Metaphors of consciousness and attention
in the brain

Bernard J. Baars

Scientific metaphors have long provided heuristic tools for approaching novel problems. Today, the neurobiology of consciousness and attention is a central concern, presenting formidable conceptual and empirical challenges. Many current ideas fit the broad theme of a theater metaphor; this idea can be worked out in detail, resulting in relevant, testable hypotheses.


Metaphors and analogies have a long history in scientific thought: for example, the Rutherford planetary analogy for atomic structure, the clockwork metaphor for the solar system, and Harvey’s pump metaphor for the heart. A metaphor can be defined as ‘the application of a word or phrase to an object or concept it does not literally denote, suggesting comparison to that object or concept.’

Heuristic metaphors are especially useful when the sciences encounter a topic that has no clear precedent, and this is the case with consciousness and attention. A classical metaphor for consciousness has been a ‘bright spot’ cast by a spotlight on the stage of a dark theater that represents the integration of multiple sensory inputs into a single conscious experience, followed by its dissemination to a vast unconscious audience. In cognitive theory, such a theater stage is called a ‘global workspace’, and implies both convergence of input and divergent dissemination of the integrated content. In this century, features of the theater metaphor have been suggested by neurobiologists from Pavlov to Crick. Indeed, nearly all current hypotheses about consciousness and selective attention can be viewed as variants of this fundamental idea, thus, its pros and cons are worth exploring.

The bright spot metaphor was extended in 1984 by Crick: he proposed a ‘searchlight of attention’ metaphor for thalamocortical interaction, specified in terms of testable hypotheses at the cellular level. As Crick wrote,

> ‘What do we require of a searchlight? It should be able to sample activity in the cortex and/or the thalamus and decide “where the action is”. It should then be able to intensify thalamic input to that region of the cortex, probably by making the active thalamic neurons in that region fire more rapidly than usual. It must then be able to turn off its beam, move to the next place demanding attention, and repeat the process. It seems remarkable, to say the least, that the nature of the reticular complex [of the thalamus] and the behavior of the thalamic neurons fit this requirement so neatly.’

Crick derived four testable hypotheses from this metaphor and, if this was its only use it could be discarded as having done its job. However, Crick suggested that ‘there may be at least two searchlights: one for the first visual area and another for all the rest.’ Thalamocortical searchlights for auditory and somatosensory cortex could be included, perhaps interacting in a mutually inhibitory fashion, so that only one sensory searchlight could be turned on at any time. But humans can be aware of more than sensory inflow; inner speech and visual imagery can compete for access to consciousness. Indeed, recent evidence indicates that inner speech involves speech-production cortex and speech-perception cortex, and that visual-projection areas participate in visual imagery. However, humans also have conscious access to ideas that might involve prefrontal activation.

Conscious contents also influence motor output, involving prefrontal, motor and anterior cingulate cortex. Because all these cortical regions interact with corresponding thalamic nuclei, the searchlight metaphor could generate testable hypotheses about the role of consciousness and attention in all these parts of the brain.

But that is not all: real searchlights are guided to their targets, suggesting executive control, and are useless without an audience to whom the contents in the illuminated spot are disseminated. In the brain, the ‘audience’ could consist of unconscious regions, such as cerebral cortex, hippocampus, basal ganglia and amygdala, that might be activated by conscious contents. The audience for a brain searchlight could also include executive or interpreter systems, such as Gazzaniga’s ‘narrative interpreter’ of the left hemisphere, and other executive regions of prefrontal cortex might receive conscious information. Thus searchlight metaphors do not stand alone, but imply a larger framework: a surrounding ‘theater’.

Cognitive models of memory have a similar set of implications: a working memory whose active items are conscious and reportable, under executive control, with an audience of memory systems to receive its contents. ‘Cognitive architectures’ are large-scale simulations that have been developed since the 1950s and have been used to model a range of behavioral tasks from chess-playing to language comprehension, memory retrieval and decision-making. Cognitive architectures resemble theaters, typically receiving input into a narrow ‘stage’ of working memory, interacting with a large ‘audience’ of semantic networks, automatic routines and memory systems. This theoretical tradition has been qualitatively related to consciousness in a framework called ‘global.
workspace theory\textsuperscript{16-18}. For example, all cognitive architectures treat active elements in working memory as reportable, but reportability is the most widely used operational definition of conscious contents. Elements outside working memory are automatic or in long-term memory, and are therefore unreportable and unconscious. Thus, cognitive architectures seem to reflect the same theater metaphor that is implicit in the searchlight notion.

