# The Epistemology of Evidence in Cognitive Neuroscience

William Bechtel
Department of Philosophy
University of California, San Diego

## 1. The Epistemology of Evidence

It is no secret that scientists argue. They argue about theories. But even more, they argue about the evidence for theories. Is the evidence itself trustworthy? This is a bit surprising from the perspective of traditional empiricist accounts of scientific methodology according to which the evidence for scientific theories stems from observation, especially observation with the naked eye. These accounts portray the testing of scientific theories as a matter of comparing the predictions of the theory with the data generated by these observations, which are taken to provide an objective portrayal of reality.

One lesson philosophers of science have learned in the last 40 years is that even observation with the naked eye is not as epistemically straightforward as sometimes assumed. What one is able to see depends upon one's training: a novice looking through a microscope may fail to recognize the neuron and its processes (Hanson, 1958; Kuhn, 1962/1970).[1] But a second lesson is only beginning to be appreciated: evidence in science is often not procured through simple observations with the naked eye, but observations mediated by complex instruments and sophisticated research techniques. What is most important, epistemically, about these techniques is that they often radically alter the phenomena under investigation. Moreover, the exact nature of this alteration is frequently poorly understood (Golgi staining is a extreme but illustrative example—100 years after the introduction by Camillo Golgi of the silver nitrate stain, we still do not understand why it binds to only a few neurons in a preparation, which is the very feature which has made it so useful in neuroanatomy). The fact that evidence is generated from altered phenomena, where the nature of the alternations is often not well understood, raises a serious question: to what degree is what is taken as evidence just the product of the alteration or in what respects does it reflect the original phenomena for which it is taken to be evidence? When scientists raise the objection that purported evidence is an artifact, they are claiming that it does not reflect the original phenomena in the intended manner but rather is a product of the mode of intervention.

---

[1]The idea that perception, the recognition of objects and events, is indirect and rests on "unconscious inferences," was clearly articulated in the 19th century by Helmholtz. A number of perceptual theorists (for example, James Gibson) have taken issue with the reference to "inferences" in perception, but the development of a neurobiological understanding of how the visual system operates clearly supports the idea that there are many processing steps between the registration of light on the retina and the recognition of the object or events seen. Accordingly, many of the same issues that arise in evaluating instruments arise in perception itself. I have argued elsewhere (Bechtel, 2000) argues for parallels between the way we settle the occasional but rare disputes about what is visually perceived and the disputes between scientists over whether an instrument has generated an artifact.

One reason philosophers until recently have not attended to the epistemic issues surrounding evidence is that they have focused on relatively established sciences. By the time scientific theories are encoded in textbooks, the controversies about the evidence adduced for them have generally been resolved and the techniques and methods have come to be treated as "black boxes" (Latour, 1987). (This does not mean that the interventions required to procure the evidence are well understood, as the case of the Golgi stain makes clear.) To appreciate the contentious nature of evidence, it is best to focus on the period when new techniques are first being introduced. At this stage charges of artifact and reasons for suspecting artifact appear frequently in the scientific literature, as do defenses against these charges.

The worries about artifacts point to a new subfield of epistemology that I refer to as the *epistemology of evidence*. The challenge for the epistemology of evidence is to understand how the instruments and techniques for producing new evidence are themselves evaluated. If researchers had independent evidence about how the techniques worked, then we would have a regress, but the procedure of epistemically evaluating evidence would be comparable to that of evaluating theories. As already noted, that is generally not the case. New instruments and techniques are introduced and relied on well before there is a theoretical understanding of how they work.

If it is not to an understanding of how the evidence is generated by the instruments and procedures, where do researchers look to evaluate whether their evidence is reliable. The somewhat surprising answer is: to the evidence itself. In general, there are three factors they focus on: (1) whether the instrument or technique is producing repeatable and well-defined results that exhibit a definite pattern, (2) the degree to which the results from one instrument or technique agree with results generated with other techniques or instruments, and (3) the degree to which the purported evidence coheres with theories that are taken to be plausible (Bechtel, 1995, 2000).

Discussions of scientific methodology often emphasize the importance of agreement of the evidence produced through use of one instrument and technique with that generated by other, generally already established, techniques. But there are two reasons to be suspicious as to whether this requirement plays as central a role as is sometimes suggested. First, the main interest in new instruments and techniques is because they produce evidence not obtained by other methods. In just these areas of greatest interest one cannot invoke agreement with other already accepted methodologies. Hence, at most the comparison with results of existing techniques represents a minimal check on the reliability of the new technique. Second, relating the evidence generated with one instrument or technique to that produced by another is frequently not straightforward but depends upon a complex set of inferences. Researchers have to develop ways to bridge between techniques, often by modifying the new technique until it produces results appropriately aligned with the older technique.

The fact that researchers have to work to bring a new technique into correspondence with results from applying more established techniques points to a different way to understand the attempt to relate new techniques with older ones—it serves as a means of calibrating the new technique. I will illustrate this in what follows. But beyond mere calibration, there is another important

consideration about the use of multiple techniques. Frequently no one technique is able to answer the questions researchers ask and evidence must be acquired by piecing together results obtained by different techniques (Bechtel, 2002). This too will be illustrated below.

I will develop this analysis of the epistemology of evidence by focusing on three of the most important sources of evidence employed in cognitive neuroscience—lesion, single-cell recording, and neuroimaging studies. The techniques and instruments discussed here are all designed to reveal the operation of mechanisms. Mechanisms are systems of component parts, each of which performs a contributory operation; these operations are coordinated into the production of an overall activity. A brain mechanism, for example, might be involved in analyzing what object or event is seen or in encoding information into long-term memory. The components will be different brain regions (systems of neurons) which carry out specific information processing operations (e.g., analyzing an object's shape). These components are typically spatially and temporally organized so as to coordinate the operation of the part to produce the activity of the whole (Bechtel, under development; Bechtel & Richardson, 1993; Machamer, Darden, & Craver, 2000).

