To appear in In R. Skipper Jr., C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson, and R. Richardson (eds.), *Philosophy and the Life Sciences: A Reader*. Cambridge, MA: MIT Press.

The Epistemology of Evidence in Cognitive Neuroscience[1]

William Bechtel
Department of Philosophy and Science Studies
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 link to 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 was once 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).[2] 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 an 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] I thank Cory Wright and Carl Craver for their very helpful comments on earlier drafts of this paper.

[2] 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) 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 investigators have a clear understanding of how interventions serve procure the evidence, 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 introduced. At this stage, charges of artifact and reasons for suspecting artifacts appear frequently in the scientific literature, as do defenses against these charges.

The worries about artifacts point to a new problem area for 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. Researchers regularly introduce and rely on new instruments and techniques well before there is a theoretical understanding of how they work.

To what do researchers look to evaluate whether their evidence is reliable if not an understanding of how the evidence is generated by the instruments and procedures? 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 that 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, 2002a). 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. A mechanism, as I used the term, is a system whose behavior produces a phenomenon in virtue of organized component parts performing coordinated component operations. 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 component parts 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 parts to produce the cognitive phenomenon in question (Bechtel & Richardson, 1993; Bechtel & Abrahamsen, under review; Bechtel, in press; Glennan, 1996, 2002; 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 when they distinguish *elementary operations* from *tasks* and argue that operations, but not tasks, are 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 cognitive phenomena in one component of a system (localizing an area for articulate speech, as discussed below). Such efforts, while not generating an understanding of how the component parts perform operations that together realize the activity of the mechanism, often play an important heuristic role. When, for example, subsequent research points to additional areas involved in performing the activity, then researchers begin to consider proposals for decomposing the task into simpler operations that then are coordinated into the performance of the overall task (Bechtel & Richardson, 1993).

Crucial to understanding a mechanism is figuring out what operations its components are performing. This typically requires some form of intervention and a means of detecting the effects of the intervention. The techniques discussed below differ in where the intervention occurs and where the effects are detected (Craver, 2002). Lesion studies intervene within the system, disabling a component of the system, and typically detect the effects in terms of changes in the overall activity of the system.[3] Both single-cell recording and neuroimaging, on the other hand, intervene on the stimulus presented to the system and detect the effect on components within the system (see Figure 1). In addition to figuring out what operations the components are performing, understanding a mechanism also requires determining how the components are organized. Although the experimental techniques discussed here can give suggestions as to the organization of the components, modeling, including mathematical and computational modeling, are often the best tools for evaluating hypotheses about organization.

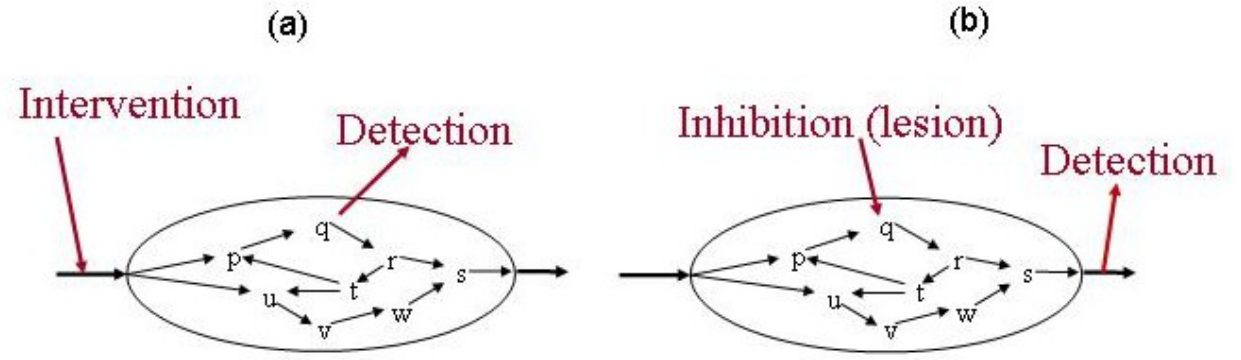

Figure 1. The locus of intervention and detection in (a) single-cell recording and neuroimaging studies and (b) lesion studies.