Theater models are also consistent with proposals for the integration of perceptual features, and for ‘convergence zones’ that combine various inputs into unified neural representations. Damasio\textsuperscript{17} has suggested that consciousness might be associated with cortical convergence zones, and theaters exist to allow numerous convergent influences to shape a coherent performance on stage that is then distributed divergently to the audience. Schacter\textsuperscript{19} notes that conscious or explicit processes involve integration across multiple dissociable subsystems, which is, metaphorically, what theaters are good for. For the widely discussed ‘binding function’ of consciousness involves yet another feature that is compatible with the theater metaphor. Gazzaniga has proposed that conscious experiences involve a ‘publicity organ’ in the ‘society’ of mind, just as a theater allows selected information to be made public\textsuperscript{20}. Finally, a vast unconscious ‘audience’ of specialized neuronal assemblies and routines is almost universal in contemporary thinking about the brain\textsuperscript{21-23}. In all of these proposals, the fundamental function of the theater architecture is to make possible novel, adaptive interactions between the sensory inflow, motor outflow and a range of knowledge sources in the brain.

**The theater metaphor: a misleading concept or useful thinking tool?**

The theater metaphor has encountered criticism from Dennett and Kinsbourne\textsuperscript{4}, who agree that it is implicit in much current thinking, but claim that it is ‘Cartesian’ and misleading. A ‘Cartesian theater’ in their view has a ‘point center’ where all sensory input converges, like the pineal gland in Descartes’ 17th century view of the brain. However, neither Crick’s thalamocortical searchlight nor cognitive architectures propose a single-point center. Rather, all current proposals involve ‘binding’, ‘convergence zones’ or ‘working memories’ for the integration of conscious input. However, Dennett and Kinsbourne maintain that there is no single place in the brain where ‘it all comes together’, as suggested by Damasio, Crick and Koch, and by others. However, recent single-cell studies by Sheinberg and Logothetis\textsuperscript{15} suggest strong convergence of conscious visual-object information in inferotemporal cortex and the superior temporal sulcus in the macaque. Approximately 90% of visual neurons in these areas respond differentially to the conscious but not the unconscious visual flow in a binocular rivalry task. Lower visual levels show low response rates to both conscious and unconscious rivaling input. Because the anterior temporal lobe integrates many visual features into object representations, it might indeed be a place where conscious visual information comes together.

Other philosophical critics maintain that consciousness could not possibly play the role attributed to it by theater hypotheses, because computers can simulate such hypotheses without consciousness. But the brain does many things differently from computers, and few scientists would rely on computers in lieu of direct evidence on the neurobiology of consciousness. Still other philosophers claim that some aspects of consciousness, such as subjectivity, might be inherently inexplicable. But that implies a misunderstanding of the scientific enterprise. The aim of the theater metaphor is to achieve a modest increase in knowledge. We cannot know today whether or not we will eventually understand a problem like subjectivity, although this might become clearer as more plausible hypotheses are tested. In summary, such philosophical challenges do not invalidate a useful thinking tool.

The criteria for productive metaphors are the same as for other scientific ideas: they should help organize existing evidence, yield testable hypotheses and suggest conceptual clarifications. For example, the terms ‘consciousness’ and ‘attention’ are conflated in much current work, but are they the same thing?\textsuperscript{24} An attractive distinction is to limit the term ‘attention’ to selective operations, while applying ‘consciousness’ to events that humans can report. Thus, attention involves the selection of targets for the searchlight to shine on, while consciousness results from illumination of the target. When reading, we do not consciously control eye movements, but we have conscious access to the results of eye movements. Likewise, we might not consciously select a certain conversation at a cocktail party, but we become aware of the results of selective operations. This distinction is already implicit in much research, but it is not applied consistently. In this article, ‘attention’ will be used for selective processes, and ‘consciousness’ for events that can be reported.

In broad terms, the theater metaphor aids the organization of basic evidence, and has yielded new, testable hypotheses.