At the core of understanding how a mechanism produces a given behavior is a decomposition of the mechanism. There are two quite different ways of decomposing a mechanism—functionally into component operations and structurally into component parts. *Localization*, as I use the term, refers to linking a component operation with a component part. It is important to emphasize what is being localized. It is not the overall cognitive performance, but the operations in terms of which the performance is to be explained. In the context of neuroimaging, Petersen and Fiez emphasize this point about what is to be localized; they distinguish *elementary operations* from *tasks* and argue that operations but not tasks are not likely to be localized in the brain:

> ". . . elementary operations, defined on the basis of information processing analyses of task performance, are localized in different regions of the brain. Because many such elementary operations are involved in any cognitive task, a set of distributed functional areas must be orchestrated in the performance of even simple cognitive tasks. . . . A functional area of the brain is not a task area: there is no "tennis forehand area" to be discovered. Likewise, no area of the brain is devoted to a very complex function; "attention" or "language" is not localized in a particular Brodmann area or lobe. Any task or "function" utilizes a complex and distributed set of brain areas" (Petersen & Fiez, 1993, p. 513).

Historically, however, research directed at discovering a mechanism often begins by localizing whole tasks or activities in one component of a system (localizing an area for articulate speech, as discussed below). Such efforts, while not generating an understanding of the mechanism directly, often play an important heuristic role. When, for example, subsequent research points to additional areas involved in performing the task, then one can begin to decompose the task into simpler operations that then are coordinated into the performance of the overall task (Bechtel & Richardson, 1993).

Before turning to the specific instruments and techniques of cognitive neuroscience, there is one additional general issue that should be noted. In trying to understand cognitive functions in the

brain, neuroscientists perform research on a variety of species. Sometimes the decision to work with a given species is driven by ethical consideration (invasive studies with humans and increasingly with higher primates, unless motivated by therapeutic needs, are deemed morally unacceptable). Other times considerations of the ease of working with a organism or the ability to secure reasonably clean data drive the decision. But the fact that different species are employed means that connections need to be made between the brains of members of different species (or between brains of different members of a species as differences are encountered there as well). Often these connections must be established indirectly in terms of the common pattern of results. For example, one way to identify human analogues of brain areas found in monkey studies is to find in neuroimaging areas that respond to the same type of stimulus for which responses are found through single-cell recording in the monkey. Although there is not space here to analyze them, such modes of calibration between species raise additional epistemic issues about the evidence that is put forward.

## 2. Deficits and Lesions

One of the oldest approaches to identifying the function of brain regions is analysis of the deficits resulting from lesions (localized damage) to those regions. Lesions can originate either from illness or injury or from neuroscientists actually destroying neural tissue.[2] Whatever the source of the lesion, the goal of this approach is to identify a psychological deficit associated with it and to infer from that what contribution the damaged area made to normal psychological function. This is an extremely difficult inferential task, as we shall see.

The classical example of using a naturally occurring lesion to make inferences about operations in the normal human brain was Paul Broca's study of a patient named LeBorgne. When Broca, already a highly regarded surgeon and founder of the Société d'Anthropoligie, first encountered LeBorgne in April 1861 he had already been hospitalized for 20 years. His initial complaint was a loss of the ability to speak although he retained the ability to make oral sounds. Among the few words he could utter was *Tan*, which he frequently employed as an epithet, and became the name by which he is often known. Subsequently, LeBorgne lost sensitivity to the right side of his body and developed paralysis. Finally, gangrene developed in his paralyzed right leg, the complaint for which Broca was brought in.

By the time Broca encountered LeBorgne a number of researchers had proposed that areas in the frontal lobe of the brain were involved in speech and language. Several decades earlier the phrenologist Franz Joseph Gall had proposed a frontal localization of speech, but his evidence,

[2] Such deliberately destructive studies are not permitted on human subjects except in what is intended as therapy. Physicians may excise brain tissue to remove tumors or to stop epileptic seizures. Typically, they will only excise tissue not thought to be playing a critical cognitive roles. But often the appraisal that a brain region is not performing a critical task is due to current ignorance. One of the best known subjects of cognitive neuroscience research is a patient known as HM. William Scoville removed major portions of his medial temporal lobe in a successful attempt to relieve intractable epilepsy. Tragically, HM was no longer able to remember any events occurring in his life after the surgery or for several years before (Scoville, 1968). His deficit resulted in a multi-decade attempt to understand the contribution of the medial temporal lobe, including the hippocampus, to memory encoding.

correlating cranial shape with the degree of development of a trait, was widely regarded as suspect by serious scientists and his resulting maps of the brain dismissed as artifacts. Subsequently Jean-Baptiste Bouillaud argued on the basis of anatomical data from a number of patients that speech was a frontal lobe function. Bouillaud was a highly regarded physician (Dean of the Faculty of Medicine in Paris and President of the Académie de Médecine), but despite his general reputation, his claim to localize speech was regarded as suspect. In addition to a general prejudice against localization in the wake of Gall, there was a conviction, stemming from Descartes, that there was a unity to the mind so that even if it was associated with the brain, it was not broken into different faculties localized in different brain areas. These doubts were buttressed by empirical evidence from other researchers who identified patients with frontal damage that did not show speech deficits and patients with speech deficits but no frontal lobe damage. Lacking even a reliable pattern, the localization claim was rejected as an artifact.

Days before Broca encountered LeBorgne, Simon Alexandre Ernest Aubertin, Bouillaud's son-in-law, had defended the frontal hypothesis for speech at a meeting of the Société d'Anthropoligie and had been severely criticized by Pierre Gratiolet. Broca was himself attracted to the hypothesis of a frontal area responsible for speech, and proposed on encountering LeBorgne to treat him as a critical test of the hypothesis. After LeBorgne died a week later, Broca performed an autopsy that revealed a massive lesion centered on the third frontal convolution of the left hemisphere, which he argued was probably the initial site of damage when LeBorgne's deficits were limited to the inability to speak. Broca went on to argue that this region was the locus of articulate speech (Broca, 1861).

Broca subsequently marshalled evidence from a number of other patients to support his claim, and refined it to maintain that control of speech was typically localized in the left frontal cortex.[3] These additional patients, together with those identified by Bouillaud, help establish a definite pattern of results, suggesting that they were not just artifacts. However, the pattern was not clean-cut as there continued to be reports of patients with speech deficits without frontal lesions and patients with frontal lesions without speech deficits. Over time the plausibility of the assignment of articulate speech, or of some information processing operation related to speech, to this area, which came to be known as *Broca's area*, increased, in part due to its close proximity to areas identified as controlling motor activities involved in articulation. The deficit from a lesion in Broca's area also came to be woven into a broader account. Two decades after Broca's work, Carl Wernicke described a different pattern of language deficit, one apparently affecting comprehension of language following lesions in a part of the temporal cortex known as *Wernicke's area*.