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 (Schaffner, 2001, reprinted in this volume). Sometimes the decision to work with a given species is driven by ethical considerations (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 an organism or the ability to secure reasonably clean data drive the decision. But the fact that different species are employed means that correspondences must be established 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 correspondences can only be established indirectly in terms of the common pattern of results. For example, one way to

[3] An approach not further discussed in this paper that has on occasion been influential in determining how the brain works is to stimulate a component and detect its effect on the overall behavior of the system. An elegant example of the use of stimulation is found in the work of William Newsome (Newsome, Britten, & Movshon, 1989). In a context of an ambiguous stimulus, microstimulation of cells in MT influenced the monkey's judgment about the direction of movement in the stimulus pattern. In this approach, as in lesion research, the intervention is within the system and detection occurs at the level of the behavior of the whole system.

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 employed in cognitive neuroscience.

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 interventions by neuroscientists that actually destroy neural tissue.[4] (The use of naturally occurring lesions is closely associated with the subfield of neuropsychology.) 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 deficit what contribution the damaged area made to normal psychological activity. This is an extremely difficult inferential task, as we shall see (see also Glymour, 1994, reprinted in this volume).

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 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 became the name by which he is often known. Subsequently, LeBorgne lost sensitivity on 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, 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 were dismissed as artifacts.[5] 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

[4] Such deliberately destructive studies are not permitted on human subjects except when intended as therapy. For example, physicians may excise brain tissue to remove tumors or to stop epileptic seizures. Typically, they will only excise tissue not thought to be playing 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.

[5] Despite his bad reputation, Gall was a very skilled neuroanatomist. Moreover, his underlying idea of correlating an activity with a feature of a brain that is engaged in performing that activity is fundamentally the same as the techniques now used in cognitive neuroscience. For further discussion, see Zawidzki and Bechtel (2004).

(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 (his case was undoubtedly hurt by his reputation as an enthusiast of Gall). In addition to a general prejudice against localization in the wake of Gall, there was a conviction, stemming from Descartes, that the mind was a unity so that even if the mind 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, prior to Broca's work, claims to the localization of language ability were generally dismissed.

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. Broca maintained that when LeBorgne's deficits were limited to the inability to speak, the damage was limited to this central area. Broca went on to contend 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 hold that control of speech was typically localized in the left frontal cortex.[6] These additional patients, together with those identified by Bouillaud, produced a definite pattern of results, suggesting that they were not just artifacts. However, the pattern was not perfect 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. A decade after Broca's work, Carl Wernicke (1874) 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*. He proposed various networks that linked his, Broca's, and other areas so as to perform various linguistic tasks.

The first half of the 20th century was dominated by skepticism about localization claims and investigators focused on the integration of the whole mind (a view ensconced in Karl Lashley's (Lashley, 1950) designation of most areas of cortex as *association areas*. The tradition of Broca and Wernicke, however, was revived in the 1950s by Norman Geschwind (1979). He advanced a model of reading aloud in which the visual areas at the back of the brain first process incoming information and pass the results to Wernicke's area where comprehension occurs. Information

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

encoding the comprehended text is then sent to Broca's area, where speech is planned. Finally, information is sent to the adjacent motor areas that direct the actual production of 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 studies have provided a principal avenue for understanding brain operations in both neuropsychology and behavioral neuroscience, especially in the period from 1950 to 1970. They were nearly the only kind of study of neural process one could carry out onhumans. The approach extended far beyond deficits of language (aphasias, apraxias), providing important clues to other cognitive phenomena such as perception and memory. For example, the inability of patients who could otherwise see to recognize an object or identify its functional significance (a condition known as *visual agnosia*) revealed the importance of areas in the temporal lobe in recognizing and categorizing what a person is seeing (see Farah, 1990, for a detailed review). Likewise, the variety of memory deficits found in amnesic patients led to the differentiation of different types of memory processing (see Squire, 1987, for a detailed review). These investigations of deficits in human patients were often coupled with lesion studies in other species, where the lesions can be experimentally induced. Accordingly, lesions in the temporal lobe were found to induce visual agnosia in monkeys (Mishkin & Pribram, 1954; Mishkin, 1966) while hippocampal lesions were found to produce deficits in spatial memory in rats (O'Keefe & Nadel, 1978).

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, researchers 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. Such was the case with Leborgne, requiring Broca (1861), for example, to engage in protracted argument as to where in that area 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). (An alternative approach to generating lesions, using chemical toxins or inhibitors, overcomes this problem. For example, tetrodotoxin (TTX) and lidocaine specifically target sodium channels and the effects are reversible, allowing researchers much more precise control over the lesions produced, but still faces the challenges that follow. The choice between various techniques designed to achieve similar results is itself a rich topic for the philosophy of experimentation.)