**Evidence for consciousness**

It has been said that there is a lack of firm evidence about consciousness, but there is a large body of relevant findings; this evidence has often been collected under other headings. Relevant evidence comes from any study that treats consciousness as an experimental variable. Crick has pointed out, for example, that before Livingstone and Hubel\textsuperscript{25}, single-cell studies of visual cortex rarely compared cortical activity in anesthetized and waking animals (F.H.C. Crick, pers. commun.). While previous studies had monitored waking visual processes, direct comparisons that allowed consciousness to be studied as a variable were difficult to find. In 1981, Livingstone and Hubel made history by pinching the tail of an anesthetized cat, thereby waking it up, and observing that this caused visual neurons to fire differently. Pinching the tail of an anesthetized cat while recording neuronal responses is one way to manipulate consciousness, but there are many others. For example, comparisons can be drawn between cortically blind and sighted parts of the visual field; between parietal neglect and normal vision; alertness compared with deep sleep, coma and anesthesia; explicit versus implicit knowledge in normal and brain damaged subjects; subliminal versus supraliminal stimulation; immediate versus long-term memory; the attended and unattended stream in dual-input tasks; and novel versus habituated stimuli or...
automatic skills. In each of these cases consciousness can be treated as an experimental variable. Some knowledge seems so obvious that it is rarely made explicit. We know, for example, that waking consciousness is biologically adaptive. Without it, vertebrates do not feed, mate, reproduce, defend their territory or young, migrate, or carry out any other survival or reproductive activity. Neurophysiologically, consciousness has pervasive effects: its characteristic electrical signature (fast, low-voltage and irregular) can be found throughout the waking brain; and in unconscious states, like deep sleep and coma, slow and coherent waves are equally widely distributed. In these respects, consciousness is not a subtle or hard-to-observe phenomenon: it is hard to ignore.

**The puzzle of conscious limited capacity in a massively parallel brain**

The behavioral and brain sciences have presented remarkably different views. Behavioral experiments on humans are used to study conscious input and voluntary motor output. They seem to show a brain that does fairly simple things, like mental arithmetic, slowly, serially, with many errors and a great deal of interference between tasks. Humans cannot perform two conscious tasks at the same time, such as talking freely while driving in traffic. Competition between such tasks depends on the extent to which they are conscious: the more they become habitual and unconscious (through practice), the less they compete. This suggests that consciousness might be responsible for capacity limits.

In contrast, direct brain observation shows a very different system, with vast, orderly forests of neurons, displaying massive parallelism, mostly unconscious in their detailed functioning and with processing capacities so large that they are difficult to estimate; the processing of any given task seems widely distributed across many brain locations. The neurobiological view of the brain is therefore quite different from the behavioral one: it is not slow, serial, mostly conscious and limited in capacity, but fast, parallel, largely unconscious and with vast capacity.

Both of these perspectives are accurate; the difference is in which aspects of the brain are observed. Until recently, psychological studies have tended to ignore the massive parallelism of the brain, and many neuroscience experiments have paid relatively little attention to the seriality, slowness and capacity limits of the conscious stream.

Given that the brain appears massively parallel, why is the conscious component so limited and serial? Would it not be adaptive to be able to do several conscious things at the same time? Certainly human ancestors might have benefited from simultaneously being able to gather food, watch for predators and keep an eye on their offspring. Yet all tasks that require consciousness compete with each other, so that only one can be done well at any given moment. These drawbacks suggest a biological tradeoff. The nervous system might be organized so that tasks at the same time, such as talking freely while driving in traffic, cannot be performed.

**Consciousness creates access**

Consciousness, although limited in capacity at any single moment, appears to offer a gateway to extensive unconscious knowledge sources in the brain. There is much behavioral evidence for this claim. Consider autobiographical memory, which is believed to involve the hippocampus: the size of long-term episodic memory is unknown, but we do know that 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. Remarkable results like this are common when we use recognition probes, that is, asking people to choose between known and new pictures. Recognition probes appear to work because they reinstate the original conscious experience of each picture. With this kind of retrieval the brain does a remarkable job, with little effort. It seems that humans create memories from the stream of perceptual input merely by paying attention, but because we are always paying attention to something, this suggests that autobiographical memory could be very large indeed. Mere consciousness of some event appears to help to store a recognizable memory of it, and when we experience it again, we can distinguish it accurately from millions of other conscious experiences; both episodic storage and retrieval seem to require consciousness.

Another example is the vocabulary of educated English speakers, which contains about 100,000 words. Although we do not use all these words in everyday speech, we can understand them. Each vocabulary item is already quite complex: for example, the Oxford English Dictionary devotes 75,000 words to the many different meanings of the single word ‘set’, but all we need to access such complex unconscious domains of knowledge is to become conscious of a word. Conscious exposure to any printed word on this page is sufficient to access its meaning, syntactic role, inner speech phonology, emotional connotations, semantic and sound associates and imagery components, and to trigger automatic inferences.

