The Cambridge Handbook of Consciousness

Edited by

Philip David Zelazo, Morris Moscovitch and Evan Thompson

University of Toronto

CAMBRIDGE
UNIVERSITY PRESS

2007

CHAPTER 8
Cognitive Theories of Consciousness

Katharine McGovern and Bernard J. Baars

Abstract

Current cognitive theories of consciousness focus on a few common themes, such as the limited capacity of conscious contents under input competition; the wide access enabled by conscious events to sensation, memory, problem-solving capacities, and action control; the relation between conscious contents and working memory; and the differences between implicit and explicit cognition in learning, retrieval, and other cognitive functions. The evidentiary base is large. A unifying principle in the midst of these diverse empirical findings is to treat consciousness as an experimental variable and, then, to look for general capacities that distinguish conscious and unconscious mental functioning. In this chapter, we discuss three classes of theories: information-processing theories that build on modular elements, network theories that focus on the distributed access of conscious processing, and globalist theories that combine aspects of these two. An emerging consensus suggests that conscious cognition is a global aspect of human brain functioning. A specific conscious content, like the sight of a coffee cup, is crucially dependent on local regions of visual cortex. But, by itself, local cortical activity is not conscious. Rather, the conscious experience of a coffee cup requires both local and widespread cortical activity.

Introduction

When consciousness became a scientifically respectable topic again in the 1980s, it was tackled in a number of different scholarly disciplines – psychology, philosophy, neuroscience, linguistics, medicine, and others. By the late 1990s, considerable interdisciplinary cooperation evolved in consciousness studies, spurred by the biennial Tucson Conferences and the birth of two new scholarly journals, *Consciousness and Cognition* and the *Journal of Consciousness Studies*. The domain of consciousness studies originated in separate disciplines, but has since become cross-disciplinary. Thus, a number of early theories of consciousness can justifiably be called purely cognitive theories of consciousness, whereas most recent theories are neurocognitive hybrids – depending on evidence from the brain as well as behavior. In this chapter, we have, for the most part, restricted discussion to cognitive or functional models of consciousness with less reference to the burgeoning neuroscientific evidence that increasingly supports the globalist position that we develop here.

Operationally Defining Consciousness

Cognitive Methods That Treat Consciousness as a Variable

There is a curious asymmetry between the assessment of conscious and unconscious processes. Obtaining verifiable experiential reports works very nicely for specifying conscious representations, but unconscious ones are much more slippery. In many cases of apparently unconscious processes, such as all the things the reader is *not* paying attention to at this moment, it could be that the "unconscious" representations may be momentarily conscious, but so quickly or vaguely that we cannot recall them even a fraction of a second later. Or suppose people cannot report a word shown for a few milliseconds: Does this mean that they are truly unconscious of it? Such questions continue to lead to controversy today. William James understood this problem very well and suggested, in fact, that there were no unconscious psychological processes at all (1890, p. 162ff.). This has been called the "zero point" problem (Baars, 1988). It should be emphasized, however, that problems with defining a zero point do not prevent scientists from studying phenomena as variables. Even today, the precise nature of zero temperature points, such as the freezing point of water, continues to lead to debate. But physicists have done extremely productive work on thermodynamics for centuries. Zero points are not the sole criterion for useful empirical variables.

The discovery that something we take for granted as a constant can be treated as a variable has led to scientific advances before. In the late 1600s, contemporaries of Isaac Newton were frustrated in their attempts to understand gravity. One key to Newton's great achievement was to imagine the presence *and the absence* of gravity, thus permitting gravity to be treated as a variable. In the same way, a breakthrough in the scientific study of consciousness occurred when psychologists began to understand that consciousness can be treated as a variable. That is, behavioral outcomes can be observed when conscious cognitions are present and when they are absent. The process of generalizing across these observations has been called *contrastive analysis* (explained below).

Beginning in the 1980s, a number of experimental methods gained currency as means of studying comparable conscious and non-conscious processes. In much of cognitive science and neuroscience today, the existence of unconscious cognitive processes, often comparable to conscious ones, is taken for granted. Table 8.1 highlights methods that have produced behavioral data relevant to the study of consciousness.

Working Definitions of "Conscious" and "Unconscious"

In the history of science, formal definitions for concepts like "heat" and "gene" tend to come quite late, often centuries after adequate operational definitions are developed. The same point may apply to conscious cognition. Although there is ongoing debate about what consciousness "really" is, there has long been a scientific consensus on its observable index of verbal report. This index can be generalized to any other kind of voluntary response, such as pressing a button or even voluntary eye movements in "locked-in" neurological patients. Experiential reports can be analyzed with sophisticated methods, such as process dissociation and signal detection. Thus, empirically, it is not difficult to assess conscious events in humans with intact brains, given good experimental conditions. We propose the

Table 8.1. Empirical methods used in the study of conscious and unconscious processes

Class of Methods	*Experimental Paradigm*	*Outcome*
Divided attention	Dichotic listening	Two dense streams of speech are offered to the two ears, and only one stream at a time can receive conscious word-level processing. Evidence suggests that the unconscious stream continues to receive some processing.
	Selective ("double exposure") viewing	When two overlaid movies are viewed, only one is perceived consciously.
	Inattentional blindness	Aspects of visual scenes to which attention is not directed are not consciously perceived; attended aspects of the same scenes are perceived.
	Binocular rivalry/dichoptic viewing (including flash suppression)	Presenting separate visual scenes to each eye; only one scene reaches consciousness, but the unconscious scene receives low-level processing.
Dual task paradigms	Driving and talking on a cell phone Rehearsing words and doing word verification	To the extent that tasks require conscious initiation and direction, they compete and degrade the performance of each other; once automatized, multiple tasks interfere less.
Priming	Supraliminal and subliminal priming Priming of one interpretation of ambiguous words or pictures	When a "prime" stimulus is presented prior to a "target" stimulus, response to the "target" is influenced by the currently unconscious nature and meaning of the "prime." Supraliminal priming generally results in a more robust effect.
Visual backward masking		When supra-threshold visual stimuli are followed immediately by visual masking stimuli (visual noise), the original stimuli are not consciously perceived, though they are locally registered in early visual cortex.
Implicit learning	Miniature grammar learning	Consciously perceived stimuli give rise to knowledge structures that are not available to consciousness.
Process dissociation and ironic effects		Participants are told to exclude certain memorized items from memory reports; if those items nevertheless appear, they are assumed to be products of non-conscious processing.
Fixedness, decontextualization, and being blind to the obvious (related to availability)	Problem-solving tasks, functional fixedness tasks (Duncker), chess playing, garden path sentences, highly automatized actions under novel conditions	Set effects in problem solving can exclude otherwise obvious conclusions from consciousness. "Breaking set" can lead to recovery of those conclusions in consciousness.

Table 8.2. Contrastive analysis in perception and imagery

Conscious Events	*Comparable Unconscious Events*
1. Perceived stimuli	1. Processing of stimuli lacking in intensity or duration, or centrally masked stimuli
	2. Preperceptual processing of stimuli
	3. Habituated or automatic stimulus processing
	4. Unaccessed versions of ambiguous stimuli/words
	5. Contexts of interpretation for percepts and concepts
	6. Unattended streams of perceptual input (all modalities)
	7. Implicit expectations about stimuli
	8. Parafoveal guidance of eye movements in reading
	9. Stimulus processing under general anesthesia
10. Images in all sense modalities	10. Unretrieved images in memory
11. a. Newly generated visual images b. Automatic images that encounter some difficulty	11. Automatized visual images
12. Inner speech: words currently rehearsed in working memory	12. Inner speech, not currently rehearsed in working memory
13. Fleetingly conscious phrases and belief statements	13. Automatized inner speech; the "jingle channel"
14. Visual search based on conjoined features	14. Visual search based on single features
15. Retrieval by recall	15. Retrieval by recognition
16. Explicit knowledge	16. Implicit knowledge

following as de facto operational definitions of conscious and unconscious that are already in very wide experimental use in perception, psychophysics, memory, imagery, and the like.

We can say that mental processes are *conscious* if they

(a) are claimed by people to be conscious; and
(b) can be reported and acted upon,
(c) with verifiable accuracy,
(d) under optimal reporting conditions (e.g., with minimum delay between the event and the report, freedom from distraction, and the like).

Conversely, mental events can be defined as *unconscious* for practical purposes if

(a) their presence can be verified (through facilitation of other observable tasks, for example); although
(b) they are not claimed to be conscious;
(c) and they cannot be *voluntarily* reported, operated upon, or avoided,
(d) even under optimal reporting conditions.

The Method of Contrastive Analysis

Using the logic of experimental research, consciousness can be treated as a controlled variable; then, measures of cognitive functioning and neural activity can be compared under two levels of the independent variable – consciousness-present and consciousness-absent. If there is no clearly unconscious comparison condition, a low-level conscious condition may be used, as in drowsiness or stimuli in background noise. The point, of course, is to have at least two quantitatively different levels for comparison.

DATA FROM CONTRASTIVE ANALYSIS

Examples of conscious versus non-conscious contrasts from studies of perception, imagery, memory, and attention appear in Table 8.2. In the left column, conscious mental events are listed; on the right are

Table 8.3. Capability contrasts between comparable conscious and non-conscious processes

Conscious Processes	*Unconscious Processes*
1. Are computationally inefficient with • Many errors • Relatively low speed • Mutual interference between conscious processes.	1. Are very efficient in routine tasks with • Few errors • High speed • Little mutual interference.
2. Have a great range of contents • Great ability to relate different conscious contents to each other • Great ability to relate conscious events to unconscious contexts • *Flexible*	2. Taken individually, unconscious processes have a *limited range* of contents • Each routine process is relatively isolated and autonomous • Each routine process is relatively context-free, operates in a range of contexts • *Fixed pattern*
3. • have high *internal consistency* at any single moment • have *seriality* over time • have *limited processing capacity*	3. The set of routine, unconscious processes, taken together, is: • *diverse*, • can operate *concurrently* • have great *processing capacity*
4. The *clearest* conscious contents are *perceptual or quasi-perceptual* (e.g., imagery, inner speech, and internally generated bodily feelings)	4. Unconscious processes are involved in *all* mental tasks, not limited to perception and imagery, but including memory, knowledge representation and access, skill learning, problem-solving, action control, etc.
5. Are associated with *voluntary* actions	5. Are associated with *non-voluntary* actions

corresponding non-conscious processes. Theoretically, we are interested in finding out what is common in conscious processing across all these cases.