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 of semantic memories, to the conclusion that it is necessary for

encoding new semantic memories). But a component operation may be important without being necessary. For example, the brain, like most biological systems, may exhibit redundancy, so that even when a component fails to perform its operation, another component performs the same or sufficiently similar operation that no overall deficit results. Or the brain may reorganize so that another component alters its operation to make up for the deficit. The brain is a constantly adapting structure in which, once an area is removed, processing in other areas is also altered. Redundancy and reorganization undercut interpreting lesion experiments as demonstrating the necessity of a component or its operation to normal performance.[7]

Typically, however, the goal is not just to learn that a region was necessary for a function, but to ascertain what it contributed to the performance of the function—that is, what elementary operation it performed. This is extremely challenging. You can gain an appreciation of the challenge involved by considering how you 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 you must shift focus from the deficits manifested in the damaged system to the activity performed by the normal system and the component operations that figure in that activity. In the case of products like radios, engineers designed the mechanisms from knowledge of what operations components could perform. With natural systems, however, 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 an elementary operation that the other does not, then it seems plausible to infer that the damaged locus performed that elementary operation. 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).

In fact, however, double dissociations can arise even if the two activities share a number of component operations. The discovery of a double dissociation encourages researchers to think

[7] 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 transcranial magnetic stimulation (TMS); 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.

comparatively about the two activities, asking what operations they might utilize in common and what different operations each requires. Such a strategy is a productive way to generate a decomposition of a task into more basic operations. Lesions disabling one of these operations and therefore the task that depends upon it can then explain the double dissociation.[8]

A fundamental challenge for researchers is that elementary operations do not simply present themselves to researchers but must be inferred. In a passage that harkens back to Petersen and Fiez's distinction between tasks and elementary operations, Feinberg and Farah characterize neuropsychology as turning to cognitive psychology to move beyond localizing tasks to focusing on the component operations (processes) underlying these tasks:

> 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).

At the core of cognitive psychology is the conception of the mind an information processing system—a machine that through a series of operations transforms information. It thus could offer neuropsychology insight into component operations.

The benefits of interaction between neuropsychology and cognitive psychology went in both directions. Cognitive psychologists had to rely on hypothesizing component information processing operations that could account for performance (or sometimes 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). Their ability to test these hypotheses was typically limited to such measures as error patterns or differences in reaction times between tasks thought to differ in single component operations. Lesion results provide a different means to test hypotheses about component operations, one that relies on different (albeit it potentially problematic) assumptions. If results of independent techniques do concur, that is one way to counter worries that either represents merely an artifact.

---

[8] Recent investigations with neural networks and other dynamical systems have shown that double dissociations can result form differentially damaging single network systems which do not employ different subsystems for performing different tasks (e.g., applying rules for pronouncing words versus looking up pronunciations in a lexicon—see (Hinton & Shallice, 1991; van Orden, Pennington, & Stone, 2001). Although such results seem to count against the claim that double-dissociations are evidence for separate systems being responsible for separately impaired activities, they are compatible with the strategy just outlined of focusing on what elementary operations figure differentially in the two tasks. The double-dissociations in these networks are generated by lesioning different parts of these networks, and it is plausible to construe these parts of the networks as responsible for different elementary operations. Moreover, even if sometimes double dissociations are not due to separate components, that does not negate the usefulness of looking for double dissociations. All research techniques are fallible, and the most one can demand is that they are probative

Even if one knows what the basic information processing operations involved in performing a cognitive task are, it remains a challenge to link a given operation to a brain region relying only on lesions. I already noted above the problems created by redundancy and reorganization. Even when these are overcome, the linkage between lesion and normal operations can be complex. A lesion may interrupt an operation by removing the part that performed it, but it may also interrupt it more indirectly by removing inputs to the responsible region or critical modulatory feedback to the responsible region. It may also create a deficit in an even more indirect manner such as by occluding blood flow to other regions or creating swelling that affects operations performed in other regions. Lesion studies themselves provide no way to resolve these worries and determine what operation the lesioned component performed in the normal organism. 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.