Understanding words seems to require the gateway of consciousness. The ability to access unconscious knowledge is conscious also applies to the vast number of automatisms that can be triggered by conscious events, including the automatic inner speech that often accompanies reading: automatic inferences in social judgments; and the automatic transformations of visual patterns on this page into letters, words and phrases. None of these automatisms are conscious in any detail, yet they are triggered by conscious events. This triggering function is hampered when conscious input is degraded by distraction, fatigue, somnolence, sedation or low signal fidelity.

Indeed, it appears that humans can access a great range of brain functions by way of conscious sensory feedback. No one knows directly which groups of vocal-tract muscles they use to say a word, but by way of conscious sensory feedback a wide variety of vocal parameters are controlled. Conscious feedback seems to create spectacular access not only to skeletal muscles, but also, in the short term to autonomic musculature. Biofeedback control of single neurons and populations of neurons almost anywhere in the brain is well established. To gain control over a single spinal motor unit we monitor its electrical activity, amplify it and play it back over headphones; in half an hour subjects have been able to play drumrolls.
using a single motor unit isolated from adjacent units. To gain control over alpha waves in occipital cortex we merely sound a tone when alpha is detected in the EEG, and shortly subjects can learn to increase the amount of alpha at will. Consciousness of sensory feedback appears to be a necessary condition for the establishment of biofeedback control, although the neural activities themselves remain entirely unconscious. It is as if consciousness of results creates access to unconscious neuronal systems that are normally inaccessible and autonomous.

**Testable hypotheses**

Some anatomical structures could function like the basic elements of a theater. They might integrate, shape, display and disseminate conscious contents, to be received and analyzed by other brain structures, and to receive feedback from them.

**Convergence zones: the ‘theater stage’**

Sensory projection areas of the posterior cortex might provide one kind of ‘theater stage’, when ‘lit up’ by attentional activation, thus displaying coherent conscious information to be distributed frontally and subcortically. In the case of visual consciousness, the first cortical projection area, V1, is an essential structure, whose lesioning leads to blindsight, that is, visual knowledge without visual consciousness. Higher visual lesions lead to selective impairment of conscious motion, color or objecthood, and thus we must include the brain areas V1–V5 and finally, IT (inferotemporal cortex) for multiple levels of visual content. Recent single-cell work by Logothetis and colleagues strongly suggests that fully integrated, conscious visual information does not emerge until the anterior pole of the temporal cortex is reached. This can be explained by the neurons in these areas responding to whole objects, combining information from previous levels. The sensory projection areas for audition and the body senses could play similar roles: even abstract conscious contents, such as meaningful ideas, often appear to be mediated by sensory indices such as words, images and sensory metaphors. Recent functional magnetic resonance imaging (fMRI) work suggests that the left prefrontal cortex might play a crucial role in semantic access. When conscious contents are distributed frontally, the anterior cingulate cortex, including the anterior cingulate, which seems to ‘light up’ during tasks that require effortful attention.

**Multiple theater stages.** If each sensory area has its own kind of consciousness, in addition to abstract and voluntary kinds of conscious involvement, how do we cope with not just one, but five or more theater stages, over which the spotlight of attention can play? One hypothesis is that the spotlight of attention can switch from visual to auditory, somatosensory, abstract or voluntary cortex in multiples of 100 ms steps. Such an arrangement would make it possible for several ‘stages’ to operate together. Each one could broadcast widely to the audience of unconscious networks as soon as the spotlight touches on it. There are other ways to get multiple global workspaces to co-operate and compete, but this is a testable first hypothesis.

**Inner speech, imagery and working memory.** Both auditory and visual consciousness can be activated internally as well as externally. Inner speech is a particularly important source of conscious auditory-phonemic events, and visual imagery is useful for solving spatial problems. They are often taken as the two basic components of cognitive working memory, and are now known to involve corresponding sensory cortex. Internally-generated somatosensory imagery reflects emotional and motivational processes, including feelings of pain, pleasure, hope, fear and sadness.

**Selective attention ‘searchlight’ control**

How are conscious contents selected? The thalamus is ideally suited for controlling sensory traffic to cortex and, among thalamic nuclei, the reticular nucleus is known to exercise inhibitory modulation over the sensory nuclei. This is indeed an expansion of Crick’s 1984 proposal for visual attention. The reticular nucleus operates under dual control of frontal executive cortex and automatic interrupts from areas such as the brain stem, emotional centers like the amygdala and limbic cortex, and pain systems. It is these attentional interrupt systems that presumably allow significant stimuli like one’s own name to ‘break through’ into consciousness in a selective listening task, when the name is spoken in the unconscious channel. Interrupt control is quite separate from frontal executive (voluntary) control. Posner suggests that effortful visual attention operates through the anterior cingulate cortex.