During the first half of the 20th century localization claims were generally suspect, but the tradition of Broca and Wernicke was revived in the 1950s by Norman Geschwind, who proposed a model of reading aloud in which information is first processed in the visual areas at the back of the brain, passed to Wernicke's area where it is comprehended, then to Broca's area where speech is planned, and finally to the adjacent motor areas which direct the actual production of

[3] For an account of Broca and the links between his work and Bouillaud, Aubertin, and Gustave and Marc Dax, see (Finger, 2000, chapter 10).

speech. In the decades since the specific assignments of information processing operations to these areas has been repeatedly challenged (but not the claim that they perform operations related to language) and additional areas have been identified that figure in language processing (Bates, 1994). The conception of language processing as involving a (quite large) number of brain areas performing different operations is now quite well established, with most of the evidence coming from analysis of patients with lesions (Bechtel, 2001b).

Lesion research raises a number of epistemic challenges. One is determining precisely what areas of the brain are injured. Until the recent introduction of imagining technology, one could only determine what areas in the human brain were damaged after the person died and an autopsy was performed. By then, though, the range of damage may have extended. This required Broca (1861), for example, to engage in protracted argument as to where the deficit was localized when LeBorgne's deficit involved only loss of articulate speech. Even with techniques for structural neuroimaging, there is still uncertainty about precisely what brain region was destroyed since brain areas do not have well delineated boundaries (Mundale, 1998).[4]

Perhaps the greatest epistemic challenge in using lesions and deficits to understand brain operation is to infer precisely what the damaged component had contributed to normal function. The most general inference is that the damaged area was in some way necessary to the normal performance (e.g., inferring from the fact that lesioning the hippocampus or hippocampal region results in anterograde amnesia, to the conclusion that it is necessary for encoding new semantic memories). One problem for this interpretation is that the brain is a constantly adapting structure so that once an area is removed, processing in other areas is also altered. When such plasticity leads to "recovery" of function, claims as to the necessity of the component that is removed are challenged.[5]

Typically, however, the goal is not just to learn that a region was necessary for a function, but what it contributed to the performance of the function—that is, what elementary operation it performed. This is extremely challenging. One can gain an appreciation of the challenge involved by considering how one might go about trying to understand how a radio (or a computer) operates by selectively removing parts and examining the resulting performance. As Richard Gregory (1968) notes, removal of a transistor from a radio may cause it to hum, yet it would be a bad inference to assume that the removed transistor was the hum suppressor. To begin to identify the operations that the damaged part actually performs one needs to shift focus from the deficits manifested in the damaged system to the process of accomplishing the activity of the normal system through carrying out simpler operations. In the case of products like

[4] See Small (19xx) who shows that the areas identified by different researchers to be Broca's area or Wernicke's area vary significantly and sometimes do not even overlap.

[5] A recently developed technique provides a strategy both for controlling more precisely the site of a lesion and preventing reorganization of processing after the lesion. This involves inducing temporary lesions through transmagnetic stimulation; it involves application of a strong but localized magnetic field at the scalp so as to disrupt the activity in the immediately underlying brain regions. Early reports (Walsh & Cowey, 1998) indicate that one can disrupt very specific functions. If so, it will allow researchers to set side these worries and focus more directly on the question of what the affected area contributed to normal function.

radios, engineers designed the mechanisms from knowledge of what operations components could perform. But in analyzing natural systems this perspective of the engineer is not available, and researchers must hypothesize what operations might figure in the performance of the overall task.

One strategy that is widely invoked in lesion research is to attempt to dissociate two mental activities by showing that damage to a given brain part may interfere with one but not another. If the two activities differ in that one requires elementary operations that the other does not, then one might infer that the damaged locus performed the additional activities. Single dissociations, however, do not show that the damaged brain part is only involved in the impaired activity, since it could be that the two activities differ in the demands they make on a component and that increased damage to the same brain part might interfere with both activities. As a result, researchers often seek double-dissociations, where damage to one area causes disruption in one activity (while leaving the other largely unaffected), and damage to another area disrupts the other activity (while leaving the first largely unimpaired). Double dissociations are often taken as compelling evidence that the two activities are performed separately in the brain (Shallice, 1988).[6] Thinking comparatively about the two activities, focusing on what operations they might utilize in common and what different operations each requires, can be a productive strategy for arriving at a decomposition of a task into more basic operations, which can then be identified with the brain areas in which lesions affect one or the other activity.

The challenge for researchers is that elementary operations are not simply given. Feinberg and Farah characterize neuropsychology as turning to cognitive psychology to understand the component operations (processes):

> Traditionally, neuropsychologists studied the localization and functional organization of *abilities*, such as speech, reading, memory, object recognition, and so forth. But few would doubt that each of these abilities depends upon an orchestrated set of *component cognitive processes*, and it seems more likely that the underlying cognitive components, rather than the task-defined abilities, are what is implemented in localized neural tissue. The theories of cognitive psychology therefore allowed neuropsychologists to pose questions about the localization and organization of the components of the cognitive architecture, a level of theoretical analysis that was more likely to yield clear and generalizable findings (Feinberg & Farah, 2000, p. 16).

Much of cognitive psychology has involved hypothesizing component operations and testing these by using such measures as the different reaction times for tasks utilizing and not utilizing the hypothetical component. Cognitive psychologists have also been able to draw on insights of researchers in artificial intelligence, who adopt an engineer's perspective in trying to conceive how to build up from simpler processes to performing an overall operation. But the bottom line is that component information processing operations are not directly apparent and must be

[6] Recent investigations with neural networks and other dynamical systems, however, have shown that double-dissociations can result form differential damaging a single system where it is known that there are not different subsystems carrying out separate tasks (e.g., applying rules for pronouncing words versus looking up pronunciations in a lexicon—see (Hinton & Shallice, 1991; Van Orden, Pennington, & Stone, 2001). Thus, double-dissociations are not foolproof indicators that there are separate systems responsible for separately impaired activities.

inferred by considering what needs to be done to perform a cognitive task with the information available. But by combining the resources of cognitive psychology and lesion experiments, researchers in each field acquire additional tools with which to constrain their hypotheses about elementary operations.

Even if one knows what are the basic information processing operations involved in performing a cognitive task, there still remains a challenge in linking a given operation to a brain region based on lesions alone. A lesion may interrupt an operation by removing the part that performed it, or it may interrupt it more indirectly by removing inputs to the responsible region or critical modulatory feedback to the responsible region. Lesion studies themselves provide no way to resolve these worries. One way to address these worries is to try to correlate activity in a region in a normal system with the tasks being performed. This is the strategy of the next two techniques I will consider.