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 18th century and definitively established by Emil du Bois-Reymond in the mid-19th century) enabled neuroscientists to study the brain in the same way investigators study other electrical systems. These include applying electrical stimuli to intervene in its operation 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. Two different ways of recording electrical activity have played major roles in cognitive neuroscience. One involves recording from electrodes placed on the scalp that pick up aggregate electrical currents.[9] These currents originate primarily from pyramidal cells that are aligned in columns in the cortex; stimulation of these cells produces ion flows into and out of the cell. When the cells are aligned spatially and activated synchronously, these ion flows create an electric field strong enough to be recorded at the scalp (Kutas & Dale, 1997). Recordings of these fields constitute an electroencephalogram (EEG). EEGs have been employed for such purposes as studying sleep-awake cycles and detecting the origins of epileptic tumors. To link the EEG to mental processing requires further analysis. One step is to relate the EEG response in time to the presentation of a given stimulus. Another is to sum the response over numerous trials to average out background information. This generates another 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). By comparing the temporal responses to different stimuli or in different task situations, researchers gain important clues as to the mental operations that are differentially invoked. A limitation of ERP research, however, is that it is extremely difficult to determine the spatial origin in the brain of the ERP signal recorded on the skull. For this reason,

[9] Hans Berger, a German psychiatrist, developed the technique for recording such signals by modifying procedures designed to record the much stronger electrical signal from heart muscle. The signals reflect an aggregate of neural activity. Berger distinguished several different patterns of waves, including what he called *alpha waves* (large-amplitude, low-frequency waves which appeared when participants closed their eyes) and *beta waves* (smaller, higher-frequency waves, which appeared as soon as the subject received sensory stimuli or was asked to solve a problem (Berger, 1929)

cognitive neuroscientists are increasingly using ERP studies together with functional neuroimaging techniques discussed in section 4.

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 sufficiently to detect it. By 1925, though, Edgar Adrian at Cambridge was able to record from a nerve in a suspended frog leg.[10] His results, however, were highly irregular, initially suggesting an artifact stemming from the recording equipment.[11] Even more puzzling was that when he laid the leg flat, the signal ceased. That anomaly was also a clue—when the muscle was suspended it was stretched, and he proposed that sensory nerves responded to the degree of stretch. He further inferred that the small oscillations he saw represented action potentials, possibly arising from single nerve fibers. This pointed to the prospect of recording from a single nerve fiber, 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). With Yngve Zotterman he pursued a strategy of slicing off spindles of a muscle connected to a nerve fiber 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 potentials were not graded but all-or-none (Adrian & Zotterman, 1926).

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 placed on the suspended muscle, the rate of firing increased, suggesting that the nerve encoded information through firing rates. This further articulated 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 them 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. Lesion studies already provided information about the primary cortical projection areas for different areas of the animal's body, but Adrian's work initially generated a surprising result: he

[10] 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)

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

found a second projection zone for a cat's paw. With these tools for recording from single cells, other researchers found additional instances of multiple projections from the senses to cortex and multiple projection areas soon became the norm. Such findings, however, generated a new question—just what is being represented in these different cortical areas? Answering this question required more than just correlating stimulation of an area of the animal's body with a response; it required a detailed model of information processing.

Single cell recording has been used most effectively in decomposing the information processing tasks performed by the visual system. As with Adrian's work, the first application 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). Single-cell recording, however, enabled researchers to do something that was not possible with lesion studies—determine the expanse of the visual field to which a single cell would respond. Haldan Keffer Hartline (1938) introduced the phrase *receptive field* for this area. One could also determine more specifically the type of stimulus to which a given cell would respond. Stephen Kuffler (1953) pioneered this inquiry by recording from retinal ganglion cells and cells in the lateral geniculate nucleus (LGN) of the thalamus (a brain region that acts as a gateway for most incoming sensory information). He found that these 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. It was easy to propose advantages to having cells respond to a difference between center and surround (e.g., it would enable an organism to detect the boundaries of objects), thereby embedding the result in a plausible theoretical perspective.

Over the half-century since Kuffler's pioneering research, single-cell recording has provided a great deal of information about the response properties of individual cells and supported inferences as to how cells in a pathway process 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). The ecological analysis that frogs needed to be sensitive to bugs in order to catch and consume them enhanced the plausibility of the claim that these cells were tuned to moving spots.

