Journal of Consciousness Studies, **4**, No. 4, 1997, pp. 292–309

IN THE THEATRE OF CONSCIOUSNESS

Global Workspace Theory, A Rigorous Scientific Theory of Consciousness.

Bernard J. Baars, The Wright Institute, 2728 Durant Ave., Berkeley, California 94704, USA. Email: baars@cogsci.berkeley.edu

Abstract: Can we make progress exploring consciousness? Or is it forever beyond human reach? In science we never know the ultimate outcome of the journey. We can only take whatever steps our current knowledge affords. This paper explores today's evidence from the viewpoint of Global Workspace (GW) theory.[1] First, we ask what kind of evidence has the most direct bearing on the question. The answer given here is 'contrastive analysis' — a set of paired comparisons between similar conscious and unconscious processes. This body of evidence is already quite large, and constrains any possible theory (Baars, 1983; 1988; 1997). Because it involves both conscious and unconscious events, it deals directly with our own subjective experience, as anyone can tell by trying the demonstrations in this article.

One dramatic contrast is between the vast number of unconscious neural processes happening in any given moment, compared to the very narrow bottleneck of conscious capacity. The narrow limits of consciousness have a compensating advantage: consciousness seems to act as a gateway, *creating access to essentially any part of the nervous system*. Even single neurons can be controlled by way of conscious feedback. Conscious experience creates access to the mental lexicon, to autobiographical memory, and to voluntary control over automatic action routines. Daniel C. Dennett has suggested that consciousness may itself be viewed as that to which 'we' have access. (Dennett, 1978) All these facts may be summed up by saying that *consciousness creates global access*.

How can we understand the evidence? The best answer today is a 'global workspace architecture', first developed by cognitive modelling groups led by Alan Newell and Herbert A. Simon. This mental architecture can be described informally as a *working theatre.* Working theatres are not just 'Cartesian' daydreams — they do real things, just like real theatres (Dennett & Kinsbourne, 1992; Newell, 1990). They have a marked resemblance to other current accounts (e.g. Damasio, 1989; Gazzaniga, 1993; Shallice, 1988; Velmans, 1996). In the working theatre, focal consciousness acts as a 'bright spot' on the stage, directed there by the selective 'spotlight' of attention. The bright spot is further surrounded by a 'fringe,' of vital but vaguely conscious events (Mangan, 1993). The entire stage of the theatre corresponds to 'working memory', the immediate memory system in which we talk to ourselves, visualize places and people, and plan actions.

Information from the bright spot is globally distributed through the theatre, to two classes of complex unconscious processors: those in the darkened theatre 'audience' mainly *receive* information from the bright spot; while 'behind the scenes', unconscious contextual systems *shape* events in the bright spot. One example of such a context is the unconscious philosophical assumptions with which we tend to approach the topic of consciousness. Another is the right parietal map that creates a spatial context for visual scenes (Kinsbourne, 1993). Baars (1983;1988; 1997) has developed these arguments in great detail, and aspects of this framework have now been taken up by others, such as the philosopher David Chalmers (1996). Some brain implications of the theory have been explored. Global Workspace (GW) theory provides the most useful framework to date for our rapidly accumulating body of evidence. It is consistent with our current knowledge, and can be enriched to include other aspects of human experience.

[1] The easiest introduction is *In the Theater of Consciousness: The Workspace of the Mind* (Baars, 1997). The most complete technical exposition is in Baars (1988). Papers since 1988 are cited in the References.

I: Introduction

You are conscious and so am I. This much we can tell pretty easily, since when we are not conscious our bodies wilt, our eyes roll up in their orbits, our brain waves become large, slow, and regular, and we cannot read a sentence like this one.

While the outer signs of consciousness are pretty clear, it is our inner life that counts for most of us. The contents of consciousness include the immediate perceptual world; inner speech and visual imagery; the fleeting present and its fading traces in immediate memory; bodily feelings like pleasure, pain, and excitement; surges of feeling; autobiographical events when they are remembered; clear and immediate intentions, expectations and actions; explicit beliefs about oneself and the world; and concepts that are abstract but focal. In spite of decades of behaviouristic avoidance, few would quarrel with this list today.

At this instant you and I are conscious of some aspects of the act of reading — the shape of *these letters* against the white texture of the page, and the inner sound of *these words.* But we are probably not aware of the touch of the chair, of a certain background taste, the subtle balancing of our body against gravity, a flow of conversation in the background, or the delicately guided eye fixations needed to see *this phrase;* nor are we now aware of the fleeting present of only a few seconds ago, of our affection for a friend, and some of our major life goals.

These unconscious elements are as important as the conscious ones, because they give us natural comparison conditions. For example: While you are conscious of words in your visual *focus*, you surely did not consciously label the word 'focus' just now as a noun; yet this sentence would be incomprehensible if highly specialized language analysers — located in the cortex of the brain, just above the left ear — did not treat 'focus' as a noun *unconsciously*. The meaning would change significantly if you understood it to be a verb or an adjective. It is revealing to compare these conscious and unconscious aspects of the same word.

On reading '*focus*', you were surely unaware of its nine alternative meanings, though in a different sentence you would instantly bring a different meaning to mind. What happened to the others? A wealth of evidence supports the notion that some of those meanings existed unconsciously for a few tenths of a second before your brain decided on the right one. Most words have multiple meanings, but only one at a time can become conscious. This seems to be a fundamental fact about consciousness.

These examples illustrate the sense of the word 'consciousness' we aim to understand — that is, *focal* consciousness of easily described events, like 'I see a printed page', or 'He imagined his mother's face'. A great body of evidence shows that conscious contents like this can be reported *as conscious* with great accuracy under the right conditions. These conditions include immediate report, freedom from distraction, and some way for the outside observer to verify the report. These are standard laboratory conditions that apply to thousands of experiments in perception, memory, attention and mental imagery. They also fit the demonstrations presented in these paragraphs.

Whenever a question about the meaning of consciousness arises, I would invite you to revisit the paragraphs above. The meaning of 'consciousness' intended here is best illustrated by your own experience. *Verifiable public report* is the key to scientific evidence, but *your experience here and now* is quite a good index to it. All of the

subjective demonstrations used here can be tested objectively, and all the objective facts can be experienced by you and me. That is why we believe we can talk about consciousness *as such*.