Despite their limitations, lesion studies have been and remain a major source of evidence in cognitive neuroscience. When a lesion to a brain region results in a determinate pattern of deficit in specific mental activities, the results appear to provide compelling evidence that the region figured in generating those activities in normal humans. But the evidence that the region was responsible for a given task remains subject to the charge of artifact unless other techniques are able to provide independent indicators of what the region is doing.

## 3. Single-Cell Recording

The discovery of the nature of electricity and that the brain in part operates on electrical principles (proposed by Luigi Galvani in the 18$^{th}$ century and definitively established by Emil du Bois-Reymond in the mid-19$^{th}$ century) enabled neuroscientists to study the brain as one would study other electrical systems, probing it with electrical stimuli or recording its electrical activity. Of these, the more probative approach for studying cognitive operations has been to record electrical activity as the brain is engaged in its different functions. One can do this in two ways. One involves recording from electrodes placed on the scalp. To understand mental processing, researchers typically attempt to correlate the resulting electroencephalogram with a specific stimulus given to the subject. By averaging over numerous trials to average out background information, researchers have developed a measure (the evoked response potential or ERP) which has been very useful in fixing the temporal pattern of neural processing (e.g., determining when attentional processes affect information processing, see Luck & Ford, 1998). Although useful in providing very fine-grained information about the timing of neural processes, it is extremely difficult to determine the spatial origin in the brain of the ERP signal recorded on the skull. This makes it very difficult to use ERPs to localize component operations in the brain.

The alternative approach to recording electrical activity in the brain is to record from individual neurons either by inserting an electrode into the neuron or by placing it next to the neuron. Although the procedures for doing this are now routine, they were challenging to work out. One problem stemmed from the weakness of the electrical signal. The combined effort of numerous investigators was required to develop instruments capable of amplifying the signal so that it could be detected, but by 1925 Edgar Adrian at Cambridge was able to record from a nerve in a

suspended frog leg.[7] His results, however, were highly irregular, initially suggesting an artifact stemming from the recording equipment.[8] Even more puzzling was that when he laid the leg flat, the signal ceased. But that was also a clue—when the muscle was suspended it was stretched, and sensory nerves would be responding to the degree of stretch. He then inferred that the small oscillations he saw represented action potentials, sometimes from single nerve fibres. This pointed to the prospect of recording from a single nerve fibre, and Adrian set that as an objective in the last sentence of his paper from the 1925 experiment: "More detailed analysis of these results is postponed until experiments have been made on preparations containing a known number of sensory endings, if possible only one" (Adrian, 1926, p. 72). That same year Adrian was joined by Yngve Zotterman and they pursued a strategy of slicing off spindles of a muscle connected to a nerve fibre until only one remained, revealing the individual action potentials traveling on that nerve. The pattern of action potentials, all of the same magnitude traveling at the same speed, confirmed what until then had only been determined indirectly—that action potential were not graded but all-or-none.

The pattern of firing was not only regular, fitting the first criterion for being reliable evidence, but fit a plausible theoretical perspective, the third criterion. Moreover, with increased weight was placed on the suspended muscle, the rate of firing increased, suggesting that the nerve encoded information through firing rates, further developing a plausible framework in which to interpret the results. Adrian and Zotterman (1926) also found that the nerve fired most when a new weight was presented, and then diminished, demonstrating that the nerve encoded changes in stimulus, not the absolute value of the stimulus. Subsequently, with Detlev Bronk, Adrian developed procedures for removing all but two or three axons to a muscle in a rabbit and recording from electrodes placed on the fragment (Adrian & Bronk, 1928, 1929). This enabled recording the action potentials sent to control the nerve. Adrian and Bronk also attached loudspeakers to their amplifiers, allowing one to listen to clicks as well as observe them on an oscilloscope.

With these tools, Adrian and numerous other investigators began to map areas of the brain where neural responses could be elicited either as the animal moved or as its body was stimulated. One of the results from Adrian's work was the discovery of a second projection zone for a cat's paw. Multiple projection areas soon became the norm. This helped focus the question—just what is being represented in these different cortical areas? Answering this question required more than just finding a correlation between stimulating an area of the animal's body and a response but a detailed model of information processing.

Perhaps the domain in which single-cell recording has been applied most profitably in working out an information processing decomposition is vision. The first use, however, was largely to confirm results that had already been obtained from lesion results—the existence of a topological map of the visual field in the primary visual cortex (Talbot & Marshall, 1941). But with single-cell recording one could look at the expanse of the visual field to which a single cell would

[7] For an account of Adrian's research, and how it drew on others such as Joseph Erlanger and Herbert Gasser, see (Finger, 2000, chapter 15)

[8] I noted above that definite results were a sign of reliable evidence. Irregular results, accordingly, suggest an artifact.

respond, which Haldan Keffer Hartline (1938) designated as the cell's *receptive field*. Stephen Kuffler (1953) extended the use of single-cell recording to identify more specifically the type of stimulus to which a cell would respond when he showed that retinal ganglion cells and cells in the lateral geniculate nucleus (LGN) responded most strongly when a stimulus was presented at the center of the receptive field and not in the surround (on center, off surround) or vice versa (off center, on surround). The determinateness of this pattern was an important indicator that the results were not artifacts but genuine indicators of the information the cells were carrying.

Over the next half-century single-cell recording has provided a great deal of information about the response properties of individual cells and sketches of how each cell processes information from earlier cells. In one of the papers in neuroscience most cited by philosophers "What the frog's eye tells the frog's brain," Jerome Lettvin and his colleagues identified retinal ganglion cells which responded to specific stimuli, including some that responded to small moving spots that the researchers characterized as bug detectors (Lettvin, Maturana, McCulloch, & Pitts, 1959). In this case the plausibility that these cells were tuned to moving spots and that the result was not an artifact was clearly enhanced by the ecological analysis that frogs needed to be sensitive to bugs in order to catch and consume them. But even more important for developing an information processing account of visual processing was Hubel and Wiesel's (1962; 1965; 1968) use of single-cell recording to map out the response characteristics of cells in primary visual cortex. Their initial assumption was that cortical cells would respond to spots of light as Kuffler had indicated for retinal and LGN cells, but their quest to find such cells was fruitless as no cells produced a clear pattern of results. The fortuitous sticking on one occasion of a slide in the projector, resulting in a bar rather than a spot of light which elicited a strong response from cortical cells, led them to explore a different type of stimulus. They eventually realized that some cells responded to specifically oriented bars of light or darkness at specific locations in their receptive field (*simple cells*) while others responded to specifically oriented bars anywhere in their receptive field (*complex cells*). Importantly, they also proposed a simple information processing model of how simple cells could compute the presence of bars from the firing pattern of multiple center-surround cells in the LGN and how complex cells could recognize the presence of bars at different locations in their receptive fields from multiple simple cells.