Capability Contrasts

The difference in mental and neural functioning between consciousness-present and consciousness-absent processing – taken across many experimental contexts – reveals stable characteristics attributable to consciousness. *Conscious processes* are phenomenally serial, internally consistent, unitary at any moment, and limited in capacity. *Non-conscious mental processes* are functionally concurrent, often highly differentiated from each other, and relatively unlimited in capacity, when taken together. Table 8.3 summaries these general conclusions.

These empirical contrasts in the capabilities of conscious and unconscious mental processes can become the criteria against which models of consciousness can be evaluated. Any adequate theory of consciousness would need to account for these observed differences in functioning. Thus, we have a way of judging the explanatory adequacy of proposals concerning the nature and functioning of consciousness. We can keep these capability contrasts in mind as we review contemporary cognitive models of consciousness.

Given the tight constraints that appear repeatedly in studies of conscious processing – that is, limited capacity, seriality, and internal consistency requirements – we might ask, Why? Would it not be adaptive to do *several* conscious things at the same time? Certainly human ancestors might have benefited from being able to gather food, be alert for predators, and keep an eye on their offspring simultaneously; modern humans could benefit from being able to drive their cars, talk on cell phones, and put on lipstick without mutual interference. Yet these tasks compete when they require consciousness, so that only one can be done well at any given moment. The question then is, Why are conscious functions so limited in a neuropsychological architecture that is so large and complex?

Functions of Consciousness in the Architecture of Cognition

A Note about Architectures

The metaphor of "cognitive architectures" dates to the 1970s when cognitive psychologists created information-processing models of mental processes. In many of these models, different mental functions, such as memory, language, attention, and sensory processes, were represented as modules, or sets of modules, within a larger information-processing system. The functional layout and the interactions of the parts of the system came to be called the cognitive architecture. We have adopted this terminology here to capture the idea that consciousness operates within a larger neuropsychological system that has many constituents interacting in complex ways.

Consciousness Serves Many Functions

William James believed that "[t]he study... of the distribution of consciousness shows it to be exactly such as we might expect in an organ added for the sake of steering a nervous system grown too complex to regulate itself (1890, p. 141)." More recently, Baars (1988) identified eight psychological functions of consciousness, which are defined in Table 8.4.

Note that each proposed function of consciousness is served through an interplay of conscious and unconscious processes. It has been argued that consciousness fulfills all eight functions by providing *access* or priority entrance into various subparts of the cognitive system (Baars, 2002). For example, the error-detection function can be accomplished only when information about an impending or actual error, which cannot be handled by "canned" automatisms, can gain access to consciousness. Subsequently, editing occurs when this conscious information is "broadcast" or distributed to other parts of the system that are capable of acting to recognize and correct it. Consciousness functions as the central distributor of information, which is used by subparts of the cognitive system or architecture.

Consciousness Creates Access

A strong case can be made that we can create access to any part of the brain by way of conscious input. For example, to gain voluntary control over alpha waves in the occipital cortex we merely sound a tone or turn on a light when alpha is detected in the EEG, and shortly the subject will be able to increase the power of alpha at will. To control a single spinal motor unit we merely pick up its electrical activity and play it back over headphones; in a half-hour, subjects have been able to play drum rolls on single motor units. Biofeedback control over single neurons and whole populations of neurons anywhere in the brain is well established (Basmajian, 1979). Consciousness of the feedback signal seems to be a necessary condition to establish control, though the motor neural activities themselves remain entirely unconscious. It is as if mere consciousness of results creates access to unconscious neuronal systems that are normally quite autonomous.

Psychological evidence leads to similar conclusions. The recognition vocabulary of educated English speakers contains about 100,000 words. Although we do not use all of them in everyday speech, we can understand each word as soon as it is presented in a sentence that makes sense. Yet each individual word is already quite complex. The *Oxford English Dictionary* devotes 75,000 words to the many different meanings of the word "set." Yet all we do as humans to access these complex unconscious bodies of knowledge is to become conscious of a target word. It seems that understanding language demands the gateway of consciousness. This is another case of the general principle that consciousness of stimuli creates widespread access to unconscious sources of knowledge, such as the mental lexicon, meaning, and grammar.

Table 8.4. Explaining the psychological functions of consciousness

Function of Consciousness	*Function Explained*
1. Definition and context-setting	By relating input to its contextual conditions, consciousness defines a stimulus and removes ambiguities in its perception and understanding.
2. Adaptation and learning	The more novelty and unpredictability to which the psychological system must adapt, the greater the conscious involvement required for successful problem solving and learning.
3. Prioritizing and access control	Attentional mechanisms exercise selective control over what will become conscious by relating input to unconscious goal contexts. By consciously relating an event or circumstance to higher-level goals, we can raise its access priority, making it conscious more often and therefore increasing the chances of successful adaptation to it.
4. Recruitment and control of thought and action	Conscious goals can recruit subgoals and behavior systems to organize and carry out flexible, voluntary action.
5. Decision-making and executive function	Consciousness creates access to multiple knowledge sources within the psychological system. When automatic systems cannot resolve some choice point in the flow of action, making it conscious helps recruit knowledge sources that are able to help make the decision; in case of indecision, making the goal conscious allows widespread recruitment of conscious and unconscious sources acting for and against the goal.
6. Error detection and editing	Conscious goals and plans are monitored by unconscious rule systems that will act to interrupt execution if errors are detected. Though we often become aware of making an error in a general way, the detailed description of what makes an error an error is almost always unconscious.
7. Reflection and self-monitoring	Through conscious inner speech and imagery we can reflect upon and to some extent control and plan our conscious and unconscious functioning.
8. Optimizing the tradeoff between organization and flexibility	Automatized, "canned" responses are highly adaptive in predictable circumstances. However, in unpredictable environments, the capacity of consciousness to recruit and reconfigure specialized knowledge sources is indispensable in allowing flexible responding.

Or consider autobiographical memory. The size of long-term episodic memory is unknown, but we do know that simply by paying attention to as many as 10,000 distinct pictures over several days without attempting to memorize them, we can spontaneously recognize more than 90% a week later (Standing, 1973). Remarkable results like this are common when we use recognition probes, merely asking people to choose between known and new pictures. Recognition probes apparently work so well because they re-present the original conscious experience of each picture in its entirety. Here the brain does a marvelous job of memory search, with little effort. It seems that humans create memories of the stream of input merely by paying attention, but because we are always paying attention to something, in every waking moment, this suggests that autobiographical memory may be very large indeed. Once again we have a vast unconscious domain, and we gain access to it using consciousness. Mere consciousness of some event helps store a recognizable memory of it, and when we experience it again, we can distinguish it accurately from millions of other experiences.

The ability to access unconscious processes via consciousness applies also to the vast number of unconscious automatisms that can be triggered by conscious events, including eye movements evoked by conscious visual motion, the spontaneous inner speech that often accompanies reading, the hundreds of muscle groups that control the vocal tract, and those that coordinate and control other skeletal muscles. None of these automatic neuronal mechanisms are conscious in any detail under normal circumstances. Yet they are triggered by conscious events. This triggering function is hampered when the conscious input is degraded by distraction, fatigue, somnolence, sedation, or low signal fidelity.

Consciousness seems to be needed to access at least four great bodies of unconscious knowledge: the lexicon of natural language, autobiographical memory, the automatic routines that control actions, and even the detailed firing of neurons and neuronal populations, as shown in biofeedback training. Consciousness seems to create access to vast unconscious domains of expert knowledge and skill.

Survey of Cognitive Theories of Consciousness

Overview

In the survey that follows, cognitive theories of consciousness are organized into three broad categories based on the architectural characteristics of the models. The first group consists of examples of *information-processing theories that emphasize modular processes*: Johnson-Laird's Operating System Model of Consciousness, Schacter's Model of Dissociable Interactions and Conscious Experience (DICE), Shallice's Supervisory System, Baddeley's Early and Later Models of Working Memory, and Schneider and Pimm-Smith's Message-Aware Control Mechanism. The second group includes *network theories that explain consciousness as patterns of system-wide activity*: Pribram's Holonomic Theory, Tononi and Edelman's Dynamic Core Hypothesis, and Walter Freeman's Dynamical Systems Approach. The third group includes globalist models that combine aspects of information-processing theories and network theories: Baars' Global Workspace Theory, Franklin's IDA as an implementation of GW theory, and Dehaene's Global Neuronal Network Theory. Theories have been selected that represent the recent history of cognitive modeling of consciousness from the 1970s forward and that account in some way for the evidence described above concerning the capability contrasts of conscious and unconscious processes.

Information-Processing Theories That Emphasize Modular Processes: Consciousness Depends on a Kind of Processing

Theories in this group emphasize the information-processing and action control aspects of the cognitive architecture. They tend to explain consciousness in terms of "flow of control" or flow of information among specialist modules.

JOHNSON-LAIRD'S OPERATING SYSTEM MODEL OF CONSCIOUSNESS

Johnson-Laird's (1988) operating system model of consciousness emphasizes its role in controlling mental events, such as directing attention, planning and triggering action and thought, and engaging in purposeful self-reflection. Johnson-Laird proposes that the cognitive architecture performs parallel processing in a system dominated by a control hierarchy. His system involves a collection of largely independent processors (finite state automata) that cannot modify each other but that can receive messages from each other; each initiates computation when it receives appropriate input from any source. Each passes messages up through the hierarchy to the operating system that sets goals for the subsystems. The operating system does not have access to the detailed operations of the subsystems – it receives

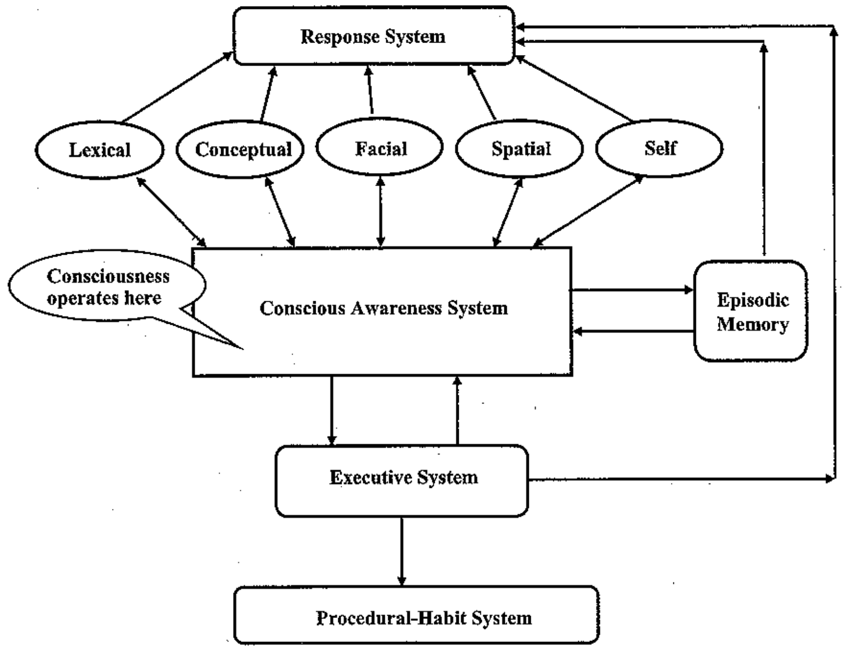

Figure 8.1. Schacter's Dissociable Interactions and Conscious Experience (DICE) Model. (Redrawn from Schacter, 1990). Phenomenal awareness depends on intact connections between the conscious awareness system and the individual knowledge modules or episodic memory. The conscious awareness system is the gateway to the executive system, which initiates voluntary action.

only their output. Likewise, the operating system does not need to specify the details of the actions it transmits to the processors – they take in the overall goal, abstractly specified, and elaborate it in terms of their own capabilities.

