be beyond the reach of present-day cognitive science, laying much of the blame for this on the frame problem. Karmiloff-Smith (1992), writing from a developmental perspective, hypothesises a stage of "representational redescription" whereby knowledge that was previously represented implicitly for a child becomes explicitly represented *to* the child. Sperber (1994) posits a "metarepresentation module", which can

[36] Forbus (2001) makes a similar recommendation.

manipulate representations about representations and concepts of concepts, allowing knowledge from multiple domains to be blended. Mithen (1996), drawing on archaeological evidence, introduces the idea of human “cognitive fluidity”, which permits “thoughts and knowledge generated by specialized intelligences [to] flow freely around the mind”.[37] Carruthers (2003) suggests that “natural language is the medium for non-domain-specific thinking, serving to integrate the outputs of a variety of domain-specific conceptual faculties”.[38]

Global workspace theory, with its commitment to multiple, parallel, specialist processes, is consistent with a modular view of the human mind. What the global workspace architecture has to offer in addition is a model of information flow that explains how an informationally unencapsulated process can draw on just the information that is relevant to the ongoing situation without being swamped by irrelevant rubbish. This is achieved by distributing the responsibility for deciding relevance to the parallel specialists themselves. The resulting massive parallelism confers great computational advantage without compromising the serial flow of conscious thought, which corresponds to the sequential contents of the limited capacity global workspace.

In the context of a global workspace architecture, there are no distinguished central processes with special privileges or powers. Rather, every high-level cognitive process can be analysed in terms of the interleaved operation of multiple, specialist, parallel processes, all of which enjoy equal access to the global workspace. Analogical reasoning, for example, can recruit the sort of capacity to integrate spatial relations allegedly located in prefrontal cortex (Waltz, *et al.*, 1999), alongside whatever other brain resources it may require. The significance of informational encapsulation is thereby much diminished, and the spectre of the frame problem is dissolved.

[37] Mithen (1996), p. 71.

[38] Carruthers (2003), p. 657.

Acknowledgments

Thanks to Igor Aleksander, Ron Chrisley, Stan Franklin, and Mercedes Lahnstein.

References

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

Baars, B.J. (1997). *In the Theater of Consciousness: The Workspace of the Mind.* Oxford University Press.

Baars, B.J. (2002). The Conscious Access Hypothesis: Origins and Recent Evidence. *Trends in Cognitive Science* 6(1), 47–52.

Baars, B.J. (2003). How Brain Reveals Mind: Neuroimaging Supports the Central Role of Conscious Experience. *Journal of Consciousness Studies* 10(9–10), 100–114.

Baars, B.J. & Franklin, S. (2003). How Conscious Experience and Working Memory Interact. *Trends in Cognitive Science* 7(4), 166–172.

Baars, B.J., Ramsoy, T.Z. & Laureys, S. (2003). Brain, Consciousness and the Observing Self. *Trends in Neurosciences* 26 (12), 671–675.

Baddeley, A.D. (1986). *Working Memory.* Oxford University Press.

Carruthers, P. (2002). The Cognitive Functions of Language. *Behavioral and Brain Sciences* 25(6), 657–674.

Carruthers, P. (2003). On Fodor’s Problem. *Mind and Language* 18(5), 502–523.

Cherniak, C. (1986). *Minimal Rationality.* MIT Press.

Courtney, S.M., Petit, L., Haxby, J.L. & Ungerleider, L.G. (1998). The Role Of Prefrontal Cortex in Working Memory. *Philosophical Transactions of the Royal Society B* 353, 1819–1828.

Dehaene, S. & Naccache, L. (2001). Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework. *Cognition* 79, 1–37.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.F., Poline, J.B. & Riviere, D. (2001). Cerebral Mechanisms of Word Masking and Unconscious Repetition Priming. *Nature Neuroscience* 4, 752–758.

Dehaene, S., Sergent, C. & Changeux, J.-P. (2003). A Neuronal Network Model Linking Subjective Reports and Objective Physiological Data During Conscious Perception. *Proceedings of the National Academy of Science* 100(14), 8520–8525.

Dennett, D. (1978). *Brainstorms*. MIT Press.

Dennett, D. (1984), Cognitive Wheels: The Frame Problem in Artificial Intelligence. In *Minds, Machines and Evolution*, ed. C.Hookway, Cambridge University Press, pp. 129–151.