Hubel and Wiesel's work exemplifies how a new technique could provide information totally unanticipated. It also was quickly accepted. What made the evidence so compelling? Of major importance was the regularity in the pattern of the data. Not only did they identify cells responsive to bars of particular orientations, but they found a regular pattern to the distribution of cells responsive to different orientations. Second, insofar as single-cell recording confirmed the topological orientation of primary visual cortex, the data was consistent with that generated from lesions studies, although it went far beyond it. Finally, with their information processing ideas, Hubel and Wiesel provided a theoretical context in which their results could be interpreted.

Over ensuing decades researchers have worked out the processing of different kinds of visual information by cells in extrastriate and temporal cortex (for a contemporary account of the steps in visual processing that has been produced primarily from single-cell recording, see van Essen and Gallant (1994); for an historical overview, see Bechtel (2001a). Similar research, beginning with Clinton Wolsey (1960) has been carried out for audition. Cell recording has also been used

to identify cells engaged in tasks further removed from the sensory and motor periphery; for example, Goldman-Rakic (1987) has identified cells that continue to fire after a stimulus has been removed when the animal must retain that information for a short interval before performing action.

As useful a technique as cell recording is, it does confront limitations. First, since the technique is primarily correlational, it requires identifying a sensory stimulus, motor response, or ongoing cognitive activity that can be correlated with the neural activity. This is most easily done close to the sensory and motor periphery. But it also depends on the luck and ingenuity of the experimenter (recall that it was the sticking of a slide in the projector that led to Hubel and Wiesel's breakthrough) in testing various stimuli. Van Essen and Gallant (1994) show that many V4 cells surprising seem to prefer non-Cartesian shapes, which would not be natural candidates to test.[9] Second, although when successful it allows researchers to identify what stimulus drives the cell, it does not reveal what contribution the cell is making to processing that information. As Marr (1982) argued, this requires an analysis of the task the cognitive system is performing and accounts in terms of elementary operations of how it is carrying out that task. Knowing what kind of stimulus different cells in a processing pathway are most responsive to can provide clues as to what operation needs to be performed for the cells later in the pathway to compute their responses from the responses of cells earlier in the pathway. Third, it assumes that electrical responses of individual cells are the proper correlate of psychological function. Increasingly, researchers are exploring the possibility that the proper correlate may be a pattern distributed over many cells. Procedures for recording from many, possibly hundreds, of cells simultaneously are now being developed, but these pose serious challenges in terms of analyzing the resulting information. Finally, and from a cognitive perspective, a very serious worry is that ethical considerations only permit single-cell recording in non-human species. This makes them unhelpful for higher cognitive functions exhibited primarily in humans.

## 4. Neuroimaging

This brings us to the newest research technique in the cognitive neuroscience arsenal, the tool that arguably is responsible for the development of a special field of cognitive neuroscience. It is also a technique that remains contentious. For a recent extremely negative appraisal of neuroimaging based on analysis of how data is gathered as well as more general doubts about whether cognitive abilities decompose into elementary operations, see Uttal (2001). Neuroimaging offers evidence of basically the same type as single-cell recording—activity in the brain that is correlated with operations a subject is performing. However, it has captured the public attention in ways that single-cell recording never did. A major reason for this is that neuroimaging in noninvasive and can be performed on normal humans while they are performing the kind of cognitive tasks that are thought to be most distinctively human, including such activities as making ethical judgments.

A host of neuroimaging techniques ranging from computerized tomography (CT or CAT scans)

[9] They also argue that cells give graded responses to stimuli and argue that researchers should not view them as feature detectors but as filters.

to magnetic resonance imaging (MRI) can provide information about structures in the brain. To image function, one needs to record a signal that is related to the activities being performed by neurons. The two techniques that have been most successful and have attracted the attention both measure blood flow.[10] The connection between neural activity (spiking) and blood flow is relatively intuitive—greater firing rates requires more energy, which requires more metabolism to provide the energy. This in turn requires a source of both oxygen and metabolites, which is provided by blood. But herein lies one of the current unknowns underlying neuroimaging—the mechanism by which greater blood flow is generated is not known. It is known that the increase exceeds oxygen demand, a critical factor in creating the blood oxygen level dependent (BOLD) contrast between conditions of heightened and less heightened neural activity. Although the details of the relationship will ultimately be important for understanding some uses of fMRI and some researchers are very concerned about the fact that the mechanism generating the signal is not understood (see Fitzpatrick & Rothman, 1999 for a report on a McDonnell Foundation sponsored conference devoted to this issue), most researchers are sufficiently confident in the relation between neural activity and measures of blood flow that they are not worried about an artifact arising at this point. This confidence is largely a result of the fact that neuroimaging has produced many determinate results that have been incorporated into information processing models of cognition.

In the case of both PET and fMRI the process of detecting the signal relies on application of principles from physics. In most PET studies, for example, radioactively labeled $H_2O$ is injected into the bloodstream. With a short half-life, it decays as it is carried by the blood, ejecting positrons as it does so. The positron will travel only a short distance until it collides with an electron, whereupon it is annihilated and emits two gamma rays directed $180^o$ opposite each other. The PET scanner contains detectors surrounding the head which record an event only when two gamma rays arrive at different locations simultaneously; a sophisticated computational approach known as tomography (from the Greek word *tomos*, which means *cut*) are then employed to determine the site of the annihilation in the cut defined by the plane of the scanner.[11] These basic processes are reasonably well understood and are not the focus of concern in the application of PET and fMRI to measuring neural activity.

Most of the epistemic issues concerning PET and fMRI turn on the connection to cognition. The challenge is posed by the fact that during the performance of a cognitive task there is blood flow throughout the brain. One might think that one could just focus on increased blood flow in the performance of a task, but that requires being able to specify the base-line from which the increase is measured. But a person is never cognitively inactive while awake. Leave a subject in a scanner with no directions to do anything specific and the subject will think about whatever she chooses to think. Beyond the problem posed by identifying a baseline, a more serious objection

[10] There are techniques for using PET to directly measure metabolism by using radioactively labeled 18-fluoro-2-deoxyglucose, but these are used primarily in diagnostic studies not functional studies.