In this model, conscious contents reside in the operating system or its working memory. Johnson-Laird believes his model can account for aspects of action control, self-reflection, intentional decision making, and other metacognitive abilities.

SCHACTER'S MODEL OF DISSOCIABLE INTERACTIONS AND CONSCIOUS EXPERIENCE (DICE)

Accumulating evidence regarding the neuropsychological disconnections of processing from consciousness, particularly implicit memory and anosagnosia, led Schacter (1990) to propose his Dissociable Interactions and Conscious Experience (DICE) model (see Figure 8.1): "The basic idea motivating the DICE model... is that the processes that mediate conscious identification and recognition – that is, phenomenal awareness in different domains – should be sharply distinguished from modular systems that operate on linguistic, perceptual, and other kinds of information" (pp. 160–161, 1990).

Like Johnson-Laird's model, Schacter's DICE model assumes independent memory modules and a lack of conscious access to details of skilled/procedural knowledge. It is primarily designed to account for memory dissociations in normally functioning and damaged brains. There are two main observations of interest. First, with the exception of coma and stupor patients, failures of awareness in neuropsychological cases are usually restricted to the domain of the impairment; these patients do not have difficulty generally in gaining conscious access to other knowledge sources. Amnesic patients, for example, do not necessarily have trouble reading words, whereas alexic individuals do not necessarily have memory problems.

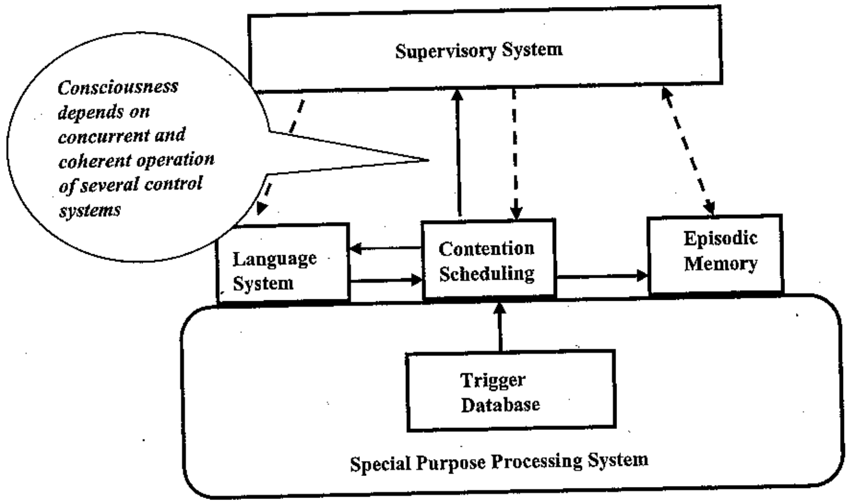

Figure 8.2. Shallice's Supervisory System Model of Conscious Processing. Solid arrows represent obligatory communications; dashed arrows represent optional communications. (Drawn from Shallice, 1988.)

Fifth, more recently an EPISODIC MEMORY component containing event-specific traces has been added to the set of control processes.

Thus, the supervisory system, contention scheduling, the language system, and episodic memory all serve higher-level or control functions in the system. As a first approximation, one of these controllers or several together might be taken as the "conscious part" of the system. However, as Shallice points out, consciousness cannot reside in any of these control systems taken individually. No single system is either necessary or sufficient to account for conscious events. Consciousness remains even when one of these control systems is damaged or disabled, and the individual control systems can all operate autonomously and unconsciously. Instead, Shallice suggests that consciousness may arise on those occasions where there is concurrent and coherent operation of several control systems on representations of a single activity. In this event, the contents of consciousness would correspond to the flow of information between the control systems and the flow of information and control from the control systems to the rest of the cognitive system.

Shallice's (1988) model aims primarily to "reflect the phenomenological distinctions between willed and ideomotor action" (p. 319). Shallice identifies consciousness with the control of coherent action subsystems and the emphasis on the flow of information among the subsystems.

BADDELEY'S EARLY AND LATER MODELS OF WORKING MEMORY: 1974 TO 2000

Working memory is a functional account of the workings of temporary memory (as distinct from long-term memory). Baddeley and Hitch (1974) first proposed their multicomponent model of working memory (WM) as an advance over single-store models, such as Short-Term Memory (STM; Atkinson & Shiffrin, 1968). The original WM model was simple, composed of a central executive with two subsystems, the phonological loop and the visuospatial sketchpad. WM was designed to account for short duration, modality-specific, capacity-limited processing of mnemonic information. It combined the storage capacity of the older STM model with an executive process that could "juggle" information between two slave systems and to and from long-term memory. The evolving model of WM has been successful in accounting for behavioral and neurological findings in normal participants and in neuropsychological patients. From the beginning, working memory,

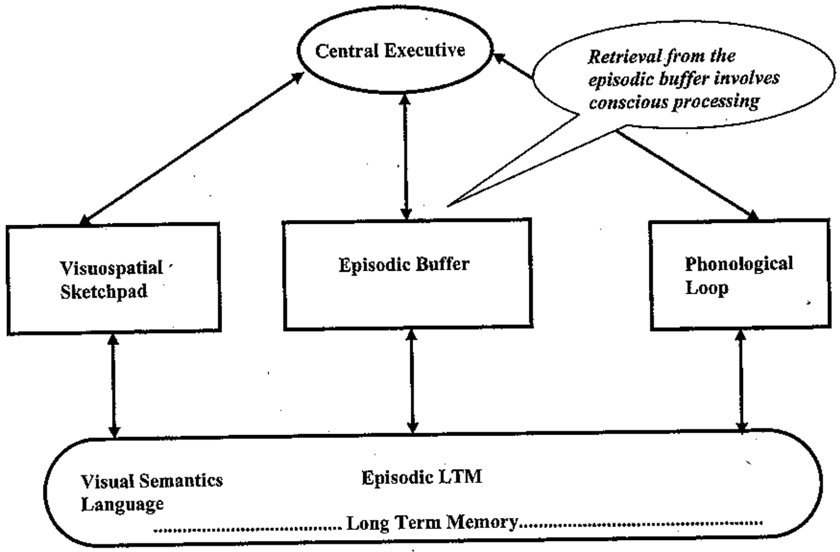

Figure 8.3. Baddeley's Model of Working Memory. This model incorporates the episodic buffer. (Adapted from Baddeley, 2000.)

particularly transactions between the central executive and the subsystems, has been associated with conscious and effortful information processing. However, these were rarely stated in terms of the question of consciousness as such.

Recently, Baddeley (2000, 2001) has proposed an additional WM component called the episodic buffer (see Figure 8.3 for a depiction of the most recent model). This addition to the WM architecture means that the central executive now becomes a purely attentional, controlled process while multimodal information storage devolves onto the episodic buffer. The episodic buffer "comprises a limited capacity system that provides temporary storage of information held in a multimodal code, which is capable of binding information from the subsidiary systems, and from long-term memory, into a unitary episodic representation. Conscious awareness is assumed to be the principal mode of retrieval from the buffer" (Baddeley, 2000, p. 417). Baddeley (2001) believes that the binding function served by the episodic buffer is "the principal biological advantage of consciousness" (p. 858).

Conscious processing in WM appears to reside in the *transactions* of the central executive with the episodic buffer (and perhaps with the visuospatial sketchpad and phonological loop), in which the central executive controls and switches attention while the episodic buffer creates and makes available multimodal information.

Baddeley's episodic buffer resembles other models of consciousness in its ability to briefly hold multimodal information and to combine many information sources into a unitary representation. A major difference between WM and other models is that WM was not proposed as a model of consciousness in general. It is restricted to an accounting of mnemonic processes – both conscious and unconscious. In addition, the WM model does not assume that contents of the episodic buffer are "broadcast" systemwide as a means of organizing and recruiting other non-mnemonic processes. No account is given of the further distribution of information from the episodic buffer, once it is accessed by the central executive.

SCHNEIDER AND PIMM-SMITH'S MESSAGE-AWARE CONTROL MECHANISM

Schneider and Pimm-Smith have proposed a model of cognition that incorporates a conscious processing component and allows widespread distribution of information from specialist modules (Schneider &

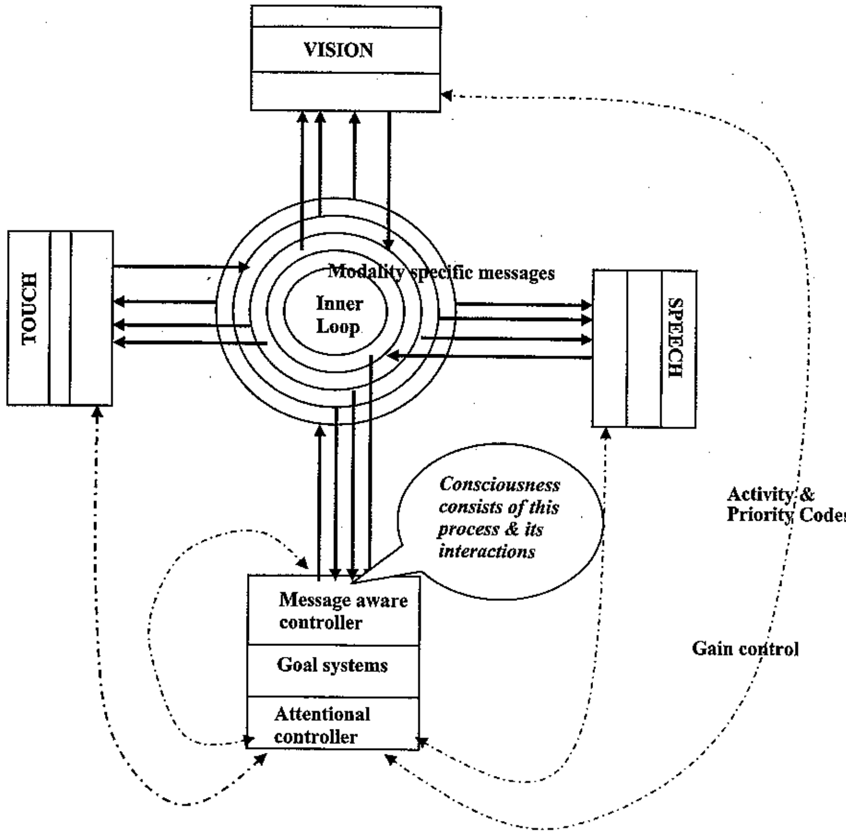

Figure 8.4. A simplified view of Schneider and Pimm-Smith's Message-Aware Control Mechanism.

