

What We Really Know About Consciousness

Review of *A Cognitive Theory of Consciousness* by Bernard Baars

Bruce Bridgeman

Department of Psychology
University of California
Santa Cruz, CA 95064
USA

bruceb@cats.ucsc.edu

Copyright (c) Bruce Bridgeman 1996
Received: May 10, 1996; Published: July 19, 1996

PSYCHE, 2(30), July 1996
<http://psyche.cs.monash.edu.au/v2/psyche-2-30-bridgeman.html>

KEYWORDS: consciousness, Global Workspace theory, functions of consciousness, modules.

REVIEW OF: Bernard J. Baars (1988) *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University. xxiii+424pp. ISBN: 0-521-42743-6. Price: US\$19.95 pbk.

Behaviorism died very slowly in American psychology. Willing to admit only stimuli and responses, narrowly defined, as the contents of its science, it was too limited to be useful in understanding higher organisms. But it seemed for decades to be the only way to eliminate subjectivism and to make psychology scientific. Cracks in behaviorist orthodoxy began to appear 40 years ago, though physiological psychologists never felt themselves bound by its restrictions. By the 1960s psychologists were openly espousing a new 'cognitive' psychology, admitting to internal states and a mental life beyond overt behavior. It was a step forward, but the block diagrams of the cognitive psychologists still had a stimulus on one end and a response on the other, just like those of the old-fashioned behaviorists. Consciousness, the great problem that got psychology started, was still viewed with alarm as the purview of charlatans and worse.

Fear of consciousness, the last remnant of behaviorist influence on psychology, has been slow to yield. Francis Crick concluded his recent book *The Astonishing Hypothesis*

(1994) with the goal "to persuade people, and especially those scientists intimately involved with the brain, that now is the time to take the problem of consciousness seriously." Bernard Baars has been taking the problem seriously for some years, and has summarized his findings in a book that is wide-ranging yet always focused on the issue of what consciousness is, what it is good for, why it is the way it is. Since psychologists have not been willing to work on problems related to consciousness, or at least not to admit it, Baars has had to assemble his evidence from a wide variety of sources. Some of the work is his own, but most comes from others. He draws on a wider range of findings than Crick, ferreting out work on such subjects as language, 'deliberate attentional resources', perception, motivation, and clinical findings, among others. All of this work is related to what for Baars is "the great, confusing and contentious nub of psychological science" (p. xv), even if the experimenters themselves didn't know it at the time.

Baars organizes this disparate literature into what he calls the 'Global Workspace Theory'. It is a thoroughly mechanistic, neurological idea, more than just another name for consciousness. The approach is thoroughly grounded in evolutionary theory, though Baars barely uses the term explicitly. He addresses questions of function: what problems does consciousness solve, how did it develop? The book is encyclopedic in its coverage and always responsible in its science. It shows by example that serious work on consciousness is both possible and productive. The book is well-written, if sometimes repetitive, and deserves the attention of anyone interested in this great and central problem.

The global workspace is a neuronal machine, a hypothesized algorithm for the workings of the highest and most general levels of brain organization. A group of modules, similar to the demons in Selfridge's 'pandemonium' model (1959), compete for access to the global workspace. Outputs of perceptual systems, needs such as hunger, goals, and emergencies all crowd around the table of the global workspace, clamoring for the attention of the organism and access to coordinated behavior. One or a small number of these modules gets selected, and dominates until its task is completed or another demon gains ascendancy. It becomes the momentary content of consciousness. Selection is based on the urgency of the request, with each module inhibiting the others. The selected module broadcasts its contents widely, recruiting parts of the brain that are needed to perform the selected function. At the same time, there is both access to memory systems and recording of the event in experiential memory. We can think of episodic memory as recording whatever is selected.

This idea is distinct from the 'Cartesian theater' that Dennett and Kinsbourne (1992) have recently discredited. There is no homunculus watching the performance in the global workspace. Once the various modules compete for access, the activity of the successful modules is distributed widely throughout the brain (Baars cites neurophysiological evidence for such a process), where other modules use the output of the "conscious" module. The global workspace is similar to Dennett and Kinsbourne's 'multiple drafts' concept, though the Baars theory is more firmly grounded in neurophysiological and

functional

analysis.

In fact, we can go a step further than Baars and look at what the global workspace must contain. The answer is -- nothing! It is a virtual workspace, a neurological metaphor. The modules talk directly with one another, and the decision about which dominates is made by mutual inhibition. No machinery has to decide where the message goes, because it goes everywhere. Baars uses the metaphor of broadcasting rather than telephone-like links for this process. Each module is under local control and responds only to its immediate inputs.

