CHAPTER 4

Global workspace theory of consciousness: toward a cognitive neuroscience of human experience?

Bernard J. Baars*

The Neurosciences Institute, 10640 John Jay Hopkins Dr., San Diego, CA 92121, USA

Abstract: Global workspace (GW) theory emerged from the cognitive architecture tradition in cognitive science. Newell and co-workers were the first to show the utility of a GW or "blackboard" architecture in a distributed set of knowledge sources, which could cooperatively solve problems that no single constituent could solve alone. The empirical connection with conscious cognition was made by Baars (1988, 2002). GW theory generates explicit predictions for conscious aspects of perception, emotion, motivation, learning, working memory, voluntary control, and self systems in the brain. It has similarities to biological theories such as Neural Darwinism and dynamical theories of brain functioning. Functional brain imaging now shows that conscious cognition is distinctively associated with wide spread of cortical activity, notably toward frontoparietal and medial temporal regions. Unconscious comparison conditions tend to activate only local regions, such as visual projection areas. Frontoparietal hypometabolism is also implicated in unconscious states, including deep sleep, coma, vegetative states, epileptic loss of consciousness, and general anesthesia. These findings are consistent with the GW hypothesis, which is now favored by a number of scientists and philosophers.

Introduction

Shortly after 1900, behaviorists attempted to purge science of mentalistic concepts like consciousness, attention, mental imagery, and voluntary control. "Consciousness," wrote John B. Watson, "is nothing but the soul of theology." But as the facts accumulated over the 20th century, all the traditional ideas of James (1890) and others were found to be necessary. They were reintroduced with more testable definitions. Memory came back in the 1960s; mental imagery in the 1970s; selective attention over the last half century; and consciousness last of all, in the last decade or so.

It is broadly true that what we are conscious of, we can report with accuracy. Conscious brain events are therefore assessed by way of reportability. We now know of numerous brain events that are reportable and comparable ones that are not. This fact invites experimental testing: why are we conscious of these words at this moment, while a few seconds later they have faded, but can still be called to mind? Why is activity in visual occipitotemporal lobe neurons reportable, while visually evoked activity in parietal regions is not? Why does the thalamocortical system support conscious experiences, while the comparably large cerebellum and basal ganglia do not? How is waking consciousness impaired after brain damage? These are all testable questions. The empirical key is to treat consciousness as a controlled variable.

A growing literature now compares the brain effects of conscious and unconscious stimulation. Precise experimental comparisons allow us to ask what conscious access does "as much." Many

*Corresponding author. Tel.: 925-283-2601; Fax: 925-283-2673; E-mail: baars@nsi.edu

DOI: 10.1016/S0079-6123(05)50004-9
techniques are used for this purpose. In visual backward masking, a target picture is immediately followed by a scrambled image that does not block the optical input, but renders it unconscious (Dehaene et al., 2001). Binocular rivalry has been used for the same reason: it shows that when two competing optical streams enter the two eyes, only one consistent interpretation can be consciously perceived at any given moment (Leopold and Logothetis, 1999). Most recently, several studies have demonstrated inattentional blindness, in which paying attention to one visual flow (e.g., a bouncing basketball) blocks conscious access to another one at the very center of visual gaze (e.g., a man walking by in a gorilla suit) (Simons and Chabris, 1999). These studies generally show that unconscious stimuli still evoke local feature activity in sensory cortex. But what is the use of making something conscious if even unconscious stimuli are identified by the brain? More than a score of studies have shown that although unconscious visual words activate known word-processing regions of visual cortex, the same stimuli, when conscious, trigger widespread additional activity in frontoparietal regions (e.g., Dehaene et al., 2001).

A rich literature has arisen comparing conscious and unconscious brain events in sleep and waking, general anesthesia, epileptic states of absence, very specific damage to visual cortex, spared implicit function after brain damage, attentional control (also see Posner, this volume), visual imagery, inner speech, memory recall, and more (Crick and Koch, 2003). In state comparisons, significant progress has been made in understanding epileptic loss of consciousness (Blumenfeldt and Taylor, 2003; Blumenfeld, this volume), general anesthesia (Fiset et al., 2001; John et al., 2001; Alkire and Fiset et al., this volume) and sleep1 (Steriade, 2001; Maquet, this volume).