Pimm-Smith, 1997). The model is an attempt to capture the adaptive advantage that consciousness adds to cognitive processing. According to Schneider and Pimm-Smith, "consciousness may be an evolutionary extension of the attentional system to modulate cortical information flow, provide awareness, and facilitate learning particularly across modalities. . . . According to [the] model, consciousness functions to monitor and transmit global messages that are generally received by the whole system serially to avoid the cross-talk problem (pp. 65, 76)." The conscious component of this model, the conscious controller, stands between the serially arriving high-level messages forwarded by the specialist modules and the attentional controller, which sends scalar messages back to the specialist modules according to their value in ongoing activities. The conscious controller is not privy to all information flowing within the system, examining data only after lower-level modules have produced invariant codes and selecting those messages that relate to currently active goals. Figure 8.4 illustrates the functional relations among components in Schneider and Pimm-Smith's model.

The message-aware control mechanism model of consciousness depends on localized specialist processors – auditory, haptic, visual, speech, motor, semantic, and spatial modules – which each have their own internal levels of processing. These modules feed their output codes serially and separately to other modules and to the consciousness controller via an inner loop. According to Schneider and Pimm-Smith, "Consciousness is a module on the inner loop that receives messages that evoke symbolic codes that are utilized to control attention to specific goals (p. 72)." Furthermore, "consciousness is the message awareness of the controller.... This message awareness allows control decisions based on the specific messages transmitted in the network rather than labeled line activity or priority codes from the various modules in the network (pp. 72–73)." There is also information sharing through non-conscious channels. The attentional controller receives activity and priority signals from specialist modules marking the availability of input through non-conscious outer loop channels, and it sends gain control information back to the specialists.

Schneider and Pimm-Smith's model has several notable characteristics: (1) information is widely distributed; (2) particular content is created by localized specialist modules, which themselves have levels of processing; (3) consciousness is identified as a separable and separate function that modulates the flow of information throughout the system: and (4) access to consciousness is through reference to goal systems. This elegant model accounts for the modulation of attention by reference both to current goals and to the recruitment of specialist modules in the service of goals through the global distribution of conscious messages.

Network Theories That Explain Consciousness as Patterns of System-Wide Activity

Although the information-processing theories in the previous section attempted to model consciousness in terms of the selection and interactions of specialized, semi-autonomous processing modules, the theories described in the present section build on networks and connectionist webs. It may be helpful to view the information-processing models as macroscopic and the network theories as microscopic. It is possible that the activities attributed to modules in the information-processing models could, in fact, be seen to be carried out by connectionist networks when viewed microscopically. As we examine network theories, it should be kept in mind that, although network theorists often make comparisons with and assumptions about networks composed of neurons, most network theories, including those described here, are in fact functional descriptions of activities that are presumed to occur in the brain. Brain networks themselves are not directly assessed.

PRIBRAM'S HOLONOMIC THEORY

Karl Pribram has developed a holographic theory (more recently, holonomic theory) of brain function and conscious processing (Pribram, 1971). He has built on the mathematical formulations of Gabor (1946), combining holography and communication theory. To state things in simple form, the holonomic model postulates that the brain is holograph machine. That is, the brain handles, represents, stores, and reactivates information through the medium of "wetware" holograms. Holograms, though complex mathematically, are familiar to us as the three-dimensional images found on credit cards and driver's licenses. Such a hologram is a photograph of an optical interference pattern, which contains information about intensity and phase of light reflected by an object. When illuminated with a coherent source of light, it will yield a diffracted wave that is identical in amplitude and phase distribution with the light reflected from the original object. The resulting three-dimensional image is what we see on our credit cards.

In a holonomic model, Pribram says, information is encoded by wave interference patterns rather than by the binary units (BITs) of computer science. Rather than encoding sensory experience as a set of features that are then stored or used in information processing, sensory input in the holonomic view is encoded as the interference pattern resulting from interacting waves of neuronal population activity. Stated in the language of visual perception, the retinal image is understood to be coded by a spatial frequency distribution over visual cortex, rather than by individual features of the visual scene. Because the surface layer of the cortex consists of an entangled "feltwork" of dendrites, Pribram suggests that cortex represents a pattern of spatial correlations in a continuous dendritic medium, rather than by discrete, localized features expressed by the firing of single neurons. The dendritic web of the surface layers of cortex is the medium in which representations are held. Functionally, it is composed of oscillating graded potentials. In Pribram's view, nerve impulses have little or no role to play in the brain web. However, dendritic potentials obviously trigger axonal firing when they rise above a neuronal threshold, so that long-distance axons would also be triggered by the dendro-dendritic feltwork.

The holonomic model finds support in the neuropsychological finding that focal brain damage does not eliminate memory content. Further, it is consistent with the fact that each sensory neuron tends to prefer one type of sensory input, but often fires to different inputs as well. Neurons do not operate individually but rather participate in different cell assemblies or active populations at different times; the cell assemblies are themselves "kaleidoscopic."

In a hologram, the information necessary to construct an image is inherently distributed. Pribram explains how the notion of distributed information can be illustrated with a slide projector (Pribram & Meade, 1999). If one inserts a slide into the projector and shows a "figure," then removes the lens from the front of the projector, there is only a fuzzy bright area. There is nothing, "no-*thing*," visible on the screen. But that does not mean there is no information in the light. The information can be re-visualized by placing a pair of reading glasses into the light beam. On the screen, one now again sees the "figure" in the slide. Putting two lenses of the eyeglasses in the beam, one sees two "figures" that can be made to appear in any part of the bright area. Thus any part of the beam of light carries all the information needed to reconstruct the picture on the slide. Only resolution is lost.

According to the holonomic theory, no "receiver" is necessary to "view" the result of the transformation (from spectral holographic to "image"), thus avoiding the homunculus problem (the problem of infinite regress, of a little man looking at the visual scene, which in turn needs a little man in its brain, and so on ad infinitum). It is the activity of the dendro-dendritic web that gives rise to the experience. Correspondingly, remembering is a form of re-experiencing or re-constructing the initial sensory input, perhaps by cuing a portion of the interference pattern. Finally, Pribram believes that we "*become aware of* our conscious experience due to a delay between an incoming pattern of signals before it matches a previously established outgoing pattern" (Pribram & Meade, 1999, p. 207).

Pribram's holonomic model is attractive in that it can account for the distributed properties of memory and sensory processing. The model makes use of the dendritic feltwork that is known to exist in the surface layer of cortex. However, the model fails to help us understand the difference between conscious and unconscious processing or the unique functions of consciousness *qua* consciousness. Pribram has not treated consciousness as a variable and cannot tell us what it is that consciousness adds to the cognitive system.

EDELMAN AND TONONI'S DYNAMIC CORE HYPOTHESIS

Edelman and Tononi's theory of consciousness (2000; see also Tononi & Edelman, 1998) combines evidence from large-scale connectivities in the thalamus and cortex, behavioral observation, and mathematical properties of large-scale brain-like networks. Based on neuropsychological and lesion evidence that consciousness is not abolished by losses of large volumes of brain tissue (Penfield, 1958), Edelman and Tononi reject the idea that consciousness depends on participation of the whole brain in an undifferentiated fashion. At the same time, they, along with many others, reject the view that consciousness depends only on local properties of neurons. Tononi and Edelman cite, for example, PET evidence suggesting that moment-to-moment awareness is highly correlated with increasing functional connectivity between diverse cortical regions (see, for example, McIntosh, Rajah, & Lobaugh, 1999). In other words, the same cortical areas seem to participate in conscious experience or not at different times, depending on their current dynamic connectivity. This idea is resonant with Pribram's description of neural assemblies being "kaleidoscopic."

The fundamental idea in Edelman and Tononi's theory is the *dynamic core hypothesis*, which states that conscious experience arises from the activity of an ever-changing functional cluster of neurons in the thalamocortical complex of the brain, characterized by high levels of differentiation as well as strong reciprocal, re-entrant interaction over periods of hundreds of milliseconds. The particular neurons participating in the dynamic core are ever changing while internal integration in the dynamic core is maintained through re-entrant connections. The hypothesis highlights the functional connections of distributed groups of neurons, rather than their local properties; thus, the same group of neurons may at times be part of the dynamic core and underlie conscious experience, whereas at other times, this same neuronal group will not be part of the dynamic core and will thus be part of unconscious processing. Consciousness in this view is not a thing or a brain location but rather, as William James argued, a process, occurring largely within the re-entrant meshwork of the thalamocortical system.

Edelman and Tononi (2000) take issue with Baars' (1988) concept of global broadcasting (see below) as a way to explain capacity limits and wide access in conscious processing. In Baars' view, the information content of any conscious state is apparently contained in the single message that is being broadcast to specialist systems throughout the brain at any one moment; information content is thus limited but widely distributed. Edelman and Tononi (2000) argue for an alternative view: that the information is not in the message, but rather in the number of system states that can be brought about by global interactions within the system itself. In place of Baars' broadcasting or theater metaphor, they offer an alternative:

> *[A] better metaphor would be . . . a riotous parliament trying to make a decision, signaled by its members raising their hands. Before counting occurs, each member of parliament is interacting with as many other members as possible not by persuasive rhetoric . . . but by simply pushing and pulling. Within 300 msec., a new vote is taken. How informed the decision turns out to be will depend on the number of diverse interactions within the parliament. In a totalitarian country, every member will vote the same; the information content of constant unanimity is zero. If there are two monolithic groups, left and right, such that the vote of each half is always the same, the information content is only slightly higher. If nobody interacts with anyone, the voting will be purely random, and no information will be integrated within the system. Finally, if there are diverse interactions within the parliament, the final vote will be highly informed (Edelman & Tononi, 2000, pp. 245–246).*

A constantly changing array of ever-reorganized mid-sized neuronal groups in a large system of possible groups has high levels of complexity and integration – characteristics of conscious states. Within this model, unconscious specialist systems are local, non-integrated neuronal groups. How the unconscious specialists are recruited into the dynamic core is not made entirely clear in the theory. Edelman and Tononi say that consciousness in its simplest form emerges in the re-entrant linkage between current perceptual categorization and value-category memory (short-term and long-term memory). Conscious experience is actually a succession of 100-ms snapshots of the current linkages that constitute the "remembered present."