The need to find comparison conditions for conscious events

A simple experiment. We can say a word mentally, and then let it fade; for about ten seconds afterwards it can still be recalled. (The reader is encouraged to try this a few times.) Our ability to retrieve the word with accuracy suggests that an unconscious memory of the conscious present must be maintained for a little while, with little loss of content. Thus we have both conscious and unconscious elements in our 'working memory'. Given that a word still exists accurately for a little while even after fading from consciousness, we can ask, 'What then is the effect of our being conscious of a word?' In effect, we have a controlled experiment, allowing us to compare the same word when it is conscious to when it is not.

Dozens of *contrastive analyses* like this can be performed over a range of phenomena. (Baars, 1983; 1988; 1997). They provide the most directly relevant body of evidence about conscious experience. In science, after all, we can only study something if we can treat it as a variable, comparing its presence to its absence. A number of historic breakthroughs in science emerged from the realization that some previously assumed constant, like atmospheric pressure or gravity, was actually a variable. We can make use of this classic scientific strategy to explore consciousness.

Notice that contrastive analysis deals directly with personal phenomenology, as anyone can demonstrate in their own experience. Repeating a word inwardly is one such case. All demonstrations in this paper involve highly reliable personal experiences.

Global Workspace theory is based entirely on well-established empirical contrasts between pairs of conscious and unconscious events. Any adequate theory of consciousness must account for the entire set of such contrasts, which is quite large. The issue is therefore highly constrained — not at all unclear, nor even controversial on the evidence.

II: The Central Puzzle: Conscious Limits vs. Unconscious Vastness

One key puzzle is especially revealing. Consciousness is closely associated with the 'limited capacity' aspects of the brain. Limited capacity mechanisms include immediate memory, the selectivity of attention, and the fact that we cannot do two demanding voluntary actions at the same time. If we look only at these limited mechanisms, the brain seems to be a rather slow, one-thing-at-a-time, error-prone organ. Most psychological models focus on this aspect of the brain.

But when the brain is observed directly it looks quite different. It seems to be a massive collection of neural nets, layers and connections, each specialized in some specific task — analysing visual shapes, maintaining body temperature, or mapping body space. The great bulk of these specialized nets operate all at once, in parallel with each other, as one great society. Their joint capacity is enormous. They are mostly unconscious, and very effective at executing routine tasks. The big puzzle is, why is the conscious aspect so limited, and the unconscious part so vast? What is the meaning of this odd conjunction?

The answer to this central puzzle will provide our basic theoretical metaphor. But first, here is the evidence in a little more detail.

The narrow limits of conscious capacity

1. Working memory

We can only keep about seven separate things in working memory. This is the famous upper bound on immediate memory, the reason why telephone numbers are generally limited to seven digits, and why psychological rating scales never exceed nine or ten steps. The seven plus or minus two limit applies to visual objects, words, numbers, colours, musical notes, and any other set of unrelated elements. It drops even below seven when we cannot rehearse the items, down to about four. In a brain of 100 billion neurons, this upper limit on working memory is fantastically small. A cheap calculator can store more numbers than a human being, in immediate memory. Why, over two billion years of biological evolution, has the brain not evolved a bigger capacity to hold numbers?

The obvious answer is that the human brain simply did not evolve to remember phone numbers, any more than human legs evolved to race along steel railroad tracks. But what is it that the brain has optimized in the puzzling trade-off between conscious and unconscious abilities?

Consider another aspect of limited conscious capacity.

2. Selectivity: consciousness is limited to only one dense stream of input

If we try to take in two dense flows of information, like two simultaneous spoken stories, one to each ear, we can understand only one at a time. That is not limited to language: when two different ball games are shown on the same screen we can also follow only one at any given time. Nor can a driver talk with a passenger while encountering heavy traffic. And when we are preoccupied with a train of thought we cannot be fully conscious of the flow of external events. Conscious involvement with one flow of information will always interrupt another.

This kind of conscious limit is one of the basic facts of life. It is the reason why students who study mathematics for many hours without distraction do better in math than those who do not. It is why the reader of this sentence can be interrupted by competing thoughts or feelings, and lose part of the meaning. At one level it is plain commonsense.

A vast collection of unconscious processes

The brain. Looking directly at the brain we see great orderly forests of neurons, each receiving input from thousands of others via bushy dendritic twigs and branches, and each sending forth its output by way of a single output limb called an axon. Neurons send out electrical pulses 40–1,000 times per second, and they are all active at the same time. A structure like the cortex is an immense, looming starship unto itself, containing by recent estimates 55 billion neurons. The interconnectedness of neurons is also remarkable; we can reach any single neuron in the brain from any other in less than seven steps. Cortical neurons project in vast elegant fibre bundles to the brain clumps nestling tightly underneath; 600 million of them project from one hemisphere to the other, and comparable numbers hang in great loops under each hemisphere to connect distant points between front and back. It is quite a beautiful and regular arrangement. The brain is massively parallel — many things are happening at the same time — largely unconscious in its details, and widely decentralized in any task.

The brain seems to show a 'distributed' style of functioning, in which the detailed work is done by millions of specialized neural groupings without specific instructions from some command centre. By analogy, the human body works cell by cell; unlike an automobile, it has no central engine that does all the work. Each cell is specialized for a particular function according to instructions encoded in its DNA, its developmental history, and chemical influences from other tissues. And the cell is of course the body's basic unit of organization. In its own way the human brain shows the same distributed style of organization.

All this is widely understood today. What is rarely remarked, however, is that *we can create access to any part of the brain using consciousness* (see below, Figure 1). To gain control over alpha waves in the 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 amount of alpha at will. To control a single spinal motor neuron we merely pick up its electrical activity and play it back over headphones; in a half-hour subjects have been able to play drumrolls on their single motor neurons! Of course we are not conscious of the details of control; conscious feedback seems to mobilize unconscious systems that handle the details. Biofeedback control over single neurons and whole populations of neurons anywhere in the brain is well established, and is considered to be 'so ubiquitous that there seems to be no form of Central Nervous System activity . . . or part of the brain that is immune to it' (Buchwald, 1974).