1At the level of cortical neurons, bursting rates do not change in deep sleep (Steriade, 2001). Rather, neurons pause together at <4 Hz between bursts. Synchronous pausing could disrupt the cumulative high-frequency interactions needed for waking functions such as perceptual continuity, immediate memory, sentence planning, motor control, and self-monitoring. It is conceivable that other unconscious states display similar neuronal mechanisms.

The global access hypothesis

The idea that consciousness has an integrative function has a long history. Global workspace (GW) theory is a cognitive architecture with an explicit role for consciousness. Such architectures have been studied in cognitive science, and have practical applications in organizing large, parallel collections of specialized processors, broadly comparable to the brain (Newell, 1994). In recent years, GW theory has been found increasingly useful by neuroscientists. The theory suggests a fleeting memory capacity that enables access between brain functions that are otherwise separate. This makes sense in a brain that is viewed as a massive parallel set of specialized processors. In such a system, coordination and control may take place by way of a central information exchange, allowing some processors — such as sensory systems in the brain — to distribute information to the system as a whole. This solution works in large-scale computer architectures, which show typical “limited capacity” behavior when information flows by way of a GW. A sizeable body of evidence suggests that consciousness is the primary agent of such a global access function in humans and other mammals (Baars, 1988, 1997, 2002). The “conscious access hypothesis” therefore implies that conscious cognition provides a gateway to numerous capacities in the brain (Fig. 1). A number of testable predictions follow from this general hypothesis (Table 1).

A theater metaphor and brain hypotheses

GW theory may be thought of as a theater of mental functioning. Consciousness in this metaphor resembles a bright spot on the stage of immediate memory, directed there by a spotlight of attention under executive guidance. Only the bright spot is conscious, while the rest of the theater is dark and unconscious. This approach leads to specific neural hypotheses. For sensory consciousness the bright spot on stage is likely to require the corresponding sensory projection areas of the cortex. Sensory consciousness in different modalities may be mutually inhibitory, within
approximately 100 ms time cycles (Baars and Franklin, 2003). Sensory cortex can be activated internally as well as externally, resulting in the "internal senses" of conscious inner speech and imagery. Once a conscious sensory content is established, it is distributed widely to a decentralized "audience" of expert networks sitting in the darkened theater, presumably using corticocortical and corticothalamic fibers. The transfer of information from conscious visual episodes to the (unconscious) hippocampal system is a clear example of such distribution of conscious information in the brain (Moscovitch, 1995). This is the primary functional role of consciousness: to allow
a theater architecture to operate in the brain, in order to integrate, provide access, and coordinate the functioning of very large numbers of specialized networks that otherwise operate autonomously. All the elements of GW theory have reasonable brain interpretations, allowing us to generate a set of specific, testable brain hypotheses about consciousness and its many roles in the brain. Some of these ideas have now received considerable empirical support (Baars, 2002; Baars et al., 2003).

The theory has been implemented in computational and neural net models and bears a family resemblance to Neural Darwinist models (Edelman, 2003). Franklin and colleagues have implemented GW theory in large-scale computer agents, to test its functionality in complex practical tasks (Franklin, 2001). IDA (for “intelligent distributed agent”), the current implementation of the extended GW architecture directed by Franklin, is designed to handle a very complex artificial intelligence task normally handled by trained human beings (also see Aleksander on machine consciousness in this volume). The particular domain in this case is interaction between U.S. Navy personnel experts and sailors who move from job to job. IDA negotiates with sailors via e-mail, and is able to combine numerous regulations, sailors’ preferences, time, location and travel considerations into human-level performance.

While it has components roughly corresponding to human perception, memory, and action control, the heart of the system is a GW architecture that allows the content or meanings of the messages to be widely distributed, so that specialized programs called “codelets” can respond with solutions to centrally posed problems. Franklin writes that “The fleshed out global workspace theory is yielding hopefully testable hypotheses about human cognition. The architectures and mechanisms that underlie consciousness and intelligence in humans can be expected to yield information agents that learn continuously, adapt readily to dynamic environments, and behave flexibly and intelligently when faced with novel and unexpected situations.” (http://csrg.cs.memphis.edu/). Similar architectures have been applied to difficult problems like speech recognition. While such auton-

omous agent simulations do not prove that GW architectures exist in the brain, they give an existence proof of their functionality. It is worth noting that few integrative theories of mind or brain show functional utility in applied settings.