Perhaps Baars and Tononi and Edelman are not so different on closer examination. Baars' (1988, 1998) model supposes that there is reciprocal exchange between the global workspace (GW) and specialist systems in the architecture of consciousness; it is difficult to see why this is different from the re-entrant linkages between neuronal groups in Edelman and Tononi's theory. Furthermore, within any one "snapshot" of the system, the pattern of dynamically linked elements in Baars' model – GW and specialists that are able to receive the particular

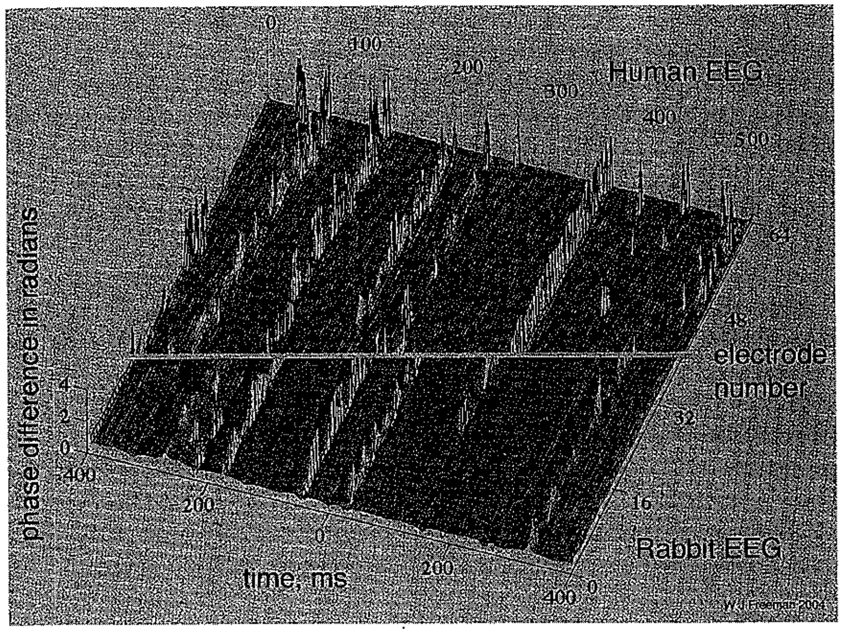

Figure 8.5. A Hilbert analysis of analytic phase differences in EEG across cortical surface measured over 400 ms in rabbit and human conscious processing. Phase differences are calculated in the beta band (12–30 Hz) for human EEG and in the gamma band (20–50 Hz) for the rabbit EEG. (With permission of the author.) (See color plates.)

message that has been disseminated – looks very much like the pattern of momentarily linked neuronal groups in Tononi and Edelman's model that are recruited in the moment depending on environmental input and value memories. Two strengths of the Tononi and Edelman model are its acknowledgment of long-distance connectivity among specialist brain regions as a characteristic of conscious processes, as well as the dynamic nature of these connections.

WALTER FREEMAN'S DYNAMICAL SYSTEMS APPROACH: FRAMES IN THE CINEMA

Like Pribram, Walter Freeman has worked to obtain empirical support for a cortex-wide dynamic neural system that can account for behavioral data observed in conscious activities. Freeman's Dynamical Systems approach to consciousness is built on evidence for repetitive global phase transitions occurring simultaneously over multiple areas of cortex during normal behavior (see Figure 8.5 Freeman, 2004; Freeman & Rogers, 2003). Freeman and his colleagues have analyzed EEGs, recorded from multiple high-density electrode arrays (64 electrodes) fixed on the cortex of rabbits and on the scalp of human volunteers. An index of synchronization was obtained for pairs of signals located at different cortical sites to detect and display epochs of mutual engagement between pairs. The measure was adapted to derive an index of global synchronization among all four cortices (frontal, parietal, temporal, and occipital) – global epochs of phase stabilization ("locking") involving all cortices under observation during conscious perceptual activity. These epochs of phase locking can be seen in the "plateaus" of global coherence in Figure 8.5. The peaks in the figure indicate momentary, global decoherence.

To understand Freeman's findings, we have to understand the basics of Hilbert analysis as it is shown in Figure 8.5. Hilbert analysis of the EEGs recorded from electrode arrays produces a three-dimensional graphical representation. In it, the phase difference between pairs of cortical electrodes within a particular EEG band is plotted against time (in milliseconds) and spatial location (represented by electrode number). The resulting plot is called a Hilbert space. The Hilbert space can be read like a topographical map. In the plot in Figure 8.5, we can see many flat plateau areas lying between peaked ridges. The plateaus represent time periods (on the order of 50 ms) in which many pairs of EEG signals from different cortical locations are found to be in phase with each other. The ridges represent very short intervals when all of these pairs are simultaneously out of phase, before returning to phase locking. These out-of-phase or decoherent epochs appear to be non-conscious transitions between moments of consciousness.

According to Freeman (2004),

> *The EEG shows that neocortex processes information in frames like a cinema. The perceptual content is found in the phase plateaus from rabbit EEG; similar content is predicted to be found in the plateaus of human scalp EEG. The phase jumps show the shutter. The resemblance across a 33-fold difference in width of the zones of coordinated activity reveals the self-similarity of the global dynamics that may form Gestalts (multisensory percepts). (Caption to cover illustration, p. i)*

Freeman's data are exciting in their ability to map the microscopic temporal dynamic changes in widespread cortical activity during conscious perception – something not found in other theories. As a theory of consciousness, the dynamical systems approach focuses primarily on describing conscious perceptual processing at the cortical level. It does not attempt to explain the conscious/non-conscious difference or the function of consciousness in the neuropsychological system. With many neurocognitive theorists, we share Freeman's question about how the long-range global state changes come about virtually simultaneously. Freeman's hypothesis of Self-organized Criticality suggests that the neural system is held in a state of dynamic tension that can change in an all-or-none fashion with small environmental perturbations. He says "a large system can hold itself in a near-unstable state, so that by a multitude of adjustments it can adapt to environments that change continually and unpredictably" (Freeman & Rogers, 2003, p. 2882).

Globalist Models That Combine Aspects of Information-Processing Theories and Network Theories

BAARS' GLOBAL WORKSPACE THEORY

A theater metaphor is the best way to approach Baars' Global Workspace (GW) theory (Baars, 1988, 1998, 2001). Consciousness is associated with a global "broadcasting system" that disseminates information widely throughout the brain. The metaphor of broadcasting explicitly leaves open the precise nature of such a wide influence of conscious contents in the brain. It could vary in signal fidelity or degree of distribution, or it might not involve "labeled line" transmission at all, but rather activation passing, as in a neural network. Metaphors are only a first step toward explicit theory, and some theoretical decision points are explicitly left open.

If consciousness is involved with widespread distribution or activation, then conscious capacity limits may be the price paid for the ability to make single momentary messages act upon the entire system for purposes of coordination and control. Because at any moment there is only one "whole system," a global dissemination capacity must be limited to one momentary content. (There is evidence that the duration of each conscious "moment" may be on the order of 100 ms, one-tenth of a second – see Blumenthal, 1977).

Baars develops these ideas through seven increasingly detailed models of a global workspace architecture, in which many parallel unconscious experts interact via a serial, conscious, and internally consistent global workspace (1983, 1988). Global workspace architectures or their functional equivalents have been developed by cognitive scientists since the 1970s; the notion of a "blackboard" where messages from specialized subsystems can be "posted" is common to the work of Baars (1988), Reddy and Newell (1974), and Hayes-Roth (1984). The global workspace framework has a family resemblance to the well-known integrative theories of Herbert A. Simon (General Problem Solver or EPAM), Allan Newell (SOAR, 1992), and John R. Anderson (ACT*, 1983). Architectures much like this have also seen some practical applications. GW theory is currently a thoroughly developed framework, aiming to explain an large set of evidence. It appears to have fruitful implications for a number of related topics, such as spontaneous problem solving, voluntary control, and even the Jamesian "self" as agent and observer (Baars, 1988; Baars, Ramsoy, & Laureys, 2003).

GW theory relies on three theoretical constructs: unconscious specialized processors, a conscious Global Workspace, and unconscious contexts.

The first construct is the *unconscious specialized processor*, the "expert" of the psychological system. We know of hundreds of types of "experts" in the brain. They may be single cells, such as cortical feature detectors for color, line orientation, or faces, or entire networks and systems of neurons, such as cortical columns, functional areas like Broca's or Wernicke's areas, and basal ganglia. Like human experts, unconscious specialized processors may sometimes be quite "narrow-minded." They are highly efficient in limited task domains and able to act independently or in coalition with each other. Working as a coalition, they do not have the narrow capacity limitations of consciousness, but can receive global messages. By "posting" messages in the global workspace (consciousness), they can send messages to other experts and thus recruit a coalition of other experts. For routine missions they may work autonomously, without conscious involvement, or they may display their output in the global workspace, thus making their work conscious and available throughout the system. Answering a question like "What is your mother's maiden name?" requires a mission-specific coalition of unconscious experts, which report their answer to consciousness. Figure 8.6 shows the major constructs in GW theory and the functional relations among them.

The second construct is, of course, the *global workspace (GW)* itself. A global workspace is an architectural capability for system-wide integration and dissemination of information. It is much like the podium at a scientific meeting. Groups of experts at such a meeting may interact locally around conference tables, but to influence the meeting as a whole any expert must compete with others, perhaps supported by a coalition of like-minded experts, to reach the podium, whence global messages can be broadcast. New links among experts are made possible by global interaction via the podium and can then spin off to become new local processors. The podium allows novel expert coalitions to form that can work on new or difficult problems, which cannot be solved by established experts and committees. Tentative solutions to problems can then be globally disseminated, scrutinized, and modified.

The evidence presented in Tables 8.2 and 8.3 falls into place by assuming that information in the global workspace corresponds to conscious contents. Because conscious experience seems to be oriented primarily toward perception, it is convenient to imagine that preperceptual processors – visual, auditory, or multimodal – can compete for access to a brain version of a global workspace. For example, when someone speaks to us, the speech stream receives preperceptual processing through the speech specialist systems before the message in the speech stream is posted in consciousness. This message is then globally broadcast to the diverse specialist systems and can become the basis for action, for composing a verbal reply, or for cuing related memories. In turn, the outcome of actions carried out by expert systems can also be monitored and returned to consciousness as action feedback.