What is odd about this, of course, is that consciousness of feedback from any neural activity is sufficient to establish control, though the great complexity of those activities is entirely unconscious. It is as if mere consciousness of results creates access to complex and unconscious systems that usually run things, free from interference. We come to similar conclusions when we look at psychological evidence. Here are two examples.

1. The mental lexicon, meaning and grammar

In a fraction of a second after you glance at a word in this sentence it is converted into a semantic code, to interpret its meaning, guided by the rules of grammar. Going from word to meaning is believed to require a large, unconscious mental lexicon. The lexicon of educated speakers of English contains about 100,000 words. While we do not use all of them in everyday speech, we can understand each one instantly, as soon as it is presented in a sentence that makes sense.

Words are complicated things. The Oxford English Dictionary, for example, devotes 75,000 words to the many different meanings of the word 'set'. 'Set' can be a verb, noun, or adjective. It can be a game in tennis or a collection of silverware. We can set a value on an antique, or build a stage set. Glamorous people make a jet-set, and unlucky ones a set of fools. The sun sets, but it is certainly not set in concrete. Mathematicians use set theory, and psychologists talk about set as readiness to act or experience something.

Consider your spontaneous experience of each of the following cases of 'set':

1. tennis: set
2. tools: set
3. put: set
4. value: set
5. stage: set

6. jet: set
7. fools: set
8. sun: set
9. theory: set
10. ready: set

Notice how rapidly we change our minds about the meaning of 'set'. In each example you probably had little difficulty understanding the intended meaning, given the first word of each pair to provide context. Simply becoming conscious of 'tools' imparts an entirely different meaning to 'set'.

This ability to combine two separate pieces of information has never been shown for unconscious input (e.g. Greenwald, 1992; MacKay, 1973). It seems that consciousness is required to integrate the meaning of each word pair. Of course everyday language always combines words: that is the role of grammar and semantics, both great unconscious bodies of knowledge. It seems that understanding language — or anything else that integrates separate bits of content — demands the gateway of consciousness. This is another case of the general principle that *consciousness creates widespread access to unconscious sources of knowledge, such as the mental lexicon, meaning and grammar.* We will see this point again.

2. Autobiographical memory

The size of long-term memory is unknown, but we do know that simply by paying attention to as many as 10,000 distinct pictures over several days, we can learn to recognize each of them *without attempting to memorize them*. Stephen Kosslyn writes,

> The capacity of our visual memories is truly staggering; it is so large that it has yet to be estimated. . . . Perhaps the most staggering results are reported by Standing (1973) who showed some of his subjects 10,000 arbitrarily selected pictures for 5 seconds each. . . . His findings showed that there is no apparent upper bound on human memory for pictures. Moreover, with immediate recall, Standing estimated that if one million vivid pictures were shown, 986,300 would be recognized if one were tested immediately afterward; even after a delay, he estimates that 731,400 would be recalled. . . . Standing collected response times when subjects decided whether they recognized pictures; the results suggested the truly astounding conclusion that subjects could search 51,180 pictures per second in long-term memory! (Kosslyn, 1980, p. 129).

Remarkable results like this are common when we probe memory by merely asking people to choose between known and new pictures. Such 'recognition tests' work well apparently because they *re-present the original conscious experience in its entirety*. Here the brain does a marvellous job of memory search, with little conscious effort. You can get an everyday sense of this impressive performance from recognizing a film seen only once, many years ago, with as sudden sense of familiarity. Often we can even predict the rest of the film based on a single scene. It seems that we create memories of the stream of experience merely by paying attention to something; but human beings are always paying attention, suggesting 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 to store a memory of it, and when we experience the same event again, it also helps us to recognize it, even out of millions of memories.

Consciousness is a gateway to vast domains of knowledge and control

In sum, it seems that consciousness gives us vast access to billions of neurons in the brain and body, to the mental lexicon, and to an inestimably large source of autobiographical memories. Mere unaided consciousness may be sufficient to create rapid learning and accurate recognition. Consciousness is also needed to trigger a great number of automatic routines that make up specific actions. All these effects of consciousness are unconscious. Consciousness may be considered as the gateway to these unconscious sources of knowledge.

Global access: an answer to the puzzle of limited capacity?

My little MacIntosh computer has a lifesaving feature for a swamped and poorly organized human user: a global search command. Just by pushing two buttons I can ask for any document by name, or even by content. No more endless searching