[11] MRI uses a strong magnetic field in which the nuclei of elements which have an odd atomic weight (e.g., hydrogen) are induced to align the axes of their spin. A brief pulse of radiowaves can perturb this alignment by tipping the orientation of spin, thereby increasing the energetic state of the nuclei. When the pulse ends, they precess back into their aligned state, releasing energy in the form of radio waves in which the frequency reflects the particular atom and its environment.

is that such an approach, if successful, would only show us areas that are involved in the performance of a task, but would not show what operation these areas perform.

One strategy for linking brain activity as measured in neuroimaging with cognition that has been widely employed is the subtractive method. This involves imaging a person while performing two different tasks thought to differ only in that one employs one (or a very small number of) cognitive operations additional to those employed by the other, and then subtracting the second image from the first. This produces a difference image, the sort of image most commonly displayed in neuroimaging papers. One then identifies the area(s) revealed in the difference image as the locus of the additional operations. For example, in a landmark early neuroimaging study, Petersen et al. (1988, 1989) subtracted the image produced when a subject read a noun and pronounced it aloud from the image produced when a subject read nouns, generated a related verb, and pronounced the verb. The researchers then identified the areas of increased blood flow in left dorsolateral prefrontal cortex as the brain areas required to perform the semantic operation of generating the verb.

The procedure is an adaptation of one initially developed by F. C. Donders (1868) for use in chronometric studies of cognitive process. In these studies, the time take to perform one task was subtracted from that required for another task and the difference was thought to reflect the time required for the additional processes required in the longer task. In its application to chronometric studies, the subtractive method was broadly criticized in the 1960s. Sternberg (1969) pointed out, for example, that it assumed that the additional cognitive activity was a pure insertion into a sequential set of processes, and this assumption might well be false. As a result, he advocated replacing the subtractive method in studies of mental chronometry by techniques which measure whether different tasks interfere with each other (for detailed discussion, see Posner, 1978).

Neuroimagers have returned to the original simple subtraction approach of Donders. Raichle (1998) defends doing so by arguing that imaging itself will reveal any changes in activation in other brain areas that would indicate interactions with other processes and not a pure insertion. There are, however, reasons for skepticism: imaging procedures will only identify statistically significant changes in activation elsewhere in the brain; if there are resulting accommodations elsewhere in the brain, they may fall below this threshold and thus not be noted. Moreover, given the computational demands of calculating responses throughout the brain, researchers often only look for significant responses in areas they already suspect of being involved in the operation in question.

So as to better appreciate the epistemic issues arising in neuroimaging research, I will turn to a specific case in which new localizationist claims were advanced on the basis of neuroimaging. Cognitive psychologists have decomposed memory processes temporally into processes of encoding, storage, and retrieval. Many researchers assumed that storage is widely distributed in cortex, but that encoding and retrieval are more focused processes whose neural mechanisms might be uncovered. Moreover, these processes could be differentiated in terms of how factors thought to affect either encoding or retrieval specifically influence overall memory performance. But attempts to relate encoding and retrieval processes to the brain through lesions faced the

limitation that any measure of memory required both encoding and retrieval, and so it was not possible to determine whether the lesion impaired encoding or retrieval processes. Neuroimaging offered the opportunity to determine what brain areas were selectively involved during encoding and retrieval.

Accordingly shortly after the introduction of PET to study cognitive activity, Endel Tulving and a number of his associates initiated a study of encoding and retrieval processes associated with episodic memory—the memory a person has of being directly involved in an event in the past (as distinct, say, from their knowledge that they were involved in the event, which they could have acquired from the reports of others). To produce varying levels of encoding, Tulving and his collaborators varied the *level of processing* of words, a manipulation that was known from purely behavioral studies to affect how well the stimuli were encoded (Craik & Tulving, 1975). Subjects either determined whether the word contained an *a*, a shallow encoding condition, or whether it referred to a living thing, a deeper encoding condition. A subsequent recognition memory task revealed that subjects recognized more of the words as being on the list when the encoding task required them to determine whether the referent was living. When the PET images made under shallow encoding conditions were subtracted from those made under deeper encoding, significantly increased activation was found in a region in left inferior prefrontal cortex extending from Brodmann's areas 45 and 46 to areas 47 and 10 (Kapur et al., 1994). To study episodic retrieval Tulving et al. (1994) presented subjects either with novel meaningful sentences or meaningful sentences they had heard 24 hours previously. Although the subjects were not required to register their recognition of the previously heard sentences, when the blood flow for novel sentences was subtracted from that for previously heard sentences, increased blood flow was found in right dorsolateral prefrontal cortex (from Brodmann area 10 through 46, 9, and 6) and bilaterally in two regions in parietal cortex. There were also regions of reduced blood flow in the temporal cortex.

The main finding of these studies was that encoding of episodic memory resulted in increased blood flow in the *left* prefrontal cortex whereas retrieval produced increased blood flow in the *right* prefrontal cortex. This led Tulving and his collaborators to propose the hemispheric encoding/retrieval asymmetry (HERA) model which asserts that:

> the left and right prefrontal cortical regions are differentially involved in episodic and semantic memory processes. Left prefrontal cortical regions are involved in retrieval of information from semantic memory to an extent that right prefrontal areas are not, at least insofar as verbal information is concerned. Left prefrontal cortical regions are involved in encoding information about novel happenings into episodic memory to an extent that right prefrontal areas are not, at least insofar as verbal information is concerned. Right prefrontal cortical regions are involved in retrieval of episodic information to an extent that left prefrontal areas are not. Right prefrontal cortical regions are involved in retrieval of episodic information to an extent that does not hold for retrieval of semantic information (Tulving, Kapur, Craik, Moscovitch, & Houle, 1994, p. 2018)

There are some important qualifications in this characterization of the HERA model to which I will return in a moment. But first it is important to note the role of subtraction in the two studies on which HERA is based. Only with subtraction did the specific areas in left and right prefrontal

cortex stand out. Buckner (1996) noted that prior to subtraction in the episodic recall task, activation was found in left prefrontal areas as well as right prefrontal areas. These areas were subtracted out since they were also elicited by presentation of novel sentences. Presumably these are areas that are involved in semantic processing of the sentences. But they may also figure in the network of areas involved in episodic retrieval. As I noted, Tulving qualified his presentation of HERA in a number of ways, one of which was to claim only that the right prefrontal areas are more involved in episodic retrieval than the areas on the left. This might be seen as a concession that the subtracted areas might contribute something, but in fact it is misleading. There is no basis for quantifying the contributions—the regions in the left hemisphere may be carrying out processes just as important to recognition as those in the right. A similar concern arises about the subtraction in the episodic encoding study. In that study, what was subtracted was the activation involved in shallow processing. But these processes may also contribute to the deep processing that is thought to be involved in encoding of episodic memory.