The global workspace is first introduced as a simple summation, an intermediate level between currently unconscious goals and more specialized processors. As the evidence accumulates, the hypothesis is progressively elaborated in a series of figures. From an engineering standpoint the figures are somewhat frustrating, because they seem to promise an information flow diagram but do not deliver it. There are boxes with outputs but no inputs and vice versa, and many of the functions are ill-defined. Still, a wealth of information is integrated within the language of the hypothesis; considered as a consistent format for describing the integrative processes that underlie conscious activity, the hypothesis has value.

Some intriguing new interpretations grow from the global workspace idea. For example, the conception of the workspace as a competitive, resonance phenomenon implies that it takes time for a module to attain dominance. And if the time needed to activate episodic memory is longer than the time needed to broadcast the initial message, information might be available very briefly but fail to affect behavior after a short delay. Thus, a natural explanation of Sperling's (1960) partial report phenomenon emerges. A subject is exposed to an array of letters for a few milliseconds; only a few of the letters can be reported, as though the subject forgets the rest of them before he can reproduce them. But if a probe follows immediately, indicating what row or column is to be reported, recall is nearly perfect. When the probe is delayed even for half a second, performance falls back to random. For Baars, the limited capacity of the global workspace broadcasting system militates against making the entire array available to consciousness. But if a high-capacity, low-level module can select a part of the array and store it in a longer-term buffer memory before it disappears, the subject could report that part with ease. In Baars' terminology (p. 220), the global workspace is used for very rapid exchange of information.

In a similar way the global workspace hypothesis can explain a myriad of other experimental effects and common experiences. Perhaps the model is too powerful -- it is difficult to imagine a situation that the workspace cannot handle. Of course the accounts in the book are explanations for existing results, so the analysis is necessarily post-hoc. Since the theory is a new one, there is no alternative but to test it at first against existing data, and then to turn to the future project of generating and testing predictions from it.

The global workspace conception should be equally useful for this task of predicting results in specific experimental contexts. Presumably, Baars or others will make falsifiable predictions from the theory and test them experimentally.

There are two schools of thought on consciousness. One holds that it has functions like any other neurological system, while the other sees it as an epiphenomenon that bestows no particular benefits. Baars quotes William James and Thomas Huxley respectively to show the long histories of the two schools. Rather than offering polemics on either side, Baars quietly goes about identifying functions of consciousness, answering the question "How does the global workspace benefit human function?" His nine major functions are summarized on p. 349.

(1) The first is definition and context-setting. The system underlying consciousness defines the input and cleans up the message, removing ambiguities by filling in from memories that can be accessed only through the broadcast mode of the global workspace. Recruiting and control functions (2) are natural corollaries of this function. Adaptation and learning (3) normally occur best when accompanied by consciousness. This may be an indirect link between learning and consciousness: just as medicine often aims simply to keep a patient alive until the body cures itself, consciousness directs attention until learning can take place. Conscious messages are also subject to (4) editing and debugging, where other modules can correct a perception or a motor act. Baars draws evidence from experiments on language mistakes and corrections for this point. Attentional mechanisms (5) control what will become conscious, relating conscious content to deeper goals.

A classic function for consciousness is (6) decision-making or executive function, with the global workspace allowing widespread recruitment of conscious and unconscious 'votes' for or against a decision.

The global workspace consists of a host of specialized modules being informed by the module broadcasting currently. These modules can also search for partial matches, supporting (7) an analogy-forming function.

Consciousness also relates to the idea of the self. Conscious imagery and inner speech enable (8) reflection upon and control of conscious and unconscious functioning. Finally, the accumulation of experience can be considered as (9) a self-system that maintains stability despite changing inner and outer conditions.

We see, in conclusion, that what had seemed to be a unified and coherent consciousness turns out to be many different mechanisms with distinct functions and modes of operation, having in common an access to the limited-capacity final common path of the global workspace and its memory mechanisms. Psychology has often found that

seemingly unified systems turn out to fragment into many components on closer examination. Examples are in memory, in perception, and now in consciousness itself. Baars' book promises to be a milestone in creating a theoretical framework for future consciousness research. It will also be a benchmark against which future theories are tested.

References

Crick, F. (1994) *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Charles Scribner's Sons.

Dennett, D., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, 15, 183-247.

Selfridge, O. (1959). Pandemonium: a paradigm for learning. In *Symposium on the Mechanization of Thought Processes*. London: HM Stationery Office.

Sperling, G. A. (1960) The information available in brief visual presentation. *Psychological Monographs*, 74, #498 (whole issue).