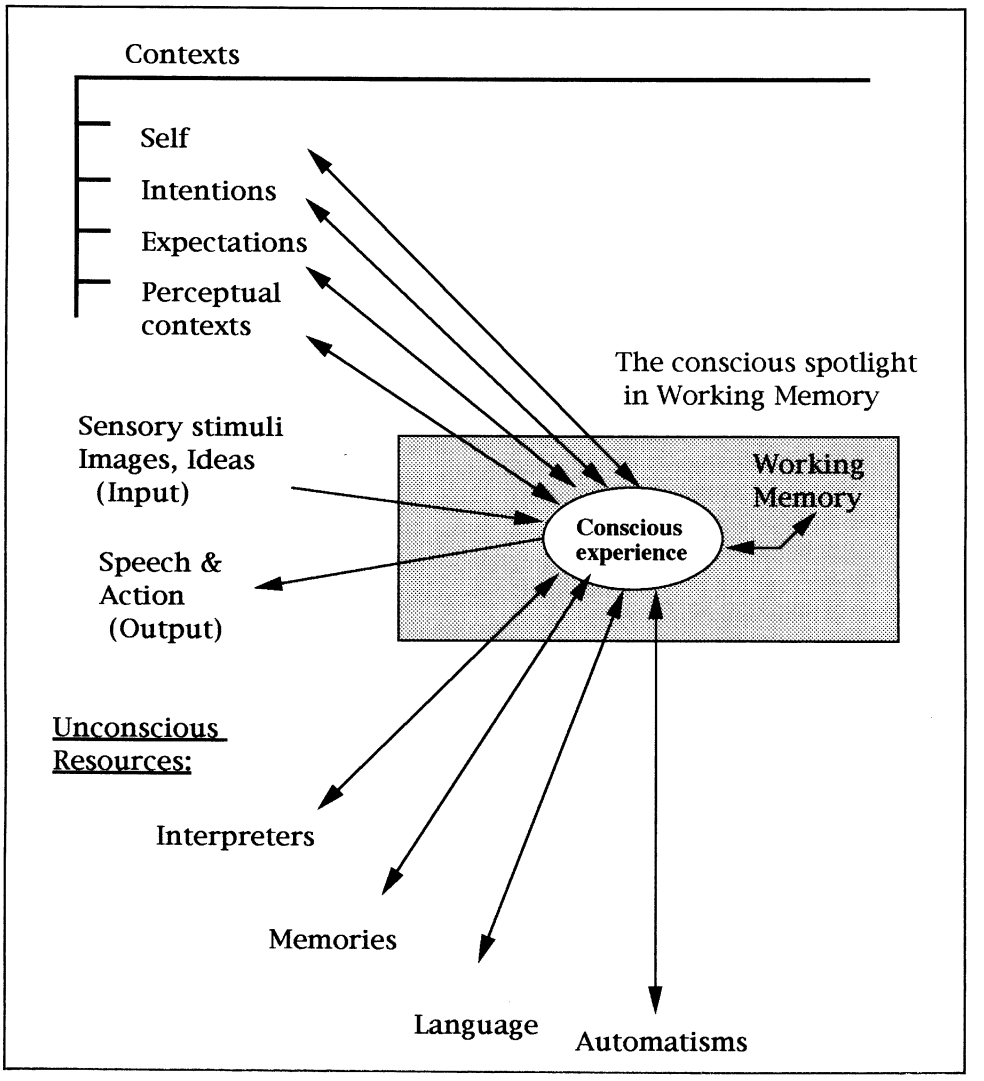

Figure 1. Consciousness creates access

through stacks of folders. It is this capacity to find 'any' name or content that makes global search so useful. While human consciousness involves more than this (cf. footnote 2, p. 17 below), it seems to have a similar capacity to create global access in a brain of 100 billion neurons (see Figure 1).

As we shall see, a 'theatre metaphor' gives us the easiest way to think about this evidence. A theatre combines very limited events taking place on stage with a vast audience, just as consciousness involves limited information that creates access to a vast number of unconscious sources of knowledge. Consciousness seems to be the publicity organ of the brain. It is a facility for *accessing, disseminating and exchanging information*, and for *exercising global coordination and control.*

III: A Working Theatre of Consciousness

In Book VII of Plato's *Republic* (Plato, 1956) we find the following allegory:

> Imagine mankind as dwelling in an underground cave . . . with necks and legs fettered, so they have to stay where they are. They cannot move their heads round because of the fetters, and they can only look forward, but light comes to them from fire burning behind them . . . What do you think such people would have seen of themselves and each other except their shadows, which the fire cast on the opposite side of the cave? (p. 312).

Plato's famous Allegory of the Cave has odd and unexpected resonances down the ages, and indeed in Asian philosophy as well. Remarkably, two and a half millennia later another observer wrote,

> A common metaphor is that of a 'spotlight' of visual attention. Inside the spotlight the information is processed in some special way. This makes us see the attended object or event more accurately and more quickly and also makes it easier to remember. Outside the 'spotlight' the visual information is processed less, or differently, or not at all (Crick, 1993, p. 62).

Plato's point was the fallibility of conscious perception compared to the eternal verities of philosophy, while Francis Crick was aiming to understand the relationship between the brain structure called the thalamus and the great cerebral cortex, both necessary for consciousness. What is the difference between Plato's fire-cast shadows and Crick's thalamic spotlight? I feel moved as much by the similarities as the differences: both are unifying conceptions of human consciousness. In fact, both seem to reflect the same underlying metaphor of our personal experience, the *theatre metaphor*. Plato's Allegory of the Cave may be more elaborate, but Crick's spotlight has a more solid basis in brain anatomy and physiology. A number of cognitive and brain scientists have suggested versions of the theatre metaphor, and in ancient times the same theme was sounded in Vedanta philosophy, at various points in Western thought, and surely by poets and philosophers in many other times and places.

Daniel Dennett and Marcel Kinsbourne have criticized one conceivable theatre model, the 'Cartesian Theatre' in which conscious experience 'comes together in a single point in the brain', much as René Descartes thought consciousness might be located in the tiny pineal gland. Descartes was looking for just *one* dimensionless point where the singular soul might connect with the brain. Dennett and Kinsbourne claim that the Cartesian Theatre cannot work, and I believe they are right. It makes no sense. There is no single point in the brain where 'it all comes together'. But no