**Sensory consciousness as a test case**

Visual consciousness has been studied in depth, and there is accepted evidence that visual features that become conscious are identified by the brain in the ventral stream of visual cortex. There, feature-sensitive cells support visual experiences of light, color, contrast, motion, retinal size, location, and object identity; small lesions can selectively abolish those conscious properties without affecting other aspects of conscious vision (Zeki, 2001; Naccache, in this volume).

However, to recollect the experience of a human face, we need the hippocampal system. To respond to it emotionally, neurons in amygdala may be activated. But hippocampus and amygdala do not seem to support conscious contents directly (Moscovitch, 2001). Thus, the ventral visual stream, which is needed for specific conscious contents, seems to influence regions that are not.

Dehaene and colleagues have shown that backward-masked visual words evoked brain activity confined to the well-known visual word recognition areas of cortex (Dehaene et al., 2001). Identical conscious words triggered higher levels of activity in these areas, but more importantly, they also evoked far more widely distributed activity in parietal and prefrontal cortex. That result has now been replicated more than a dozen times, using different brain imaging techniques and different methods for comparing conscious and unconscious input. Such methods have included binocular rivalry (Sheinberg and Logothetis, 1997), inattentive blindness (Rees et al., 1999), neglect and its extinction (Rees et al., 2002), and different sense modalities, such as audition (Portas et al., 2000), pain perception (Rosen et al., 1996), and sensorimotor tasks (Haier et al., 1992; Raichle et al., 1994). In all cases, conscious sensory input evoked wider and more intense brain activity than identical unconscious input.
Complementary findings come from studies of unconscious states. In deep sleep, auditory stimulation activates only primary auditory cortex (Portas et al., 2000). In vegetative states following brain injury, stimuli that are ordinarily loud or painful activate only the primary sensory cortices (Laureys et al., 2000, 2002). Waking consciousness is apparently needed for widespread of input-driven activation to occur. These findings support the general notion that conscious stimuli mobilize large areas of cortex, presumably to distribute information about the input.

**Inner speech, imagery, and working memory**

Both auditory and visual consciousness can be activated endogenously. Inner speech is a particularly important source of conscious auditory-phonemic events, and visual imagery is useful for spatial memory and problem-solving. The areas of the left hemisphere involved in outer speech are now known to be involved in inner speech as well (Paulesu et al., 1993). Likewise, mental imagery is known to involve visual cortex (Kosslyn et al., 2001). Internally generated somatosensory imagery may reflect emotional and motivational processes, including feelings of psychological pain, pleasure, hope, fear, sadness, etc. (Damasio, 2003). Such internal sensations may communicate to other parts of the brain via global distribution or activation. Prefrontal executive systems may sometimes control motor activities by evoking motivational imagery, broadcast from the visual cortex, to activate relevant parts of motor cortex. Parts of the brain that play a role in emotion may also be triggered by global distribution of conscious contents from sensory cortices and insular cortex. For example, the amygdala appears necessary to recognize visual facial expressions of fear and anger. Thus, many cortical regions work together to transform goals and emotions into actions (Baars, 1988).

**The attentional spotlight**

The sensory “bright spot” of consciousness involves a selective attention system, the ability of the theater spotlight to shine on different actors on the stage. Like other behaviors like breathing and smiling, attention operates under dual control, voluntary, and involuntary. Voluntary attentional selection requires frontal executive cortex, while automatic selection is influenced by many areas, including the brain stem, pain systems, insular cortex, and emotional centers like the amygdala and peri-aqueductal grey (Panksepp, 1998). Presumably, these automatic attentional systems that allow significant stimuli to “break through” into consciousness, as when a subject’s name is sounded in an otherwise unconscious auditory source.

**Context and the first-person perspective**

When we step from a tossing sailboat onto solid ground, the horizon can be seen to wobble. On an airplane flight at night passengers can see the cabin tilting on approach to landing, although they are receiving no optical cues about the direction of the plane. In those cases unconscious vestibular signals shape conscious vision. There are numerous examples in which unconscious brain activities can shape conscious ones, and vice versa. These unconscious influences on conscious events are called “contexts” in GW theory (Fig. 1). Any conscious sensory event requires the interaction of sensory analyzers and contextual systems. In vision, sensory contents seem to be produced by the ventral visual pathway, while contextual systems in the dorsal pathway define a spatial domain within which the sensory event is defined. Parietal cortex is known to include allocentric and egocentric spatial maps, which are not themselves objects of consciousness, but which are required to shape every conscious visual event. There is a difference between the disorders of content systems like the ventral stream, compared to damaged context systems. In the case of ventral stream lesions, the subject can generally notice a missing part of normal experience; but for damage to context, the brain basis of expectations is itself damaged, so that one no longer knows what to expect, and hence what is missing. This may be why parietal neglect is so often accompanied by a striking loss of knowledge about one’s body space.
(Bisiach and Geminiani, 1991). Patients suffering from right parietal neglect can have disturbing alien experiences of their own bodies, especially of the left arm and leg. Such patients sometimes believe that their left leg belongs to someone else, often a relative, and can desperately try to throw it out of bed. Thus, parietal regions seem to shape contextually both the experience of the visual world and of one’s own body. Notice that neglect patients still experience their alien limbs as conscious visual objects (a ventral stream function); they are just disowned. Such specific loss of contextual body information is not accompanied by a loss of general intelligence or knowledge.