Obviously the abstract GW architecture can be realized in a number of different ways in the brain, and we do not know at this point which brain structures provide

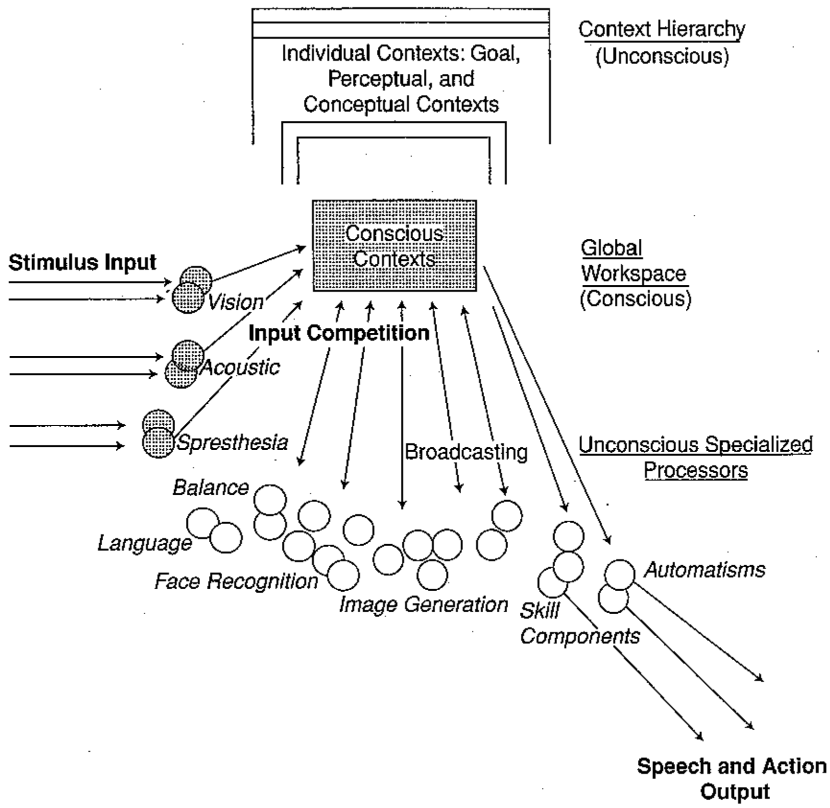

Figure 8.6. Global Workspace Architecture: Basic constructs and their relations.

the best candidates. Although its brain correlates are not entirely clear at this time, there are possible neural analogs, including the reticular and intralaminar nuclei of the thalamus, one or more layers of cortex, long-range cortico-cortico connections, and/or active loops between sensory projection areas of cortex and the corresponding thalamic relay nuclei. Like other aspects of GW theory, such neural candidates provide testable hypotheses (Newman & Baars, 1993). All of the neurobiological proposals described in this chapter provide candidates (Freeman, 2004; Dehaene & Naccache, 2001; Edelman & Tononi, 2000; Tononi & Edelman, 1998), and some have been influenced by GW theory.

Context, the third construct in GW theory, refers to the powers behind the scenes of the theater of mind. Contexts are coalitions of expert processors that provide the director, playwright, and stagehands behind the scenes of the theater of mind. They can be defined functionally as knowledge structures that constrain conscious contents without being conscious themselves, just as the playwright determines the words and actions of the actors on stage without being visible. Conceptually, contexts are defined as preestablished expert coalitions that can evoke, shape, and guide global messages without themselves entering the global workspace.

Contexts may be momentary, as in the way the meaning of the first word in a sentence shapes an interpretation of a later word like "set," or they may be long lasting, as with life-long expectations about love, beauty, relationship, social assumptions, professional expectations, worldviews, and all the other things people care about. Although contextual influences shape conscious experience without being conscious, contexts can also be set up by conscious events. The word "tennis" before "set" shapes the interpretation of "set," even when "tennis" is already gone from consciousness. But "tennis" was

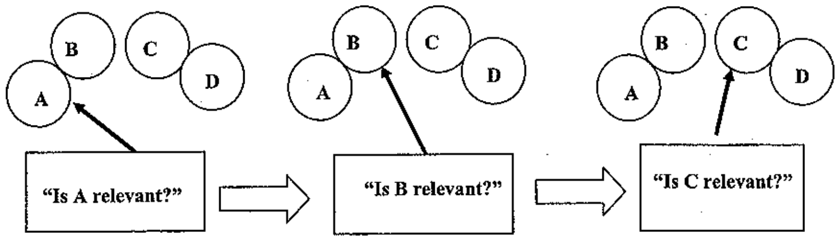

Figure 8.7. A naive processor approach to environmental novelty.

initially conscious and needed to be conscious to create the unconscious context that made sense of the word "set."

Thus conscious events can set up unconscious contexts. The reader's ideas about consciousness from years ago may influence his or her current experience of this chapter, even if the memories of the earlier thoughts do not become conscious again. Earlier experiences typically influence current experiences as contexts, rather than being brought to mind. It is believed for example that a shocking or traumatic event earlier in life can set up largely unconscious expectations that may shape subsequent conscious experiences.

SHANAHAN: AN ANSWER TO THE MODULARITY AND FRAME PROBLEMS

Shanahan and Baars (2005) suggest that the global workspace approach may provide a principled answer to the widely discussed "modularity" and "frame" problems. Fodor (1983) developed the view that cognitive functions like syntax are performed by "informationally encapsulated" modules, an idea that has some empirical plausibility. However, as stated by Fodor and others, modules are so thoroughly isolated from each other that it becomes difficult to explain how they can be accessed, changed, and mobilized on behalf of general goals. A closely related difficulty, called the frame problem, asks how an autonomous agent can deal with novel situations without following out all conceivable implications of the novel event. For example, a mobile robot on a cart may roll from one room to another. How does it know what is new in the next room and what is not, without explicitly testing out all features of the new environment? This task quickly becomes computationally prohibitive. Shanahan and Baars (2005) point out that the following:

> *What the global workspace architecture has to offer... 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....*

Compare the naive processor's inefficient approach (depicted in Figure 8.7) with a massively parallel and distributed global workspace approach (depicted in Figure 8.8) to dealing with environmental novelty.

The key point here is that the GW architecture permits widely distributed local responsibility for processing global signals. As was pointed out above, conscious and non-conscious process differ in their capabilities – they are two different modes of processing that, when combined, offer powerful adaptive possibilities.

FRANKLIN'S IDA AS AN IMPLEMENTATION OF GW THEORY

Stan Franklin and colleagues (Franklin, 2001; Franklin & Graesser, 1999) have developed a practical implementation of GW theory in large-scale computational agents to test its functionality in complex practical tasks.

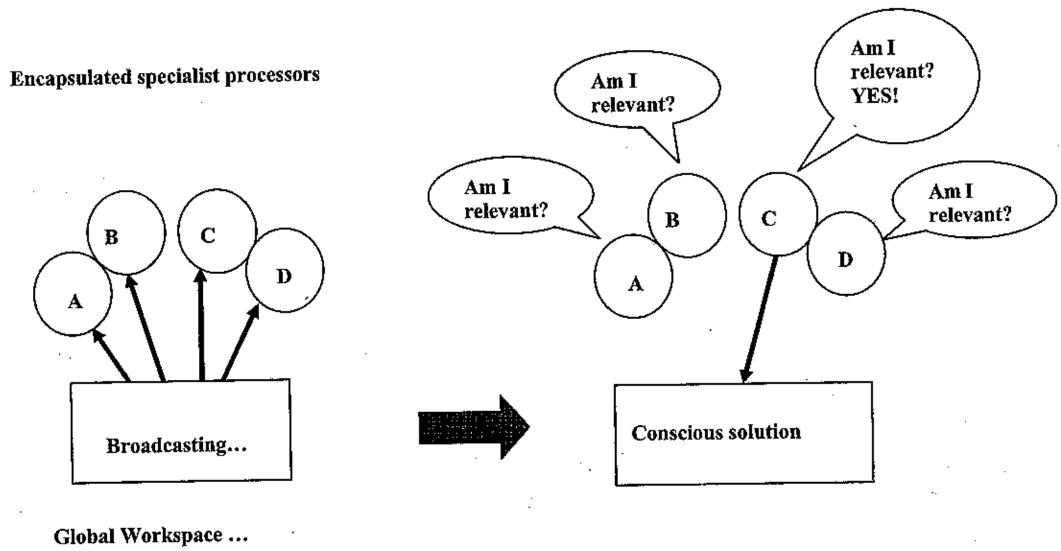

Figure 8.8. A GW approach to environmental novelty.

IDA, or Intentional Distribution 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 (see Chapter 7). The particular domain in this case is interaction among U.S. Navy personnel experts and sailors who move from job to job. IDA interacts with sailors via e-mail and is able to combine numerous regulations, sailors' preferences, and time, location, and travel considerations into human-level performance. Although it has components roughly corresponding to human perception, memory, and action control, the heart of the system is a GW architecture that allows input messages to be widely distributed, so that specialized programs called "codelets" can respond with solutions to centrally posed problems (see Figure 8.9).

Franklin writes, "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" (see http://csrg.cs.memphis.edu). Although agent simulations do not prove that GW architectures exist in the brain, they demonstrate their functionality. Few if any large-scale cognitive models can be shown to actually perform complex human tasks, but somehow the real cognitive architecture of the brain does so. In that respect, the test of human-level functionality is as important in its way as any other source of evidence.