Tulving and his collaborators are quite cognizant that the areas they have identified in their neuroimaging studies may only perform a component operation in the encoding or retrieval process and that the whole process may involve a complex network of other areas each performing a different operation underlying encoding or retrieval. In their study of episodic encoding they specifically comment:

> Although left prefrontal activation was associated with enhanced memory performance in our study, it seems quite unlikely that this region constitutes the locus of memory storage. It is more plausible that the left inferior prefrontal structures identified are part of a more complex network of cortical and subcortical structures that subserve memory functions (Kapur et al., 1994, p. 2010).

My contention so far is that some of the areas relevant to encoding or retrieval may be hidden by reliance on subtraction.

A second concern that arises from the subtraction design is characterizing the additional task that differentiates the focal task and the subtracted task. Tulving and his collaborators explicitly note that the areas they identify as involved in episodic encoding overlap substantially with those Petersen et al. identified as involved in semantic processing. Tulving proposes that the two tasks are in fact linked—semantic processing is one form of deeper processing that enhances encoding. But then the question naturally arises as to whether the areas activated in Tulving's studies play any role in encoding *per se*. Adina Roskies raised this issue in a commentary that appeared in the same issue as the Tulving studies:

> While it is possible that this brain area is directly involved in memory processes, it is also possible and consistent with other literature that this area is specifically involved in semantic processing or response generation, both of which occur in this task, and that the enhancement of recognition that is reported could be due to synaptic changes in other brain regions, even those lacking direct connections with prefrontal cortex (Roskies, 1994, p. 1990).

Another qualification in Tulving et al.'s presentation of HERA is that the results may only hold for verbal information. Employing a refinement that allows researchers to analyze separately encoding episodes for which there is later recall from those for which recall fails (thereby

avoiding reliance on depth of processing as a surrogate for successful encoding), Gabrieli and his colleagues demonstrated right inferior frontal cortex activation, as well as bilateral hippocampal activations, when pictures were substituted for words in an encoding studies (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998). Likewise, Kelley et al. (1998) showed right prefrontal activity in encoding of unfamiliar faces and bilateral activations with line-drawings of nameable objects. Together these suggest that the left hemisphere activations in Kapur et al. may be a result of the use of verbal materials, and may have to do more with their semantic processing than episodic memory encoding.

Although I have focused on a number of criticisms that have been advanced with respect to HERA, it remains a focal hypothesis in the literature, one based solely on neuroimaging results. Its critics, relying on different interpretations of these and other neuroimaging results, contend that it is an artifact. To appreciate how this controversy has developed, let's return to the three criteria I have maintained that scientists employ in evaluating artifacts. First, there is little doubt that Tulving and his collaborators produced quite determinate results, ones that could be replicated by others. In the domain of inquiry, there were few other approaches to the problem of identifying neural mechanisms. In particular, there was little evidence that lesions in the prefrontal areas these researchers identified produced deficits in either encoding or retrieval of episodic memories. So that left the results vulnerable. But perhaps the greatest vulnerability stemmed from the fact that the results did little to advance a detailed mechanism of encoding and retrieval. Such advance would require some account of the processes involved in encoding or retrieval, and to date there is little in the literature to suggest what these might be. But this need not be the basis of a negative assessment of the enterprise. Gabrieli, for example, reflecting on the growing number of sites that seem to be activated in encoding and retrieval studies, in addition to those advanced by Tulving, comments:

> "it has been difficult for psychologists to define multiple, specific processes that mediate episodic encoding or retrieval. . . . Yet it is virtually certain that there are multiple encoding and retrieval processes which vary according to materials and task demands. From this perspective, the variability in imaging findings suggests that future imaging studies may provide an impetus not only for more precise process-structure mappings but also for a new level of rigor and precision in understanding the psychological organization of episodic memory" (Gabrieli, 2001, p. 281).

It is too soon to tell whether the evidence of episodic encoding in the left prefrontal cortex and episodic retrieval in the right prefrontal cortex is an artifact or reflects actual neural mechanisms. The pattern of results Tulving and his collaborators produced is clear and replicable, but it is neither supported by other techniques nor yet a component of a detailed information processing theory. It thus gives us a perspective on what factors render scientific evidence vulnerable to the charge of artifact. Other results obtained with neuroimaging, however, do meet these requirements. For example, a PET study by Haxby et al. (1991) indicating separate visual pathways for processing information about object identity and spatial location were consistent with earlier results (Mishkin, Ungerleider, & Macko, 1983) using lesions in monkeys. Moreover, the separation of different processing pathways also fit into a developing theoretical framework for analyzing perceptual processing. On the other hand, the scientific advance represented by Haxby et al. study was primarily to show that the pattern identified in monkeys

held in humans as well and so was less radical than that offered by Tulving and his collaborators. When a new technique is used to develop results beyond what other techniques have made available, the risk of artifact is commensurately greater.

## 4. Conclusions

Evidence in science is frequently the focus of contest. This is true in cognitive neuroscience as in other sciences. I have focused on three of the principal sources of evidence linking cognitive processes to the brain—lesion studies, single-cell recording, and neuroimaging. In each case the results advanced involve manipulation of the phenomena under study and the results must be evaluated as to whether they reflect in the appropriate manner the phenomena being studied or only the mode of intervention used to produce the evidence. I have tried to show how three different features of the evidence itself are appealed to in determining whether results are artifacts. First is whether there is a definite pattern in the results that can be procured reliably. Second is whether the results are consistent with results produced by other techniques. Third is whether the results fit into a coherent theoretical account. Of these, the first and third are especially important, while the second often figures in calibrating a technique or in generating complementary information that helps establish a theoretical framework for interpreting the results.

References

Adrian, E. D. (1926). The impulses produced by sensory nerve endings. Part I. *Journal of Physiology (London), 61*, 49-72.

Adrian, E. D., & Bronk, D. W. (1928). The discharge of impulses in motor nerve fibres. Part I. Impulses in single fibres of the phrenic nerve. *Journal of Physiology, 66*, 81-101.