Edelman, G.M. & Tononi, G. (2000). *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books.

Engel, A.K., Fries, P., Köning, P., Brecht, M. & Singer, W. (1999). Temporal Binding, Binocular Rivalry, and Consciousness. *Consciousness and Cognition* 8, 128–151.

Erman, L.D. & Lesser, V.R. (1975). A Multi-Level Organization for Problem Solving Using Many, Diverse, Cooperating Sources of Knowledge. In *Proceedings Fourth International Joint Conference on Artificial Intelligence (IJCAI 75)*, pp. 483–490.

Falkenheimer, B., Forbus, K. & Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence* 41, 1–63.

Fodor, J.A. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J.A. (1983). *The Modularity of Mind*. MIT Press.

Fodor, J.A. (1987). Modules, Frames, Fridgeons, Sleeping Dogs, and the Music of the Spheres. In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, ed. Z.W.Pylyshyn, Ablex, pp. 139–149.

Fodor, J.A. (2000). *The Mind Doesn't Work That Way*. MIT Press.

Fodor, J.A. & Pylyshyn, Z.W. (1988). Connectionism and Cognitive Architecture: A Critique. *Cognition* 28, 3–71.

Forbus, K. (2001). Exploring Analogy in the Large. In *The Analogical Mind: Perspectives from Cognitive Science*, eds. D.Gentner, K.J.Holyoak & B.N.Kokinov, MIT Press, pp. 23–58.

Forbus, K., Gentner, D. & Law, K. (1994). MAC/FAC: A Model of Similarity-Based Retrieval. *Cognitive Science* 19, 141–205.

Franklin, S. & Graesser, A. (1999). A Software Agent Model of Consciousness. *Consciousness and Cognition* 8, 285–301.

Franklin, S. & Graesser, A. (2001). Modelling Cognition with Software Agents. In *Proc. 23rd Annual Conference of the Cognitive Science Society*.

Freeman, W.J. & Rogers, L.J. (2003). A Neurobiological Theory of Meaning in Perception Part V: Multicortical Patterns of Phase Modulation in Gamma EEG. *International Journal of Bifurcation and Chaos* 13(10), 2867–2887.

Gabaix, X. & Laibson, D. (2003). A New Challenge for Economics: "The Frame Problem". In *The Psychology of Economic Decisions, Volume One: Rationality and Well-Being*, eds. I.Brocas & J.D.Carrillo, Oxford University Press, pp. 169–183.

Garey, M.R. & Johnson, D.S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H.Freeman & Co.

Gentner, D. (1983). Structure-mapping: A Theoretical Framework for Analogy. *Cognitive Science* 7: 155-170.

Gentner, D. (2003). Why We're So Smart. In *Language in Mind*, ed. D.Gentner & S.Goldin, MIT Press, pp. 195–235.

Haselager, W.F.G. & van Rappard, J.F.H. (1998). Connectionism, Systematicity, and the Frame Problem. *Minds and Machines* 8(2), 161–179.

Hayes-Roth, B. (1985). A Blackboard Architecture for Control. *Artificial Intelligence* 26(3), 251–321.

Holyoak, K.J. & Thagard, P. (1989). Analogical Mapping by Constraint Satisfaction. *Cognitive Science* 13, 295–355.

Hummel, J.E. & Holyoak, K.J. (1997). Distributed Representations of Structure: A Theory of Analogical Access and Mapping. *Psychological Review* 104(3), 427–466.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press.

Keane, M.T., Ledgeway, T. & Duff, S. (1994). Constraints on Analogical Mapping: A Comparison of Three Models. *Cognitive Science* 18, 387–438.

Keane, M.T. & Eysenck, M.W. (2000). *Cognitive Psychology: A Student's Handbook (4th edition)*. Psychology Press.

Kolodner, J.L. (1993). *Case-Based Reasoning*. Morgan Kaufmann.

Kreiman G, Fried, I. & Koch, C. (2002). Single-Neuron Correlates of Subjective Vision in the Human Medial Temporal Lobe. *Proceedings of the National Academy of Science* 99(12), 8378-8383.