DEHAENE'S GLOBAL NEURONAL NETWORK THEORY

Stanislas Dehaene and his colleagues (Dehaene & Naccache, 2001; Dehaene, Kerszberg, & Changeux, 1998) have recently proposed a *global neuronal workspace theory* of consciousness based on psychological and neuroscientific evidence quite similar to that cited by Baars and others. Dehaene and colleagues identify three empirical observations that any theory of consciousness must be able to account for: "namely (1) a considerable amount of processing is possible without consciousness, (2) attention is a prerequisite of consciousness, and (3) consciousness is required for some specific cognitive tasks, including those that require durable information maintenance, novel combinations of operations, or

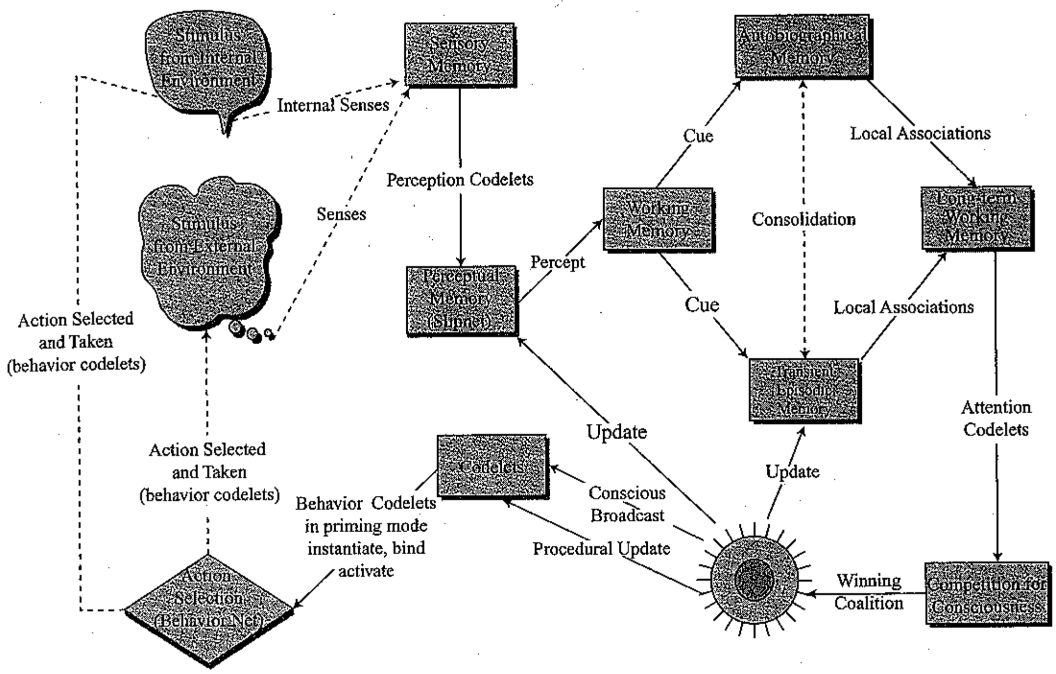

Figure 8.9. Franklin's IDA Model.

the spontaneous generation of intentional behavior" (p. 1). The Dehaene and Naccache model depends on several well-founded assumptions about conscious functioning.

The first assumption is that non-conscious mental functioning is modular. That is, many dedicated non-conscious modules can operate in parallel. Although arguments remain as to whether psychological modules have immediate correlates in the brain, Dehaene and Naccache (2001) say that the "automaticity and information encapsulation acknowledged in cognitive theories are partially reflected in modular brain circuits." They tentatively propose that "a given process, involving several mental operations, can proceed unconsciously only if a set of adequately interconnected modular systems is available to perform each of the required operations" (p. 12; see Figure 8.10).

The second assumption, one shared by other cognitive theories, is that controlled processing requires an architecture in addition to modularity that can establish links among the encapsulated processors. Dehaene et al. (1998) argue that a distributed neural system or "workspace" with long-distance connectivity is needed that can "potentially interconnect multiple specialized brain areas in a coordinated, though variable manner" (p. 13).

The third assumption concerns the role of attention in gating access to consciousness. Dehaene and Naccache (2001) review evidence in support of the conclusion that considerable processing can occur without attention, but that attention is required for information to enter consciousness (Mack & Rock, 1998). They acknowledge a similarity between Michael Posner's hypothesis of an attentional amplification (Posner, 1994) and their own proposal. Attentional amplification explains the phenomena of consciousness as due to the orienting of attention, which causes increased cerebral activation in attended areas and a transient increase in their efficiency. According to Dehaene & Nacache (2001),

> *[I]nformation becomes conscious ... if the neural population that represents it is mobilized by top-down attentional amplification into a brain-scale state of coherent activity that involves many neurons distributed throughout the brain. The long-distance*

connectivity of these 'workspace neurons' can, when they are active for a minimal duration, make the information available to a variety of processes including perceptual categorization, long-term memorization, evaluation, and intentional action. (p. 1)

An implication of the Dehaene and Naccache model is that consciousness has a granularity, a minimum duration of long-distance integration, below which broadcast information will fail to be conscious.

It is worth noting a small difference between Baars' version of global workspace and that of Dehaene and colleagues. (Dehaene & Naccache, 2001; Dehaene et al., 1998). They believe that a separate attentional system intervenes with specialized processors to allow their content to enter the global workspace and become conscious. Baars (1998), on the other hand, sees attention not as a separate system but rather as the name for the process of gaining access to global workspace by reference to long-term or current goals. Clearly, further refinement is needed here in thinking through what we mean by attention or an attentional system as separate from the architecture of consciousness, in this case, varieties of GW architecture.

Dehaene, Sargent, and Changeux (2003) have used an implementation of the global neuronal workspace model to successfully simulate attentional blink. Attentional blink is a manifestation of the all-or-none characteristic of conscious processing observed when participants are asked to process two successive targets, T1 and T2. When T2 is presented between 100 and 500 ms after T1, the ability to report it drops, as if the participants' attention had "blinked." During this blink, T2 fails to evoke a P300 potential but still elicits event-related potentials associated with visual and semantic processing (P1, N1, and N400). Dehaene et al. (2003) explain,

Our simulations aim at clarifying why some patterns of brain activity are selectively associated with subjective experience. In short, during the blink, bottom-up activity, presumably generating the P1, N1, and N400 waveforms, would propagate without necessarily creating a global reverberant state. However, a characteristic neural signature of long-lasting distributed activity and g-band emission, presumably generating the P300 waveform, would be associated with global access. (p. 8520)

In the simulation, a network modeled the cell assemblies evoked by T1 and T2 through four hierarchical stages of processing, two separate perceptual levels and two higher association areas. The network was initially assigned parameters that created spontaneous thalamocortical oscillations, simulating a state of wakefulness. Then, the network was exposed to T1 and T2 stimulation at various interstimulus intervals (ISI). T1 excitation was propagated bottom-up through all levels of the processing hierarchy, followed by top-down amplification signals that resulted in sustained firing of T1 neurons. Dehaene et al. (2003) hypothesized that this sustained firing and global broadcasting may be the neural correlate of conscious reportability. In contrast, the activation evoked by T2 depended closely on its timing relative to T1. For simultaneous and long ISIs, T2 excitation evoked sustained firing. Importantly, when T2 was presented during T1-elicited global firing, it evoked activation only in the low-level perceptual assemblies and resulted in no global propagation. Dehaene and colleagues conclude that this detailed simulation has provided tentative links between subjective reports and "objective physiological correlates of consciousness on the basis of a neurally plausible architecture" (2003, p. 8524).

The Globalist Argument: An Emerging Consensus

In the last two decades, a degree of consensus has developed concerning the role of consciousness in the neuropsychological architecture. The general position is that consciousness operates as a distributed and flexible system offering nonconscious expert systems global accessibility

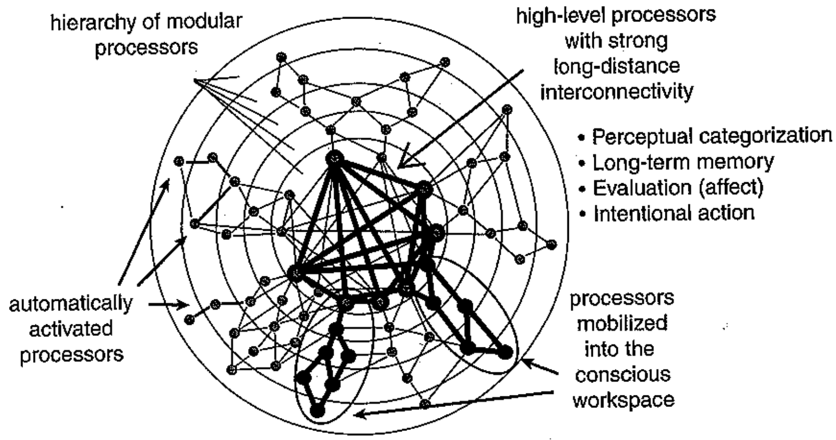

Figure 8.10. A global neuronal network account of conscious processes (Dehaene & Naccache, 2001, p. 27).

to information that has a high concurrent value to the organism. Although consciousness is not itself an executive system, a global distribution capacity has obvious utility for executive control, in much the way that governments can control nations by influencing nation-wide publicity.

Excerpted below are the views of prominent researchers on consciousness revealing considerable agreement.

- **Baars (1983):** "Conscious contents provide the nervous system with *coherent, global information.*"
- **Damasio (1989):** "Meaning is reached by time-locked multiregional retroactivation of widespread fragment records. Only the latter records can become contents of consciousness."
- **Freeman (1991):** "The activity patterns that are formed by the (sensory) dynamics *are spread out over large areas of cortex,* not concentrated at points. Motor outflow is likewise *globally distributed*.... In other words, the pattern categorization does not correspond to the selection of a key on a computer keyboard but to an induction of a global activity pattern." [Italics added]
- **Tononi and Edelman (1998):** "The dynamic core hypothesis avoids the category error of assuming that certain local, intrinsic properties of neurons have, in some mysterious way, a privileged correlation with consciousness. Instead, this hypothesis accounts for fundamental properties of conscious experience by linking them to global properties of particular neural processes" (p. 1850).
- **Llinas et al. (1998):** "... *the thalamus represents a hub from which any site in the cortex can communicate with any other such site or sites*.... temporal coincidence of specific and non-specific thalamic activity generates the functional states that characterize human cognition" (p. 1841).
- **Edelman and Tononi (2000):** "When we become aware of something... it is as if, suddenly, many different parts of our brain were privy to information that was previously confined to some specialized subsystem.... the wide distribution of information is guaranteed mechanistically by *thalamocortical and corticocortical reentry*, which facilitates the interactions among distant regions of the brain" (pp. 148–149).
- **Dennett (2001):** "Theorists are converging from quite different quarters on a version of the global neuronal workspace model of consciousness" (p. 42).
- **Kanwisher (2001):** "... it seems reasonable to hypothesize that awareness of a particular element of perceptual information must entail not just a strong enough neural representation of information, but

also access to that information by most of the rest of the mind/brain."

- **Dehaene and Naccache (2001):** "We propose a theoretical framework... the hypothesis of a global neuronal workspace.... We postulate that this global availability of information through the workspace is what we subjectively experience as the conscious state."
- **Rees (2001):** "One possibility is that activity in such a distributed network might reflect stimulus representations gaining access to a 'global workspace' that constitutes consciousness" (p. 679).
- **John et al. (2001):** "Evidence has been steadily accumulating that information about a stimulus complex is distributed to many neuronal populations dispersed throughout the brain."
- **Varela et al. (2001):** "... the brain ... transiently settling into a globally consistent state ... [is] the basis for the unity of mind familiar from everyday experience."
- **Cooney and Gazzaniga (2003):** "Integrated awareness emerges from modular interactions within a neuronal workspace.... The presence of a large-scale network, whose long-range connectivity provides a neural workspace through which the outputs of numerous, specialized, brain regions can be interconnected and integrated, provides a promising solution... In the workspace model, outputs from an array of parallel processors continually compete for influence within the network" (p. 162).
- **Block (2005):** "Phenomenally conscious content is what differs between experiences as of red and green, whereas access conscious content is information which is 'broadcast' in the global workspace."