Adrian, E. D., & Bronk, D. W. (1929). The discharge of impulses in motor nerve fibres. Part II. The frequency of discharge in reflex and voluntary contractions. *Journal of Physiology, 66*, 119-151.

Adrian, E. D., & Zotterman, Y. (1926). The impulses produced by sensory nerve endings. Part 2: The response of a single end-organ. *Journal of Physiology, 61*, 151-171.

Bates, E. (1994). Modularity, domain specificity and the development of language. *Discussions in Neuroscience, 10*(1 and 2), 136-153.

Bechtel, W. (1995). Deciding on the data: Epistemological problems surrounding instruments and research techniques in cell biology, *PSA 1994* (Vol. 2, pp. 167-178).

Bechtel, W. (2000). From imaging to believing: Epistemic issues in generating biological data. In R. Creath & J. Maienschein (Eds.), *Biology and epistemology* (pp. 138-163). Cambridge, England: Cambridge University Press.

Bechtel, W. (2001a). Decomposing and localizing vision: An exemplar for cognitive neuroscience. In W. Bechtel & P. Mandik & J. Mundale & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 225-249). Oxford: Basil Blackwell.

Bechtel, W. (2001b). Linking cognition and brain: The cognitive neuroscience of language. In W. Bechtel & P. Mandik & J. Mundale & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 152-171). Oxford: Basil Blackwell.

Bechtel, W. (2002). Aligning multiple research techniques in cognitive neuroscience: Why is it important? *Philosophy of Science, 69*, S48-S58.

Bechtel, W. (under development). Explanation without Laws: The Mechanist Alternative.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as scientific research strategies*. Princeton, NJ: Princeton University Press.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science, 281*, 1185-1187.

Broca, P. (1861). Remarque sur le Siege de la Faculte Suivies d' une Observation d' Aphemie. *Bulletins de la Societe Anatomique de Paris, 6*, 343-357.

Buckner, R. L. (1996). Beyond HERA: Contributions of specific prefrontal brain areas to long-term memory retrieval. *Psychonomic Bulletin and Review, 3*(2), 149-158.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294.

Donders, F. C. (1868). Over de snelheid van psychische processen. Onderzoekingen gedaan in het Pysiolish Laboratorium der Utrechtsche Hoogeschool: 1868-1869. *Tweede Reeks, 2*(92-120).

Finger, S. (2000). *Minds behind the brain: A history of the pioneers and their discoveries*. Oxford: Oxford University Press.

Fitzpatrick, S. M., & Rothman, D. L. (1999). New approaches to functional neuroenergetics. *Journal of Cognitive Neuroscience, 11*, 467-471.

Gabrieli, J. D. E. (2001). Functional neuroimaging of episodic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 49-72). Cambridge, MA: MIT Press.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In J. M. Brookhart & V. B. Mountcastle & S. R. Geiger (Eds.), *Handbook of Physiology: The Nervous System* (Vol. 5, pp. 373-417). Bethesda, Maryland: American Physiological Society.

Gregory, R. L. (1968). *Models and the localization of functions in the central nervous system.* . Paper presented at the In C. R. Evans and A. D. J. Robertson, eds., Key papers in cybernectics. London: Butterworth. Pp. 91-102.

Hanson, N. R. (1958). *Patterns of discovery*. Cambridge: Cambridge University Press.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate retina. *American Journal of Physiology, 113*, 59-60.

Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch, P., Schapiro, M. B., & Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences, USA, 88*, 1621-1665.

Hinton, G. E., & Shallice, T. (1991). Lesioning a connectionist network: Investigations of acquired dyslexia. *Psychological Review, 98*, 74-95.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London), 160*, 106-154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology, 195*, 229-289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London), 195*, 215-243.

Kapur, S., Craik, F. I. M., Tulving, E., Wilson, A. A., Houle, S., & Brown, G. M. (1994). Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. *Proceedings of the National Academy of Sciences (USA), 91*, 2008-2111.

Kelley, W. L., Miezin, F. M., McDermott, K., Buckner, R. L., Raichle, M. E., Cohen, N. J., & Petersen, S. E. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobes for verbal and nonverbal memory encoding. *Neuron, 20*, 927-936.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology, 16*, 37-68.

Kuhn, T. S. (1962/1970). *The Structure of Scientific Revolutions* (Second ed.). Chicago: University of Chicago Press.

Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers, 47*, 1940-1951.

Luck, S. J., & Ford, M. A. (1998). On the role of selective attention in visual perception. *Proceedings of the National Academy of Sciences (USA), 95*, 825-830.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1-25.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*(414-417).

Mundale, J. (1998). Brain mapping. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science*. Oxford: Basil Blackwell.

Petersen, S. E., & Fiez, J. A. (1993). The processing of single words studied with positron emission tomography. *Annual Review of Neuroscience, 16*, 509-530.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Raichle, M. E. (1998). Behind the scenes of functional brain imaging: A historical and physiological perspective. *Proceedings of the National Academy of Sciences, 95*, 765-772.

Roskies, A. L. (1994). Mapping memory with positron emission tomography. *Proceedings of the National Academy of Sciences (USA), 91*, 1989-1991.

Scoville, W. B. (1968). Amnesia after bilateral medial temporal-lobe excision: Introduction to case H.M. *Neuropsychologia, 6*, 211-213.

Shallice, T. (1988). *From neuropsychology to mental structure*. New York: Cambridge University Press.

Talbot, S. A., & Marshall, W. H. (1941). Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Opthalmology, 24*, 1255-1263.

Tulving, E., Kapur, S., Craik, F. I., M., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Sciences (USA), 91*, 2016-2020.

Tulving, E., Kapur, S., Markowitsch, H. J., Craik, F. I. M., Habib, R., & Houle, S. (1994). Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. *Proceedings of the National Academy of Sciences (USA), 91*, 2012-2015.

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT Press.

van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13*, 1-10.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? Inductive methods and isolable systems. *Cognitive Science, 25*, 111-172.

Walsh, V., & Cowey, A. (1998). Magnetic stimulation studies of visual cognition. *Trends in Cognitive Sciences, 2*(3), 103-110.

Wolsey, C. (1960). Organization of the cortical auditory system: a review and a synthesis. In G. L. Rasmussen & W. F. Windle (Eds.), *Neural mechanisms of the auditory and vestibular systems* (pp. 165-180). Springfield, IL: Charles C. Thomas.