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. After long, frustrating investigations, an accident provided a clue. Hubel was changing the slide in the project when it stuck, casting a bar rather than a spot of light. This elicited a strong response from the cell from which they were recording, which had previously been quiescent. (Hubel, 1982) After this fortuitous finding, they began to test bars of light and eventually determined that some cells in primary visual cortex 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 (see Figure 2).

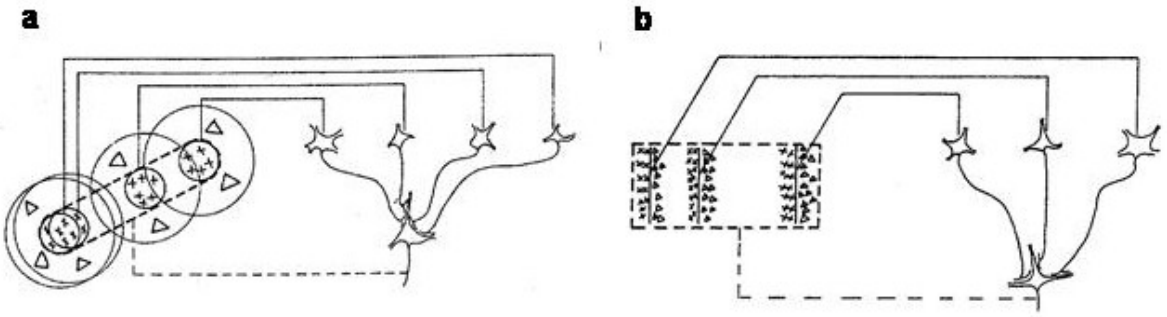

Figure 2. Hubel and Wiesel's information processing proposals. In (a) simple cells identify a bar from the output of center-surround cells in the LGN. The cell on the right serves as an and-gate, firing only if the four cells shown on the left are all activated by the stimulus. In (b) complex cells identify bars anywhere in their receptive fields. The cell on the left is an or-gate, firing if any of the simple cells on the left are activated by simple cells on the left that respond to a vertical bar).

Hubel and Wiesel's work exemplifies how a new technique could provide totally unanticipated information. Their results were 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 resulted in identification of different processing areas involved in 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 an action.

As useful a technique as cell recording is, it does confront limitations from the cognitive neuroscientist's perspective of trying to understand the overall operation of the mechanism

responsible for specific cognitive activities. First, since the technique is primarily correlational, it requires identifying a sensory stimulus, motor response, or ongoing cognitive activity to which the neural activity can be correlated. This is most easily done close to the sensory and motor periphery. 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 surprisingly seem to prefer irregular (non-Cartesian) shapes, which would not be natural candidates to test.[12] Second, although when successful the technique 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 correlates 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 data. Finally, and from a cognitive perspective, a very serious worry is that ethical considerations only permit single-cell recording in non-human species. This poses challenges for employing this technique to study higher cognitive functions exhibited primarily in humans.

4. Neuroimaging

This brings us to one of the newest research techniques 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.[13] The type of evidence neuroimaging provides is of the same form as that generated by single-cell recording—activity in the brain that is correlated with cognitive operations a person 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 is noninvasive and imaging can be done on normal humans while they are engaging in cognitive tasks that are thought to be most distinctively human, including abstract reasoning tasks or making ethical judgments. (The conditions for imaging are far from optimal—laying on one's back in a cylindrical tube that is making loud clanking noises. Thus, it is not possible now to image people as they perform cognitive operations in the normal contexts of their lives. This restricts the kind of cognitive tasks that can be imaged as well as raises worries about whether people perform the tasks in the same way in the scanner as in real life—a concern about the ecological validity of the studies.)

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

[13] 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). I discuss and evaluate Uttal's objections in (Bechtel, 2002b).

At the turn of the 20th century, X-rays revolutionized the medical sciences by providing non-invasive images of bone structure. (Bones blocked the transmission of X-rays, resulting in light images on photographic plates.) X-rays, however, could not provide information about brain structures since brain tissue blocks the transmission of very view X-rays. Investigators interested in anatomical structures in the brain therefore developed alternative techniques such as computerized tomography (CT or CAT scans) and magnetic resonance imaging (MRI); the latter, in particular, provides high-resolution images of brain tissue.