Libet, B. (1996). Neural Processes in the Production of Conscious Experience. In *The Science of Consciousness: Psychological, Neuropsychological, and Clinical Reviews*, ed. M.Velmans, Routledge, pp. 96–117.

Llinás, R., Ribary, U., Contreras, D. & Pedroarena, C. (1998). The Neuronal Basis for Consciousness. *Philosophical Transactions of the Royal Society B* 353, 1841–1849.

McCarthy, J. & Hayes, P.J. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In *Machine Intelligence 4*, eds. D.Michie & B.Meltzer, Edinburgh University Press, pp. 463-502.

Mithen, S. (1996). *The Prehistory of the Mind.* Thames & Hudson.

Newell, A. (1962). Some Problems of Basic Organization in Problem-Solving Systems. In *Proceedings of the Second Conference on Self-Organizing Systems*, pp. 393–342.

Nii, H.P. (1986). The Blackboard Model of Problem Solving. *AI Magazine* 7(2), 38–53.

Pinker, S. (1997). *How the Mind Works*. Penguin.

Pockett, S. (2002). On Subjective Back-Referral and How Long It Takes to Become Conscious of a Stimulus: A Reinterpretation of Libet's Data. *Consciousness and Cognition* 11(2), 144–161.

Pylyshyn, Z.W. (ed.) (1987). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Ablex.

Reiter, R. (2001). *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. MIT Press.

Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B. & Varela, F. (1999). Perception's Shadow: Long-Distance Synchronization of Human Brain Activity. *Nature* 397, 430–433.

Russell, S. & Wefald, E. (1991). *Do the Right Thing: Studies in Limited Rationality*. MIT Press.

Sandewall, E. (1994). *Features and Fluents: The Representation of Knowledge about Dynamical Systems, Volume 1*. Oxford University Press.

Selfridge, O. (1959). Pandemonium: A Paradigm for Learning. In *Proceedings of the Symposium on Mechanisation of Thought Processes*, H.M.S.O. London (pp. 511–529).

Shanahan, M.P. (1997). *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*. MIT Press.

Shanahan, M.P. (2003). The Frame Problem. In *The Macmillan Encyclopedia of Cognitive Science*, ed. L.Nadel, Macmillan, pp. 144-150.

Shanahan, M.P. (2004). The Imaginative Mind: Rehearsing Trajectories Through an Abstraction of Sensorimotor Space. Submitted to *Behavioral and Brain Sciences*.

Simon, H. (1957). *Models of Man*. Wiley.

Smolensky, P. (1988). On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences* 11(1): 1–74.

Sperber, D. (1994). The Modularity of Thought and the Epidemiology of Representations. In *Mapping the Mind: Domain Specificity and Culture*, eds. L.A.Hirschfeld & L.A.Gelman, Cambridge University Press, pp. 39–67.

Thagard, P., Holyoak, K.J., Nelson, G. & Gochfeld, D. (1990). Analog Retrieval by Constraint Satisfaction. *Artificial Intelligence* 46, 259–310.

Tooby, J. & Cosmides, L. (1992). The Psychological Foundations of Culture. In *The Adapted Mind*, eds. J.H.Barkow, L.Cosmides & J.Tooby, Oxford University Press, pp. 19–136.

van Gelder, T. (1997). The Dynamical Hypothesis in Cognitive Science. *Behavioral and Brain Sciences* 21(5): 615–628.

Varela, F., Lachaux, J.-P., Rodriguez, E. & Martinerie, J. (2001). The Brainweb: Phase Synchronization and Large-Scale Integration. *Nature Reviews Neuroscience* 2, 229–239.

Waltz, J.A., Knowlton, B.J., Holyoak, K.J., Boone, K.B., Mishkin, F.S., de Menezes Santos, M., Thomas, C.R. & Miller, B.L. (1999). A System for Relational Reasoning in Human Prefrontal Cortex. *Psychological Science* 10(2), 119–125.

Waltz, J.A., Lau, A., Grewal, S.K. & Holyoak, K.J. (2000). The Role of Working Memory in Analogical Mapping. *Memory and Cognition* 28(7), 1205–1212.

Zilberstein, S. & Russell, S.J. (1993). Anytime Sensing, Planning and Action: A Practical Model for Robot Control. In *Proceedings 1993 International Joint Conference on Artificial Intelligence (IJCAI 93)* (pp. 1402–1407).