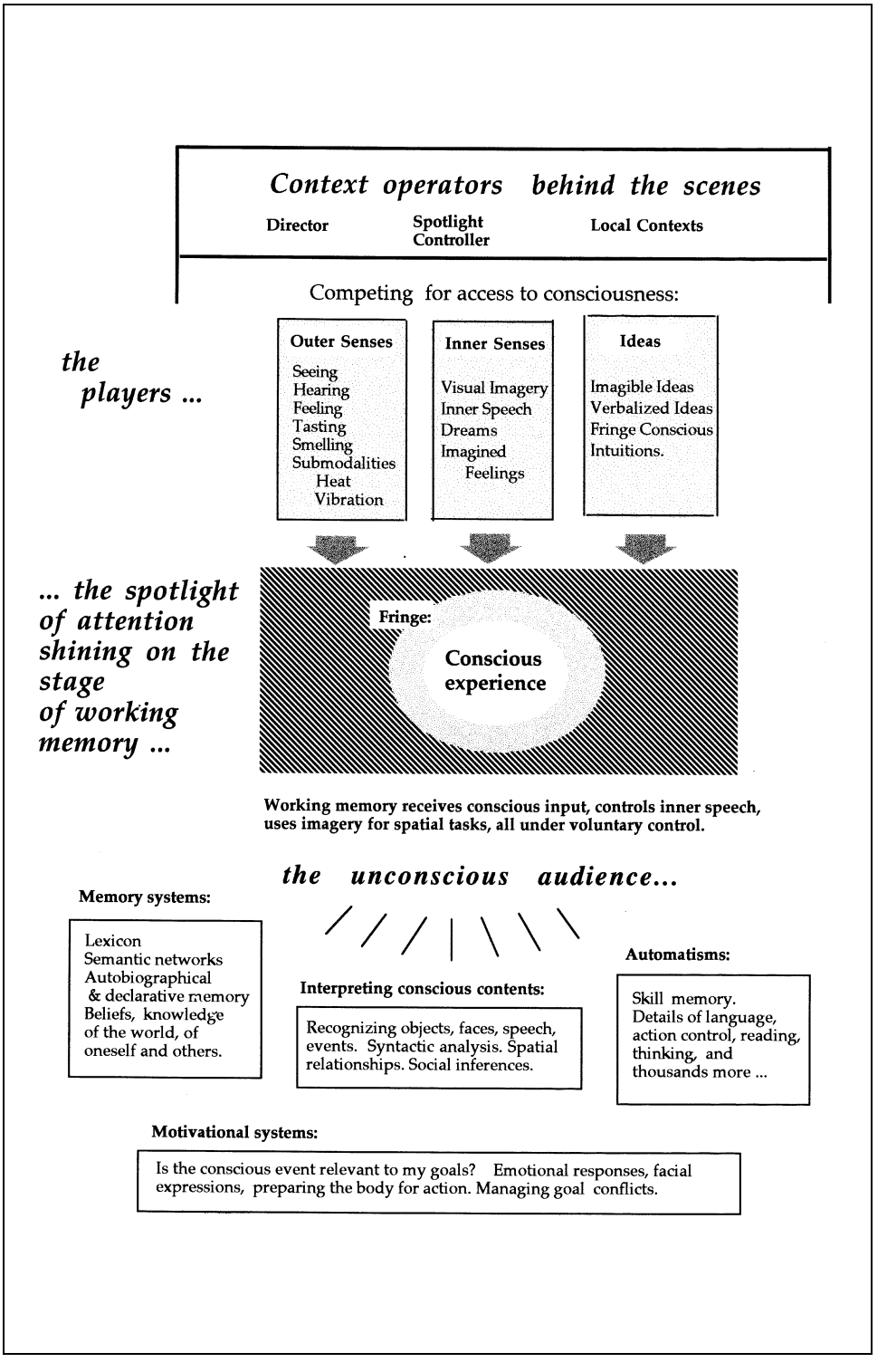

Figure 2. A theatre metaphor for conscious experience

one in science today suggests a Cartesian Theatre. Certainly none of the cognitive theatre models that have been proposed since the 1950s suffer from these defects. Nor do movie theatres converge on a single dimensionless point. Theatres work just fine in the real world, and provide helpful metaphors for exploring human experience. If such a metaphor becomes misleading at some point, we should simply walk away from it and look for something better.

The universality of theatre models

As it happens, all of our unified models of mental functioning today are theatre metaphors; it is essentially all we have. Cognitive architectures developed by Alan Newell, Herbert A. Simon, John R. Anderson and others resemble theatres. All are equipped with working memories that are limited in capacity. All involve 'active' elements, much like the conscious elements of working memory, though without using the word 'consciousness.' And all have large sets of unconscious mechanisms, whether they are called productions, long-term memory, or procedural memory (Newell, 1990; Anderson, 1983). These theories have been developed over the last forty years based on a vast range of evidence, from studies of chess players to arithmetic problem-solving, mental rotation of visual images to action skills. A remarkable group of distinguished scientists have devoted careers to these integrative conceptions of human cognition.

The theatre of consciousness is simple in concept. Figure 2 shows a diagram, with a stage, an attentional spotlight shining on the stage, actors to represent the contents of conscious experience, an audience, and a few invisible people behind the scenes, who exercise great influence on whatever becomes visible on stage.

The stage receives sensory and abstract information, but only events in the spotlight shining on the stage are completely conscious. The actor in the spotlight frets and struts his hour upon the stage, directed by the playwright and director, against a background created by scene setters. These behind-the-scenes influences are *context operators*, unconscious systems that shape conscious events. The spotlight selects the most significant actors on stage, and once lit up, their messages are distributed to an audience consisting of all the unconscious routines and knowledge sources — the vast array of unconscious tools we use to adapt to the world.

Here is perhaps the most important point. Notice that the theatre in Figure 2 shows both *convergent input* and *divergent output.* Onto the stage converge the actors and their speeches, the make-up artists and scene designers, the playwrights, directors and acting coaches. Every dramatic moment, each syllable spoken on stage reflects this convergence of input. Yet as soon as a syllable is pronounced (and this is the aim of the whole enterprise, of course) it floats out to the audience with effects that are largely unknown, but which depend on each listener making of it whatever they will. A message is 'broadcast' globally, but it is interpreted locally in the mind of each audience member. The director and playwright, listening backstage, also take in the global messages to hone the next performance. In sum, there is massive convergence of information onto the stage, but once it has come together there, it spreads divergently to the audience.

Scientific metaphors

Aren't metaphors pretty crude for thinking about difficult scientific problems? Actually, they have a long and honourable history in science, as tools to help us make the leap from the known to the unknown. The clockwork metaphor of the solar system helped astronomers in the 16th century understand the interplay of the planets. William Harvey's discovery that the heart pushed blood through the veins and arteries, much like a pump, has been a useful metaphor for several centuries and it is still used today. Physicists about 1900 found the image of the atom as a tiny solar system a helpful starting point for understanding subatomic structure. Many scientific theories begin as humble metaphors.

Obviously we don't expect to find tiny theatres in the brain. We may, however, be able to find structures that display and disseminate conscious contents. The conscious parts of the brain seem to include the sensory areas of the cortex, perhaps some surrounding areas, supported by a few subcortical structures; and fleetingly, perhaps, 'amodal' cortex, which does not give rise to sensory qualities. Together they may provide the 'stage' for the unconscious audience in the rest of the brain. The theatre concept helps us to think about brain functioning in an interesting way.

Metaphors must be used with care. They are always partly wrong, and we should keep an eye out for occasions when they break down. Yet they often provide the best starting point we can find.