Imaging operations (function) as opposed to parts (structure) requires recording a signal that is related to the operations being performed in a given brain area. The two techniques that have been most successful and have attracted the most attention—positron emission tomography (PET) and functional magnetic resonance imaging (fMRI)—both measure blood flow.[14] The connection between neural activity (spiking) and blood flow is relatively intuitive—greater firing rates require more energy, which requires more metabolism to provide the energy. This in turn requires both oxygen and metabolites, which are provided by blood. But herein lies one of the current unknowns underlying neuroimaging—the mechanism by which greater blood flow is generated. What investigators do know is that the increase in blood flow exceeds oxygen demand, a critical factor in creating the blood oxygen level dependent (BOLD) contrast that provides the signal in fMRI. The details of the relationship will likely ultimately be important for understanding particular uses of fMRI and some current researchers are very concerned about the current deficit in our understanding (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. A 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 that 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*) is then employed to determine the site of the annihilation in the cut defined by the plane of the scanner.[15] 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.

---

[14] 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.

[15] 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 whose frequency reflects the particular atom and its environment.

Most of the epistemic issues concerning PET and fMRI turn on the connection to cognitive operations. The challenge stems from 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 (see Gusnard & Raichle, 2001). While people are awake, they are always cognitively active. Leave people in a scanner with no direction to do anything specific and they will think about whatever they choose. Beyond the problem posed by identifying a baseline, a more serious objection 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 called the *subtractive method*.[16] This involves imaging a person while performing two different tasks. One of these is thought to employ one (or a very small number of) cognitive operations additional to those employed by the other. Researchers then subtract the activation measured during the simpler task from that measured during the more complex task. The result is a difference image showing the increased or decreased activations measured during the second task. The level of difference (or the statistical significance level associated with the difference) is color coded in the pictures of the brain commonly shown in neuroimaging papers. The area(s) revealed in the difference image are regarded as the locus of the additional operations.

The subtractive method is well illustrated in a landmark early neuroimaging study. Steven Petersen and his colleagues (Petersen, Fox, Posner, Mintun, & Raichle, 1988; Petersen, Fox, Posner, Mintun, & Raichle, 1989) subtracted the activation produced when participants read nouns aloud from that produced when they read the nouns silently, generated verbs associated with them and pronounced these aloud.[17] Note that the second task involves two of the same component operations as the simpler task—reading the noun and pronouncing a word aloud—but adds an additional component—generating the verb. The increased activation, so the reasoning goes, reflects the performance of the additional task. The investigators found increased activation in three areas: anterior cingulated cortex, left dorsolateral prefrontal cortex, and the right inferior lateral cerebellum. They provided alternative explanations of the anterior cingulated and cerebellum increases and interpreted the left dorsolateral prefrontal cortex as the brain area responsible for the semantic operation of generating the verb.

This procedure is an adaptation of one initially developed by F. C. Donders (1868) for use in chronometric studies of cognitive process. In chronometric studies, the researcher subtracts the time taken to perform one from that required for another task and the difference is thought to reflect the time required for the additional operations required in the longer task. In its application to chronometric studies, the subtractive method was broadly criticized in the 1960s. Saul Sternberg (1969) pointed out, for example, that it assumed that the additional cognitive

[16] Alternative approaches are being developed (see, for example, Frith & Friston, 1997)

[17] Petersen and his colleagues also performed two preliminary subtractions: (1) looking at a fixation point from passively reading words and (2) passively reading words from reading them aloud.

activity was a pure insertion into a sequential set of processes, and this assumption might well be false. The additional operations may involve interactions with the operation in the simpler task, causing them, for example, to be performed more slowly. As a result, the increased reaction time may not reflect simply the additional operation. He advocated replacing the subtractive method in studies of mental chronometry with more complex 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 although neuroimaging would seem to confront the same problem of assuming a pure insertion of the additional operation (e.g., generating a verb). The additional operation could conceivably interact with and alter the operations performed in the simpler task (reading the noun or pronouncing a word). Marcus Raichle (1998) defends use of the subtractive method by arguing that imaging itself will reveal any changes in activation in other brain areas that would indicate interactions with other processes. 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 of significance 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 and this may lead them to miss less the increased activations stemming from interactions.