Although debate continues about the functional character of consciousness, the globalist position can be summarized in the following propositions:

1. The architecture of consciousness comprises numerous, semi-autonomous specialist systems, which interact in a dynamic way via a global workspace.
2. The function of the workspace is global distribution of information in order to recruit resources in the service of current goals.
3. Specialist systems compete for access to the global workspace; information that achieves access to the workspace obtains system-wide dissemination.
4. Access to the global workspace is "gated" by a set of active contexts and goals.

DISSENTING VIEWS

The globalist position argues that consciousness provides a momentary unifying influence for a complex system through global distribution and global access. In this sense, consciousness may be said to have unity. Alternative views chiefly depart from the globalist position on this point. They argue, in one way or another, that consciousness is not fundamentally unified.

One alternative view is that of Marcel (1993) who argued for "slippage" in the unity of consciousness. In part, he made his case based on his observation that different reporting modalities (blink vs. finger tap) could produce conflicting reports about conscious experience. Marcel took this to indicate that consciousness itself is not unified in any real sense.

Marcel's argument bears some similarity to Dennett's "multiple drafts" argument (Dennett, 1991). Dennett pointed to the puzzle posed by the phi phenomenon. In the phi phenomenon, we observe a green light and a red light separated by a few degrees in the field of vision as they are flashed in succession. If the time between flashes is about one second or less, the first light flashed appears to move to the position of the second light. Further, the color of the light appears to change midway between the two lights. The puzzle is explaining how we could see the color change before we see the position of the second light. Dennett hypothesizes that the mind creates different analyses or narratives (multiple drafts) of the scene at different moments from different sensory inputs. All of the accounts are available to influence behavior and report. A given scene can give rise to more than one interpretation. In contrast, with global broadcasting models, Dennett says there is no single version of the scene available anywhere in the psychological system.

Simiarly, Zeki (2001, 2003) has argued on neurological grounds that there is "disunity" in the neural correlates of consciousness. With many others, Zeki notes that the visual brain consists of many separate, functionally specialized processing systems that are autonomous with respect to one another. He then supposes that activity at each node reaches a perceptual endpoint at a different time, resulting in a perceptual asynchrony in vision. From there, Zeki makes the inference that activity at each node generates a microconsciousness. He concludes that visual consciousness is therefore distributed in space and time, with an organizing principle of abstraction applied separately within each processing system. It remains to be seen whether Zeki's microconsciousnesses can be examined empirically via contrastive analysis and whether the microconsciousnesses are necessarily conscious or simply potentially conscious. The globalist position would argue that all neural processing is potentially conscious, depending on the needs and goals of the system. Clearly, this is a point for future discussion.

Conclusion

This chapter suggests that current cognitive theories have much in common. Almost all suggest an architectural function for consciousness. Although the reader's experience of *these words* is no doubt shaped by feature cells in visual cortex, including word recognition regions, such local activity is not sufficient for consciousness of the words. In addition, some widespread functional brain capacity is widely postulated. Direct functional imaging evidence for that hypothesis is now abundant. In that sense, most current models are globalist in spirit, which is not to deny, of course, that they involve multiple local specializations as well. It is the *integration* of local and global capacities that marks these theoretical approaches. Given the fact that scientists have only "returned to consciousness" quite recently, this kind of convergence of opinion is both surprising and gratifying.

Future work should focus on obtaining neuroscientific evidence and corresponding behavioral observations that can address global access as the distinguishing feature of consciousness. Additional work could contribute simulations of the kind offered by Dehaene, Sargent, and Changeux (2003), supporting the plausibility of all-or-none global propagation of signals as models of the neurocognitive architecture of consciousness, and of Franklin, documenting the real-world potential of global workspace architectures as intentional agents. Further work is also needed to resolve the issue of whether consciousness is all-or-none, as Baars, Freeman, and Dehaene and his colleagues argue, or whether there are multiple drafts (Dennett, 1991) or microconsciousnesses (Zeki, 2001, 2003) playing a role in the architecture of consciousness (see also Chapter 15).

References

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. Spence (Eds.), *Advances in the psychology of learning and motivation: Research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Baars, B. J. (1983). Conscious contents provide the nervous system with coherent, global information. In R. J. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 3, pp. 41–79). New York: Plenum Press.

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Baars, B. J. (1998). *In the theater of consciousness: The workspace of the mind*. New York: Oxford University Press.

Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences*, 6(1), 47–52.

Baars, B. J., Ramsoy, T., & Laureys, S. (2003). Brain, conscious experience, and the observing self. *Trends in Neurosciences*, 26(12), 671–675.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423.

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist*, 56(11), 851–864.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Basmajian, J. (1979). *Biofeedback: Principles and practice for the clinician*. Baltimore: Williams & Wilkins.

Block, N. (2005). Two neural correlates of consciousness, *Trends in Cognitive Sciences*, 9, 46–52.

Blumenthal, A. L. (1977). *The process of cognition*. Englewood Cliffs, NJ: Prentice-Hall.

Cooney, J. W., & Gazzaniga, M. S. (2003). Neurological disorders and the structure of human consciousness. *Trends in Cognitive Sciences*, 7 (4), 161–165.

Damasio, A. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25–62.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79, 1–37.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences USA*, 95, 14529–14534.

Dehaene, S., Sargent, C., & Changeux, J. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Science USA*, 100(14), 8520–8525.

Dennett, D. (1991). *Consciousness explained*. Boston: Back Bay Books.

Dennett, D. E. (2001). Are we explaining consciousness yet? *Cognition*, 79, 221–237.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness*. New York: Basic Books.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Franklin, S. (2001). Conscious software: A computational view of mind. In V. Loia & S. Sessa (Eds.), *Soft computing agents: New trends for designing autonomous systems* (pp. 1–46). Berlin: Springer (Physica-Verlag).

Franklin, S., & Graesser, A. (1999). A software agent model of consciousness. *Consciousness and Cognition*, 8, 285–301.

Freeman, W. J. (2004). Origin, structure, and role of background EEG activity. Part 1. Analytic amplitude. *Clinical Neurophysiology*, 115, 2077–2088 (Including issue cover).

Freeman, W. J. (1991). The physiology of perception. *Scientific American*, 264, 78–85.

Freeman, W. J., & Rogers, L. (2003). A neurobiological theory of meaning in perception. Part V. Multicortical patterns of phase modulation in gamma EEG. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 13(10), 2867–2887.

Gabor, D. (1946, November). Theory of communication. *Journal of the IEE (London)*, 93(26), 429–457.

Hayes-Roth, B. (1984). A blackboard model of control. *Artificial Intelligence*, 16, 1–84.

James, W. (1890). *The principles of psychology*. New York: Holt.

John, E. R., Prichep L. S., Kox, W., Valdes-Sosa, P., Bosch-Bayard, J., Aubert, E., Tom, M., di Michele, F., & Gugino, L. D. (2001). Invariant reversible qeeg effects of anesthetics. *Consciousness and Cognition*, 10, 165–183.

Johnson-Laird, P. N. (1988). A computational analysis of consciousness. In A. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 357–368). Oxford: Clarendon.

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition*, 79, 89–113.

Llinas, R., Ribary, U., Contreras, D., & Pedroarena, C. (1998). The neuronal basis of consciousness. *Philosophical Transaction of the Royal Society, London*, 353, 1841–1849.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Marcel, A. (1993). Slippage in the unity of consciousness. *CIBA Foundation Symposium*. 174, 168–80; discussion, pp. 180–186.

Mandler, G. A. (1975). Consciousness: Respectable, useful and probably necessary. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, NJ: Erlbaum.

McIntosh, A. R., Rajah, M. N., & Lobaugh, N. J. (1999). Interactions of prefrontal cortex in relation to awareness in sensory learning. *Science*, 284, 1531–1533.

Milner, B., & Rugg, M. D. (Eds.), (1992). *The neuropsychology of consciousness*. London: Academic Press.

Newell, A. (1992). SOAR as a unified theory of cognition: Issues and explanations. *Behavioral and Brain Sciences*, 15(3), 464–492.

Newman, J., & Baars, B. J. (1993). A neural attentional model for access to consciousness: A global workspace perspective. *Concepts in Neuroscience*, 4(2), 255–290.

Norman, D. A., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behaviour* (CHIP Report No. 99). San Diego: University of California.

Penfield, W. (1958). *The excitable cortex in conscious man*. Springfield, IL: Thomas.

Posner, M. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences USA*, 91, 7398–7403.

Pribram, K. H. (1971). *Languages of the brain: Experimental paradoxes and principles in neuropsychology*. Englewood Cliffs, NJ: Prentice-Hall.

Pribram, K., & Meade, S. D. (1999). Conscious awareness: Processing in the synaptodendritic web. *New Ideas in Psychology*, 17(205), 214.

Reddy, R., & Newell, A. (1974). Knowledge and its representations in a speech understanding system. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 256–282). Potomac, MD: Erlbaum.

Rees, G. (2001). Seeing is not perceiving. *Nature Neuroscience*, 4, 678–680.

Schacter, D. L. (1990). Toward a cognitive neuropsychology of awareness: Implicit knowledge and anosognosia. *Journal of Clinical and Experimental Neuropsychology*, 12(1), 155–178.

Schneider, W., & Pimm-Smith, M. (1997). Consciousness as a message-aware control mechanism to modulate cognitive processing. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 65–80). Mahwah, NJ: Erlbaum.

Shallice, T. (1972). The dual functions of consciousness. *Psychological Review*, 79(5), 383–393.

Shallice, T. (1978). The dominant action system: An information processing approach to consciousness. In K. S. Pope & J. L. Singer (Eds.), *The stream of consciousness: Scientific investigations into the flow of experience* (pp. 117–157). New York: Plenum.

Shallice, T. (1988). Information-processing models of consciousness: Possibilities and problems. In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 305–333). Oxford: Clarendon Press.

Shanahan, M., & Baars, B. J. (2005), Applying global workspace theory to the frame problem. *Cognition*, 98, 157–176.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 525, 207–222.

Tononi, G., & Edelman, G. (1998). Consciousness and complexity. *Science*, 282, 1846–1851.

Varela, F., Lachaux, J., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Neuroscience*, 2, 229–239.

Zeki, S. (2001). Localization and globalization in conscious vision. *Annual Review of Neuroscience*, 24, 57–86.

Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Science*, 7(5), 214–218.