The parts of a theatre

1. The stage of working memory

Here is a little demonstration to help focus again on the facts. Try to stop your inner speech for ten seconds (timing yourself by looking at a clock). I find it impossible to do for more than five seconds or so. It is the simplest possible demonstration, but it shows how dependent we are on the flow of inner speech, which is one dimension of our working memory. Now think about travelling from home to work, remembering to stop off at the supermarket on the way home . . . can you see it? It is difficult not to 'see' concrete spatial descriptions. That is the domain of working imagery. Over the last few decades we have gathered a body of solid evidence about the verbal and visual components of working memory, but one thing has not changed: both components are remarkably limited in their capacity to retain information.

As we pointed out above, we can still retrieve words and images from working memory (WM) after they have faded, for some seconds. Thus WM is mostly in the dark at any single time. Yet we need to be aware of the 'active' elements in working memory, including sensory input, rehearsed and imagined items, items we act on voluntarily (by rehearsing a phone number, for instance), and those we plan to act on. We may say that WM input, output, and manipulated items in WM apparently need to be conscious.

2. The bright spotlight of attention

We can shift at will among the numbers we are keeping in working memory. In imagining the supermarket on the way home, we can focus at will on the various items for sale, the apples or the soap. The contents of consciousness can be guided, both voluntarily and spontaneously, like a bright spot on the stage of working

memory. As we showed before, WM is not entirely conscious. We can add this feature to our theatre metaphor. Let's say that *'the theatre has a powerful spotlight of attention, and only events in the bright spot on stage are strictly conscious'*.

3. The actors trying to get into the bright spot

All working memories show *competition* between different 'actors', the potential thoughts, images or sensations that try to reach the stage (Figure 2). The more conscious involvement is required for any actor, the more it will compete against the others.

This image of an actor trying to gain access to the spotlight of attention is too simple, of course. Sensory systems have a vast range of contents, from a single star on a dark night to a fast-moving ball game in a crowded stadium. There is much evidence to show that the 'actors' can be decomposed into single sensory neurons, and recomposed into complex multimodal events involving millions of sensory cells. Metaphors have their limits, and this is clearly one of them.

4. Context is set behind the scenes

Any experience is shaped by unconscious contexts, just as events on stage are shaped by people behind the scenes. Much attentional selection is spontaneous and unconscious, as if commands from behind the scenes influence the direction of the spotlight. All perceptual systems are shaped by unconscious factors; our visual perception of depth is shaped by the unconscious assumption that light comes from above. Lighting someone's face from below can make it unrecognizable. Above we have shown the difference in our understanding of the word 'set' when it is preceded by 'put' vs. 'tennis'. That kind of context-sensitivity is universal in language, perception, action control, memory, problem-solving, etc.

Even conceptual assumptions can act as unconscious contexts. Each of us is run by beliefs that are unconscious at the time they shape our thoughts and actions. Consider the famous mind–body debate. Ask someone about their own freedom of action, and they will claim some kind of free-will mentalism. Ask them about taking a physical aspirin for a mental headache, and they will smoothly switch to dualism. Ask college students whether their minds can be understood in terms of neurons, and they will probably agree to physicalistic reductionism. Each of these positions involves a cluster of beliefs, most of which are unconscious most of the time. Yet they shape our conscious thoughts every second of the day.

The director. Executive functions seem to be contextual in just this way. They seem to make use of the conscious bright spot, even when they are not conscious themselves. It is believed that human working memory is guided by some sort of executive system that makes decisions guided by goals. The decision to rehearse a telephone number may not be entirely conscious. We rarely have much access to the reasons why we do anything, and when we are forced to guess we are often wrong. Thus it seems that the theatre director also works invisibly behind the scenes. Such executive functions are apparently located in frontal cortex; damage to this part of the brain leads to a predictable loss of ability to guide one's actions by long-term goals.

5. The audience

Consciousness is the gateway to a vast unconscious collection of specialized knowledge. All unified models of cognition today suggest some sort of unconscious audience: It may be called long-term memory or a set of automatic 'productions', but it consists of multiple specialized capacities that are not conscious. We have illustrated three examples above, and we will explore some implications below.

Psychologists have become convinced that the real work in navigating through the problem spaces of our lives is done unconsciously for most of us most of the time. This is a counterintuitive idea. Intuitively we tend to think of our 'selves' as being 'in charge' of our actions, our bodies, even our thoughts. But the nervous system prefers a different style of operating, one that is more *distributed*, in which most of the work is done in a decentralized way, by local processors. Executive control does exist, but it seems to take place by way of those distributed specialized capacities. In saying a word, 'we' have some overall control, but the delicate and subtle execution of speech articulation is largely unconscious.

Some implications

(a) Learning as a magical process. The idea that consciousness is a gateway — something that creates access to a vast unconscious mind — has interesting implications for understanding learning. It suggests that learning just requires us to 'point' our consciousness at some material we want to learn, like some giant biological camera, and the detailed analysis and storage of the material will take place unconsciously. Given a conscious target, it seems as if learning occurs magically, without effort or guidance, carried out by some skilled squad of unconscious helpers.

For two decades cognitive scientists have tried to model expert knowledge in medicine, physics, and computer programming. The big lesson of those years is that expert knowledge is highly *domain specific*: medical knowledge, for example, is so different from computer programming that almost nothing in one domain applies to the other. And yet as human beings we do only one thing with whatever we need to learn: we just bring it to consciousness, and specialized learning somehow occurs. Language activates an utterly different part of the brain than visual events, which are yet different from planning and feeling, action control, learning what foods taste good, and hundreds of other specialized mechanisms for interpreting conscious information. We direct our attention to the formula $x = y + 3$, play with its elements and rules, and somehow, with no *detailed* conscious encoding of the information, we acquire the ability to understand and use it. We learn to see new visual patterns simply by paying attention to a set of X-rays or a series of Flemish paintings. We learn to hear in new ways merely by listening to bird songs or symphonies. Paying attention — becoming conscious of some material — seems to be the sovereign remedy for learning anything, applicable to many very different kinds of information. It is the universal solvent of the mind

How can we think about this with the theatre notion? It seems that all the learning engines are located in the darkened audience of the theatre. Conscious events on stage seems to elicit automatic learning in certain unconscious neural assemblies. It is the audience members that do the learning, just as in the real theatre.