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 operations of encoding, storage, and retrieval. Storage is assumed to be widely distributed in cortex, but many researchers have pursued the hypothesis that encoding and retrieval are more localized 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. Attempts to relate encoding and retrieval processes to the brain through lesions, however, 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. (An episodic memory is different from a person's knowledge of being involved in the event, which the person could have acquired from the reports of others). To use the subtractive method, Tulving and his collaborators needed to develop two tasks to compare and they choose tasks that differed in 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). For the shallow encoding conditions, participants were asked whether the word contained an *a*, whereas for the deeper encoding condition they were asked whether it referred to a living thing. A subsequent recognition memory task revealed that participants recognized more of the words

when the encoding task required them to determine whether the referent was living, confirming that better encoding occurred in the deeper encoding condition. 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 (Kapur et al., 1994). To study episodic retrieval Tulving presented participants either with novel meaningful sentences or with meaningful sentences that they had heard 24 hours previously. Although the participants 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 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)

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. Tulving in fact 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 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 refined imaging procedure 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), John 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 against HERA, it remains a focal hypothesis in the literature, one whose evidential support is provided solely by 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 and appraise it, 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. Second, Tulving's

results receive little corroboration from other techniques. In particular, lesion studies, from which one might have hoped to acquire corroborating evidence, did not provide any support since lesions in the prefrontal areas do not generate deficits in either encoding or retrieval of episodic memories. So that left the results vulnerable. Third, the results did little to advance a detailed mechanism of encoding and retrieval. Such advance would require some account of the nature of the operations involved in encoding or retrieval, and to date there is little in the literature to suggest what these might be.

These shortcomings count against HERA specifically. They are not, though, grounds for pessimism about the overall enterprise of invoking neuroimaging and its potential to help develop mechanistic models of memory. Recent neuroimaging studies have identified yet more brain locations that show increased activation in encoding and retrieval studies in addition to those first targeted by Tulving. Reflecting on these results, Gabrieli 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).

That is to say, imaging may turn out to be a powerful discovery strategy—as imaging discovers multiple areas involved in a cognitive task, it motivates inquiry into what operations these areas perform. Moreover, it can also serve as a tool for answering such questions. As the number of imaging studies using different tasks increases, it becomes possible to conduct meta-analyses that reveal the range of tasks in which a given brain area is activated (Cabeza & Nyberg, 2000). By considering what operations might be common to these various tasks, imaging researchers can contribute to the project of identifying new component operations. If the neuroimaging results do result in new proposals as to the operations involved in memory encoding and retrieval, and if these results are corroborated by other approaches such as lesion studies (perhaps using TMS to create temporary lesions), neuroimaging will realize its promise as a tool for developing improved accounts of mental mechanisms.

The epistemic issues raised by PET and fMRI are of critical importance since these techniques, together with ERP and lesion studies, are the primary tools for studying the neural foundations of cognitive processing in humans. Their strength is that, unlike lesion studies, they can correlate activity in the brain with cognitive operations a person is performing. Although concerns remain about the origins of the signal measured, PET and fMRI are generally regarded as providing highly credible information about neural activity. The main epistemic issues concern the procedures for relating this information to cognitive operations. A primary procedure for doing this is the subtractive method and I have focused on some of the epistemic issues it raises. The strategies for evaluating the results of this technique are the same as I identified earlier—the determinateness of the results, consistency with results obtained by other techniques, and coherence with theoretical perspectives. One sign of determinateness is the stability of brain areas showing increased activation across tasks and studies involving the same operation. Although earlier analyses raised doubts about such stability (Poeppel, 1996), more recently

others have found considerable agreement between studies (Corbetta, 1998; Cabeza & Nyberg, 2000). Some of the most celebrated results of imaging studies identify brain areas active in a cognitive task different from those whose damage resulting in deficits on those tasks. In one respect, such differences are what one hopes for—the motivation for developing a new technique is to discover something new. The verb-generate studies of Petersen et al. generated what was at the time an unexpected finding that left prefrontal cortex was the site of semantic processing and Tulving studies indicated frontal activity in encoding episodic memories whereas lesions there do not seem to impair such encoding. The differences in results then prompt attempts to generate consistency, often by calibrating and interpreting the different results in light of each other. The final measure is that results fit into coherent theoretical perspectives, especially accounts of how neurocognitive mechanisms work. In addition to linking brain regions with component operations of cognitive processing, this requires accounts that agree with the known neuroarchitecture of the brain and show how the component operations figure in performing a variety of cognitive tasks. This is an ongoing effort in cognitive neuroscience that appears to be making rapid progress, which in turn supports the credibility of the empirical evidence imaging produces.