**Receiving regions: the ‘audience’**

Which brain regions receive conscious information? We have already listed some possibilities. Consciousness seems to be needed to access at least four bodies of unconscious knowledge: (1) autobiographical memory, which is believed to require the hippocampus; (2) the lexicon of natural language, thought to involve speech perception areas of both hemispheres; (3) automatic routines that control actions, requiring motor and prefrontal cortex, basal ganglia and cerebellum; and (4) the detailed firing of neurons and neuronal populations by way of sensory feedback. In addition, (5) the amygdala is also known to receive information about visual facial expressions, (6) Area 46 of the prefrontal cortex contains another visual map, and neurons in this area are believed to support one kind of working memory.

**Broadcasting of selected contents: ‘speaking to the audience’**

How is conscious information disseminated? Sensory conscious events from posterior cortex might be broadcast frontally and subcortically. Because there are many spatial maps throughout the brain, the ‘trade language’ of the brain could consist of activated maps co-ordinated by temporal oscillations. High fidelity is important to such broadcasting, which implicates the ‘labeled line’ system of the brain. Labeled line fibers emerging from posterior sensory cortex include corticocortical axon bundles, the arcuate fasciculi and the posterior portions of the corpus callosum. A second major system of high-fidelity transmission operates via the thalamus, including the mediodorsal nuclei that project to prefrontal cortex.

Labeled-line fibers also connect to subcortical structures, including the limbic brain, hippocampi, amygdalae and basal ganglia, all of which are known to have precise spatial maps. Because such connections are typically bidirectional, it seems plausible that labeled line tracts establish activation loops, lasting for up to tens of seconds. Significant conscious events can be renewed by inner speech, by visual imagery, or by conscious emotional feeling states, thus re-initializing...
such activity loops. Storage of such activated information in long-term memory might occur via NMDA synapses\(^{40,41}\).

**Unconscious systems that shape conscious events: ‘backstage’**

How are conscious contents shaped? Behind the stage in a theater are many people who shape and influence the performance without themselves being visible: they include the playwright, makeup artists and the stage director. There are analogous ‘contextual’ systems in the brain that shape conscious contents while being unconscious. In the visual system, sensory contents seem to be produced by the ventral visual pathway, whereas unconscious contextual systems in the dorsal pathway define a spatial object-centered framework within which the sensory event is defined. There is a major difference between damage to content regions compared with contextual areas: in the case of lesioned content systems such as the ventral pathway, the subject can generally notice a missing aspect of normal experience; for damaged context systems, one no longer knows what to expect. Without a spatial framework for vision, it is hard to define what might be missing. This might be why parietal neglect is so often accompanied by anosognosia, a massive loss of knowledge about one’s body space\(^{42}\).

**Narrative observer and executive systems: the ‘stage director’**

How do conscious events influence decision-making and motor control? Gazzaniga\(^9\) describes conditions under which split-brain patients encounter conflict between right and left hemisphere functions. Such patients often use the left hemisphere to talk to themselves, sometimes attempting to force the right hemisphere to obey its commands. When that proves impossible, the left hemisphere might rationalize or reinterpret events. The left-brain ‘narrative interpreter’ receives its own sensory inflow from the right visual field, so that it ‘observes’ a conscious flow of visual information. The right hemisphere might have a parallel executive interpreter that observes its own conscious flow from the left visual field. Although the right-brain observer does not speak, it might be able to deal better with anomaly via irony, jokes and other emotional strategies. Each interpretive system can control its own voluntary motor functions and thus there is an obvious analogy with a stage director, who observes events on stage and orders changes where needed. It is possible that full consciousness does not exist without the participation of such self systems, which might be centered in prefrontal cortex.

**Concluding remarks**

Many proposals about brain organization and consciousness reflect a single underlying theme that can be labeled the ‘theater metaphor’. In these views the overall function of consciousness is to provide very widespread access to unconscious brain regions. Such access is needed for global activation, co-ordination and control. The theater metaphor yields testable hypotheses about perceptual binding, thalamocortical integration, working memory and selective attention, multimodal convergence zones, aspects of hemispheric specialization, and much more.

**Selected references**


**Erratum**

In the Book Review by David P. Carey of ‘Consciousness Lost and Found’ by Lawrence Weiskrantz, which was published in the January issue of TINS (Vol. 21, pp. 49–50), the reference list was incorrect. The correct reference list is as follows:


We apologize to the author and readers.