(b) Implicit learning. Children learning language do not label the words they hear as nouns or verbs. Rather, they pay attention to speech sounds, and as a consequence, the underlying regularities are learned *implicitly.* We rarely become conscious of abstract patterns — the regularities of grammar, the harmonic progressions of a symphony, or the delicate brushwork of Vermeer. Most knowledge is tacit knowledge; most learning is implicit.

Nevertheless, even in implicit learning we must be conscious of the material from which we derive the unconscious pattern. A child learning the rules of grammar must listen quite consciously to a long series of spoken sentences. Thus the weight of evidence today suggests that *all learning requires conscious access to what is to be learned.* In theatre terms, it is as if we merely place an actor in the spotlight, and the audience will silently remember his or her speech.

(c) Automatisms steer us through the world. Try to read the following words *without* hearing them in your inner speech:

inchoate
Pappa Doc
infundibulum

The mere visual experience of a printed word seems to trigger automatic inner speech for most of us. A large part of the unconscious consists of complex automatic processes, that are triggered by conscious 'priming' events. Again, it seems as if events in the spotlight of attention automatically trigger complex uncontrolled events that take place in the audience.

(d) Problem-solving functions. Consciousness creates access to unconscious problem-solving. The famous 'incubation process' in mathematics involves a conscious question, unconscious work on the problem, and a conscious emergence of the solution. But we can see the same three-stage pattern in answering everyday questions: 'What is your mother's maiden name?' ' What is 20 x 13?' In each case, there is a brief pause, and then, without conscious work on the question, the answer appears. It is as if unconscious algorithms about one's mother, or about multiplying numbers, are recruited by a conscious appeal for an answer. The theatre analogy is clear: we only need to have an actor proclaim a question, and special problem solvers in the audience go to work to solve the problem without further conscious involvement. When an answer is found, it is often 'returned' to consciousness, as if an audience member mounts the stage to announce the answer.

(e) Priming: consciousness is used to set the context for future events. Remember the list of words (above), each followed by 'set'? Each of those words served to 'prime' a certain interpretation of 'set'. And indeed, the title of this section helped to prime your interpretation of this sentence at this very moment. But of course you are not conscious of the section title at this very moment, because doing that would interfere with what you are conscious of now. One of the remarkable features of conscious experiences is how they can trigger unconscious contexts that help to interpret later conscious events. It is as if some actors have the function of announcing changed circumstances, that will shape our understanding of the next scene, like the witches in *Macbeth*.

(f) Consciousness also creates access for the self. In a deep and brilliant observation, Daniel Dennett (1978) has written that

> That of which I am conscious is that to which I have *access*, or (to put the emphasis where it belongs), that to which *I* have access . . .

'I' have *access* to perception, thought, memory, and body control. Each of us would be mightily surprised if we were unable to gain conscious access to some vivid recent memory, some sight, smell or taste in the immediate world, or some well-known fact about our own lives such as our name. The 'self' involved in conscious access is sometimes referred to as the self as observer. William James called it the knower, the 'I'.

This is of course our common intuition, one that was not seriously doubted until this century, when philosophers like Gilbert Ryle found ways to question it. Empirically we now know that Ryle was wrong. The 'narrative interpreter' found in the left frontal cortex of split-brain patients, does indeed receive conscious sensory information (Gazzaniga, 1993). The narrative interpreter takes in and comments on that information, acts upon it, and is able to describe it; in general it displays just the kind of access to conscious input that one would expect from the everyday notion of self. There may be a complementary but inarticulate 'self' in the right frontal cortex. Without such 'access of the observing self', 'we' cannot obtain information from the world, from memory, or from imagined ideas of the future. Without frontal cortex 'we' cannot exercise voluntary control over skeletal muscles or inner speech. When the observing self is eclipsed by psychogenic fugue or multiple personality disorder, victims report 'time loss' — as if the eclipse in the observing self has also caused consciousness to disappear, for weeks or even months. The observing self seems to be a necessary framework for conscious experience.

IV: What Can We Do with a Working Theatre? The Functions of Consciousness

Theatres are useful. Their fundamental features are found in classrooms, television screens, broadcasting, armies, bulletin boards, bureaucracies and business organizations. It seems that evolutionary biology has discovered the same style of functioning as well — not surprisingly, given the many homologies that exist among animal species, which are often rediscovered in human artifacts.

The evidence suggests that consciousness is a major biological adaptation. If so, it may have not just one but many adaptive functions. The bloodstream carries nutrients, removes metabolic wastes, distributes hormones, transports immune cells, and much more. Major biological adaptations tend to gather multiple functions.

The evidence suggests the following nine functions for consciousness.

1. Adaptation and learning function

The more novelty we encounter, the more conscious involvement is needed for successful learning and problem-solving. In the language of Jean Piaget, the more we must accommodate to new information, the more we need to be consciously involved with it.

2. Definitional and contextualizing function

Every conscious experience is shaped by contextual factors, which are not conscious when they do so. The reader's experience at this very moment is influenced by prior ideas about our topic, now out of consciousness. Content and context are twin issues, ultimately inseparable.

3. Access to a self system

The problem of subjectivity may demand a deeper understanding of how the 'observing self' interacts with conscious contents. The self seems to monitor and control access to conscious events, maintaining continuity across different situations and life circumstances. As described above, 'I' seem to have *access* to perception, thought, memory and body control. There is now a body of evidence that supports this common sense belief. Self and consciousness seem to be mutually necessary systems.

4. Prioritizing and access control function

By persuading smokers that the act of lighting a cigarette is life-threatening over the long term, public health agencies aim to increase smokers' conscious involvement with smoking, intending thereby to create more opportunities for change in a largely automatic habit. By consciously relating some event to higher-level goals (like good health), we can make it conscious more often and thereby increase the chances of successful adaptation.

5. Recruitment and control function

Conscious goals can recruit unconscious routines to carry out actions in new ways, as shown by the extraordinary range of single neurons and neuron populations that can be controlled with conscious biofeedback.

6. Decision-making and executive function

The theatre is not an executive, but just as governments aim to control news media, self-systems located in the frontal cortex probably exercise their control by means of conscious 'publicity'. It is well established, for example, that athletes can improve performance by consciously imagining desired actions. The theatre may also be used to make decisions: an actor can call out a question to the audience, which may then respond with specialized knowledge that is not otherwise available.