5. Conclusions

Evidence in science is frequently the focus of contest. This is true in cognitive neuroscience as in other sciences. I have examined the development of 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 researchers must evaluate the results as to whether they reflect in the appropriate manner the component operations in cognitive mechanisms. Otherwise, critics can dismiss the results as only artifacts resulting from the mode of intervention used to produce them. Such assessment typically does not depend upon in-depth understanding of how the technique generates results but on the nature of the results produced. I have tried to show how researchers commonly appeal to three different features of the evidence itself to determine whether results from these techniques of cognitive neuroscience 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.

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

Bechtel, W. (2002b). Decomposing the mind-brain: A long-term pursuit. *Brain and Mind, 3*, 229-242.

Bechtel, W. (in press). *Discovering cell mechanisms*. Cambridge: Cambridge University Press.

Bechtel, W., & Abrahamsen, A. (under review). Explanation: A mechanist alternative.

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

Berger, H. (1929). Über daas Elektroenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten, 87*(527-570).

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). Remarques sur le siége de la faculté du langage articulé, suivies d'une observation d'aphemie (perte de la parole). *Bulletin de la Société Anatomique, 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.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*, 1-47.

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences, USA, 95*, 831-838.

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.

Craver, C. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science, 69*, S83-S97.

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

Farah, M. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press.

Feinberg, T. E., & Farah, M. J. (2000). A historical perspective on cognitive neuroscience. In M. J. Farah & T. E. Feinberg (Eds.), *Patient-based approaches to cognitive neuroscience* (pp. 3-20). Cambridge, MA: MIT Press.

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.

Frith, C. D., & Friston, K. J. (1997). Studying brain function with neuroimaging. In M. Rugg (Ed.), *Cognitive neuroscience* (pp. 169-195). Cambridge, MA: MIT Press.

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.

Geschwind, N. (1979). Specializations of the human brain. *Scientific American, 238*(3), 158-168.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 50-71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342-S353.

Glymour, C. (1994). Methods of cognitive neuropsychology. *British Journal for the Philosophy of Science, 45*, 815-835.

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. In A. D. J. Robertson (Ed.), *Key papers in cybernetics* (pp. 91-102). London: Butterworth.

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nature Reviews/Neurosience, 2*, 685-694.

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.

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

Hubel, D. H. (1982). Evolution of ideas on the primary visual cortex, 1955-1978: A biased historical account. *Bioscience Reports, 2*, 435-469.

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.

Kutas, M., & Dale, A. (1997). Electrical and magnetic readings of mental function. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 197-242). Cambridge, MA: MIT Press.

Lashley, K. S. (1950). In search of the engram. *Symposia of the Society for Experimental Biology, IV. Physiological Mechanisms in Animal Behaviour*, 454-482.

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.

Marr, D. C. (1982). *Vision: A computation investigation into the human representational system and processing of visual information*. San Francisco: Freeman.

Mishkin, M. (1966). Visual mechanisms beyond the striate cortex. In R. W. Russel (Ed.), *Frontiers in physiological psychology*. New York: Academic.

Mishkin, M., & Pribram, K. H. (1954). Visual discrimination performance following partial ablations of the temporal lobe: I. Ventral vs. lateral. *Journal of Comparative and Physiological Psychology, 47*, 14-20.

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

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature, 341*, 52-54.

O'Keefe, J. A., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.

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

Petersen, S. E., Fox, P. J., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing single words. *Journal of Cognitive Neuroscience, 1*(2), 153-170.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(18 February), 585-588.

Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain and Language, 55*(3), 317-351.

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.

Schaffner, K. A. (2001). Extrapolation from animal models: Social life, sex, and super models. In P. McLaughlin (Ed.), *Theory and method in the neurosciences* (Vol. Pitt-Konstanz Colloquium 5). Pittsburgh: Pittsburgh University Press.

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.

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Sternberg, S. (1969). The discovery of processing stages: Extension of Donders' method. *Acta Psychologica, 30*, 276-315.

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.

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.

Wernicke, C. (1874). *Der aphasische Symptomenkomplex: eine psychologische Studie auf anatomischer Basis*. Breslau: Cohn and Weigert.

Woolsey, 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.

Zawidzki, T. W., & Bechtel, W. (2004). Legacy: decomposition and localization in cognitive neuroscience. In C. E. Erneling & D. M. Johnson (Eds.), *The Mind As a Scientific Object*. Oxford: Oxford University Press.