Vogele and Fink (2003) suggest that parietal cortex is involved in the first-person perspective, the viewpoint of the observing self. When subjects are asked to adopt the visual perspective of another person, parietal cortex became differentially active.

**Self-systems**

Activation by of visual object regions by the sight of a coffee cup may not be enough to generate subjective consciousness of the cup. The activated visual information may need to be conveyed to executive or self-systems, which serve to maintain constancy of an inner framework across perceptual situations. When we walk from room to room in a building, we must maintain a complex and multileveled organization that can be viewed in GW theory as a higher-level context. Major goals, for example, do not change when we walk from room to room, but conscious perceptual experiences do. Gazzaniga (1996) has found a number of conditions under which split-brain patients encounter conflict between right and left hemisphere executive and perceptual functions. He has proposed the existence of a “narrative self” in the left frontal cortex, based on split-brain patients who are clearly using speech output in the left hemisphere to talk to themselves, sometimes trying to force the right hemisphere to obey its commands. When that proves impossible, the left hemisphere will often rationalize the sequence of events so as to repair its understanding of the interhemispheric conflict. Analogous repairs of reality are observed in other forms of brain damage, such as neglect. They also commonly occur whenever humans are confronted with major, unexpected life changes. The left-hemisphere narrative interpreter may be considered as a higher-level context system that maintains expectations and intentions across many specific situations. Although the inner narrative itself is conscious, it is shaped by unconscious contextual influences.

If we consider Gazzaniga’s narrative interpreter of the dominant hemisphere to be one kind of self-system in the brain, it must receive its own flow of sensory input. Visual input from one-half of the field may be integrated in one visual hemicortex, as described above, under retinotopic control from area V1. But once it comes together in late visual cortex (presumably in inferotemporal object regions), it needs to be conveyed to frontal areas on the dominant hemisphere, in order to inform the narrative interpreter of the current state of perceptual affairs. The left prefrontal self system then applies a host of criteria to the input, such as “did I intend this result? Is it consistent with my current and long-term goals? If not, can I reinterpret it to make sense in my running account of reality?” It is possible that the right hemisphere has a parallel system that does not speak but that may be better able to deal with anomalies via irony, jokes, and other emotionally useful strategies. The evidence appears to be good that the isolated right prefrontal cortex can understand such figurative uses of language, while the left does not. Full consciousness may not exist without the participation of such prefrontal self systems.

**Relevance to waking, sleeping, coma, and general anesthesia**

Metabolic activity in the conscious resting state is not uniformly distributed. Raichle et al. (2001) reported that mesiofrontal and medial parietal areas, encompassing precuneus and adjacent posterior cingulate cortex, can be posited as a tonically active region of the brain that may continuously gather information about the world around, and possibly within, us. It would appear
to be a default activity of the brain. Mazoyer et al. (2001) also found high prefrontal metabolism during rest. We will see that these regions show markedly lower metabolism in unconscious states. Laureys (1999a, b, 2000) and Baars et al. (2003) list the following features of four unconscious states, that are causally very different from each other: deep sleep, coma/vegetative states, epileptic loss of consciousness, and general anesthesia under various agents. Surprisingly, despite their very different mechanisms they share major common features. These include: (i) widely synchronized slow waveforms that take the place of the fast and flexible interactions needed for conscious functions; (ii) frontoparietal hypometabolism; (iii) widely blocked functional connectivity, both corticocortical and thalamocortical; and (iv) behavioral unconsciousness, including unresponsiveness to normally conscious stimuli. Fig. 2 shows marked hypofunction in the four unconscious states compared with conscious controls, precisely where we might expect it: in frontoparietal regions.