7. Error-detection and editing function

Suppose this paragraph contains an error. Often you will become conscious of the error, even though the process that detects it is unconscious. It seems that conscious input is monitored by unconscious systems, which will act to interrupt the flow if errors are detected. The knowledge of what makes an error an error is rarely conscious.

8. Reflective and self-monitoring function

Through inner speech and imagery we can reflect upon, monitor, problem-solve and try to modify our own functioning.

9. Optimizing the trade-off between organization and flexibility

Unconscious, habitual or inborn, 'canned' responses are highly adaptive in predictable situations. However, in unpredictable conditions the capacity of consciousness to organize existing knowledge in new ways is indispensable.

In sum, consciousness appears to be the main way in which the nervous system adapts to novel, challenging and informative events in the world.

V: Summary and Conclusions

An array of evidence is beginning to reveal the role of consciousness in the nervous system, at least in outline. Conscious experience seems to create access to many independent knowledge sources in the brain, most of them quite unconscious. Humans seem to have a larger repertoire of uses for consciousness — including language and long-term planning, self-monitoring and self-reflection, inner speech, metaphor, symbolic representation of experience and deliberate use of imagery. When it comes to sensory consciousness, however, the brain shows little difference between humans and many other mammals.

In the view presented here, *global access* may be a necessary condition for consciousness; but in the nature of science we simply do not know at this time what would be the truly *sufficient* conditions. Some other conditions seem to be necessary, including conscious access to a self system (Baars, 1988; 1997).[2]

For many, the emerging scientific evidence will raise ethical and philosophical implications. This is not the place to explore them. But just as the scientific study of pain perception, for example, has relevance to medical ethics, so does the issue of consciousness in general. That is not to say that a better scientific understanding will somehow dictate values, but rather that people inevitably bring their values to whatever we learn about such a central human concern.

We began this paper by asking whether we can make progress on a traditionally contentious question like the nature of our own experience. The question is answered in the affirmative; but in classic scientific fashion, each answer, no matter how tentative, raises many more new questions.

The book of thought about human experience opens at the very beginnings of human philosophy, with the Greeks and the philosophers of the Himalayas more than two millennia ago. Almost every century since that time has added new pages to this work in progress. Brain science and psychology are only the latest contributors in a long and distinguished lineage. Though they have barely begun their work, there is little doubt even now that we are seeing the beginnings of a new chapter in human thought.

[2] A more complete account seems to require at least five more plausible conditions for conscious experience. They may be added to an extended global workspace architecture (Baars, 1988; 1997). They are:

i. *Internal consistency* of conscious contents at any given moment;
ii. *Informativeness* — conscious input always involves a demand for adaptation on the nervous system (Baars, 1988, Ch. 5);
iii. *Self access:* conscious events require access by a self-system (Baars, 1988; 1997);
iv. *Minimum sensory involvement*: conscious events may require sensory or imaginal events for some minimum duration;
v. *Contextualization*: conscious events are always shaped by invisible prior constraints at every level of representation.

References

Anderson, John R. (1983), *The Architecture of Cognition* (Cambridge, MA: Harvard University Press).

Baars, B.J. (1983), 'Conscious contents provide the nervous system with coherent, global information', in *Consciousness and Self-regulation, Vol. 3*, ed. R. Davidson, G. Schwartz and D. Shapiro (New York: Plenum).

Baars, B.J. (1988), *A Cognitive Theory of Consciousness* (New York: Cambridge University Press). [Available from University Microfilm, Ann Arbor, MI, USA (Tel. 1-800-521-0600, order number 2016099). Second edition in preparation.]

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

Buchwald, J.S. (1974), 'Operant conditioning of brain activity — an overview', in *Operant Conditioning of Brain Activity*, ed. M.M. Chase (Los Angeles: University of California Press).

Chalmers, D. J. (1996), *The Conscious Mind: In Search of a FundamentalTheory* (New York: Oxford University Press).

Crick, Francis (1993), *The Astonishing Hypothesis: The Scientific Search for the Soul* (New York: Scribner's).

Dennett, D.C. (1978), *Brainstorms* (Cambridge, MA: MIT Press/Bradford).

Damasio, A.R. (1989), 'Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition', *Cognition*, **39**, pp. 25–62.

Dennett, Daniel C. (1978), 'Toward a cognitive theory of consciousness', in *Brainstorms* (Cambridge, MA: Bradford/MIT Books).

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

Gazzaniga, M.S. (1993), 'Brain mechanisms and conscious experience', in *Experimental and Theoretical Studies of Consciousness*, CIBA Foundation Symposium 174 (New York: Wiley).

Greenwald, A. (1992), 'New Look 3, unconscious cognition reclaimed', *American Psychologist*, **47** (6), pp. 766–79.

Kinsbourne, M. (1993), 'Integrated cortical field model of consciousness', in *Experimental and Theoretical Studies of Consciousness*. CIBA Foundation Symposium 174 (New York: Wiley).

Kosslyn, Stephen M. (1980), *Image and Mind* (Cambridge, MA: Harvard University Press).

MacKay, D.G. (1973), 'Aspects of a theory of comprehension, memory, and attention', *Quarterly Journal of Experimental Psychology*, **25**, pp. 22–40.

Mangan, Bruce (1993), 'Taking phenomenology seriously: The 'fringe' and its implications for cognitive research', *Consciousness and Cognition*, **2**, (2) pp. 89–108.

Plato (1956), *Great Dialogues of Plato*, trans. W.D.H. Rouse (New York: New American Library).

Newell, Alan (1990), *Unified Theories of Cognition* (Cambridge, MA: Harvard University Press).

Shallice, T.R. (1988), 'Information-processing models of consciousness: possibilities and problems', in *Consciousness in Contemporary Science*, ed. A.J. Marcel & E. Bisiach (Oxford: Oxford University Press).

Standing, L. (1973), 'Learning 10,000 pictures', *Quarterly Journal of Experimental Psychology*, **25**, pp. 207–22.

Velmans, M. (ed. 1996), *The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews* (London: Routledge).

Paper received 14 November 1996.