In a related study, John and co-workers showed marked quantitative electroencephalogram (EEG) changes between conscious, anesthetic, and pre-anesthetic (conscious) states (John et al., 2000). At loss of consciousness, gamma power decreased while lower frequency bands increased in power, especially in frontal leads. Loss of consciousness was accompanied by a significant drop in coherence between homologous areas of the two hemispheres, and between posterior and anterior regions of each hemisphere. However, there was hypersynchronous activity within anterior regions. The same basic changes occurred across all six anesthetics, and reversed when patients regained consciousness (see John, in this volume).

From the viewpoint of globalist theories, the most readily interpretable finding is the coherence drop in the gamma range after anesthetic loss of consciousness. It suggests a loss of coordination between frontal and posterior cortex, and between homologous regions of the two hemispheres. The authors also suggest that the anteriorization of low frequencies “must exert a profound inhibitory influence on cooperative processes within (frontal) neuronal populations. This functional system then becomes dedifferentiated and disorganized” (p. 180). Finally, the decoupling of the posterior cortex with anterior regions suggests “a blockade of perception” (p. 180). These phenomena appear to be consistent with the GW notion that widespread activation of nonsensory regions is required for sensory consciousness.

The role of frontoparietal regions in conscious contents and states

Could it be that brain regions that underlie the contextual functions of Fig. 1 involve frontal and parietal regions? In everyday language, the “observing self” may be disabled when those regions are dysfunctional and long-range functional connectivity is impaired. Frontoparietal association areas have many functions, but several lines of evidence suggest that they could have a special relationship with consciousness, even though they do not support the sensory contents of conscious experience directly. (i) Conscious stimulation in the waking state leads to frontoparietal activation, but unconscious input does not; (ii) in unconscious states, sensory stimulation activates only sensory cortex, but not frontoparietal regions; (iii) the conscious resting state shows high frontoparietal metabolism compared with outward-directed cognitive tasks; and (iv) four causally very different unconscious states show marked functional decrements in the same areas. Although alternative hypotheses must be considered, it seems reasonable to suggest that “self” systems supported by these regions could be disabled in unconscious states. From the viewpoint of the narrative
observer, this would be experienced as subjective loss of access to the conscious world. Unconscious states might not necessarily block the objects of consciousness; rather, the observing subject might not be at home. Which in turn may serve coordination and control. Consciousness is the gateway to the brain.

Uncited References

Fiset et al. (1999); Frackowiak et al. (2004); Franklin (2000); Freeman (2003); Tononi and Edelman (1998).

Acknowledgments

I gratefully acknowledge support from the Neurosciences Institute and the Neurosciences Research Foundation (10640 John Jay Hopkins Drive, San Diego, CA 94549, USA). I thank Steven Laureys, Gerald M. Edelman, Stan Franklin, Bjorn Merker, Anil Seth, Douglas Nitz, E. Roy John, and Walter Freeman for helpful discussions.
References


**Queries and / or remarks**

<table>
<thead>
<tr>
<th>Query No</th>
<th>Details required</th>
<th>Author's response</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQ1</td>
<td>Please confirm the change in the year of publication.</td>
<td></td>
</tr>
<tr>
<td>AQ2</td>
<td>Please provide the reference.</td>
<td></td>
</tr>
<tr>
<td>AQ3</td>
<td>Please confirm the change in the year of publication.</td>
<td></td>
</tr>
<tr>
<td>AQ4</td>
<td>Please provide the reference.</td>
<td></td>
</tr>
<tr>
<td>AQ5</td>
<td>Please confirm the change in the year of publication.</td>
<td></td>
</tr>
<tr>
<td>AQ6</td>
<td>Please provide the reference.</td>
<td></td>
</tr>
<tr>
<td>AQ7</td>
<td>Please check the citation of 'notes' 1 and 2 in the text.</td>
<td></td>
</tr>
<tr>
<td>AQ8</td>
<td>Please check the insertion of initials.</td>
<td></td>
</tr>
<tr>
<td>AQ9</td>
<td>Please provide the place of publication.</td>
<td></td>
</tr>
<tr>
<td>AQ10</td>
<td>Please provide the names of remaining authors where 'et al' is referred in the ref. list.</td>
<td></td>
</tr>
<tr>
<td>UC. ref.</td>
<td>If uncited references are appear under the section 'uncited references,' please cite them at relevant places in the text.</td>
<td></td>
</tr>
</tbody>
</table>