

Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience

William Bechtel (bechtel@twinearth.wustl.edu)

Philosophy-Neuroscience-Psychology Program, Washington University
Campus Box 1073, St. Louis, MO 63130 USA

Robert N. McCauley (philrnm@emory.edu)

Department of Philosophy, Emory University, Atlanta, GA 30322 USA

Abstract

Functionalists in philosophy of mind traditionally raise two major arguments against the type identity theory: (1) psychological states are *multiply realizable* so that there are no one-to-one mappings of psychological states onto neural states and (2) the most that evidence could ever establish is the *correlation* of psychological and neural states, not their identity. We defend a variant on the traditional type identity theory which we call *heuristic identity theory* (HIT) against both of these objections. Drawing its inspiration from scientific practice, heuristic identity theory construes identity claims as hypotheses that guide subsequent inquiry, not as conclusions of the research.

Introduction

Functionalists in philosophy argue that the type identity theory advances an unjustifiably strong account of the metaphysics of mind. Ironically, one of the first proponents of using functional criteria to identify mental states, David Armstrong, viewed functional analysis as a means for *supporting* the identity theory. The dominant versions of functionalism, though, reject the type identity of mental and physical states, since their relations are many-to-many or at least one-to-many, not one-to-one. This is known as the *multiple realizability* objection to the identity theory.

The identity theory faces another objection to the effect that empirical investigations can never establish anything more than a correlation between mental events and physical events. We shall call this the *correlation* objection. Recent discussions of consciousness (Chalmers, 1996) have pressed this objection anew. The general argument is that, at best, neurophysiological approaches isolate brain states that correlate with conscious states. They cannot justify identifying these neural states with the conscious states (especially in light of their disparate properties). They can only establish their correlation.

A richer appreciation of the course of scientific research over time and of the thoroughly hypothetical character of all identity claims in science argues for a heuristic conception of the identity theory. Identity claims typically play a *heuristic* role in science. Scientists adopt them as hypotheses in the course of empirical investigation to guide subsequent inquiry--rather than settling on them merely as the results of such inquiry. Defending *heuristic identity theory* (HIT) from both the multiple realizability and

correlation objections, we will argue that mapping at least some mental states (viz., many that figure in scientific psychology) one-to-one with physical states is a perfectly normal part of research in cognitive neuroscience and that the results often provide ample support for these hypotheses.

HIT versus the multiple realizability objection: The role of comparative studies

Underlying the multiple realizability objection is the assumption that looking across species will yield type differences in brains despite the type identities of mental states (Putnam, 1967). Perusing research in cognitive neuroscience, though, casts doubts on that assumption.

It seems obvious that when individuals from different species are in the same mental state, their neural states will differ. After all, even within the mammalian order, brains from different species clearly look different. Thus, it may prove surprising to learn that the neurobiological practice of identifying brain areas and brain processes¹ is and historically has been a *comparative* endeavor (Bechtel & Mundale, 1999). To appreciate how a comparative approach informs neurobiological proposals, consider the examples that follow (two historical and one contemporary).

The first involves research on mapping the brain into functionally relevant areas by using cytoarchitectural tools, a project now largely associated with Korbinian Brodmann, but pursued at the turn of the century by many others (e.g., Oskar and Cécile Vogt, Constantin von Economo). A key foundation for Brodmann's demarcation of areas was his demonstration that cortex generally consists of six layers, for which he reported comparative studies involving fifty-five species. He distinguished areas on the basis of the relative thickness of these layers (e.g., layer 4 was very thick in areas 1, 2, and 3, but much thinner in area 4) and the particular types of neurons found (e.g., pyramidal cells). In identifying brain areas, Brodmann worked comparatively; besides his well-known map of the human cortex (figure 1), he generated maps for the lemur, flying fox, rabbit,

¹Although philosophers often speak of brain states, neuroscientists are not generally interested in states but in brain areas and brain processes.

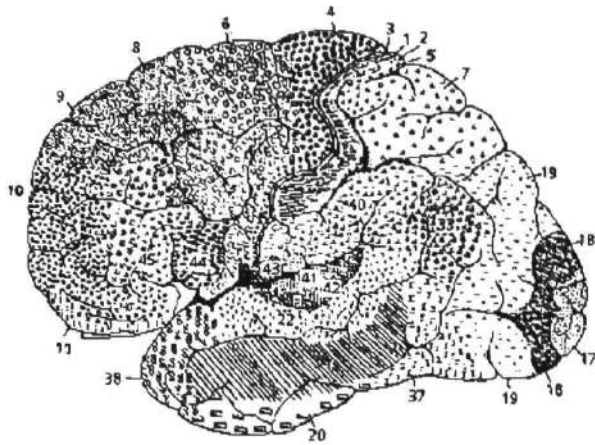


Figure 1: Brodmann's (1909) cytoarchitectural map of areas of the human cerebral cortex.

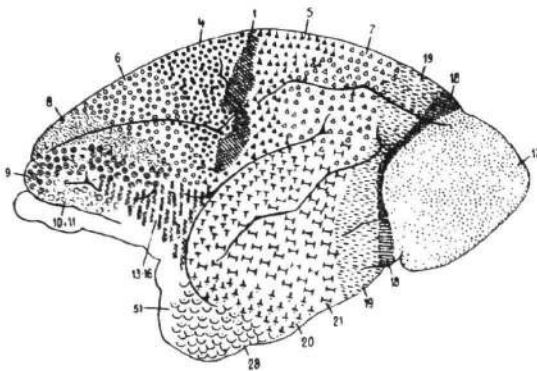


Figure 2: Brodmann's (1909) cytoarchitectural map of the lemur.

and others. (See figure 2 for an example.) For Brodmann, success in finding comparable areas in different species despite differences in brain shapes and in the relative location of areas was pivotal in establishing the reality of distinct functionally relevant areas in the brain. (Brodmann, 1909/1994).

Although Brodmann's goal was to identify functionally relevant brain areas, neuroanatomical techniques generally do not suffice to do so. Before the rise of functional brain imaging, neuroscientists primarily relied on lesion and electrophysiological techniques. David Ferrier (1876) used electrophysiological techniques in the 1870s, employing mild electrical stimulation to map brain areas in a large number of species, including monkey, dog, jackal, cat, rabbit, guinea pig, and rat. He utilized a numbering scheme shown in Table 1 to record the functional character of the responses elicited (figure 3). Although he was unable to use this technique on humans, Ferrier's goal was to extrapolate his results to humans and in his final chapter he proposed how areas found in other species project onto the human cortex (figure 4).

Historically, neuroscientific practice routinely involved identifying brain areas and processes across broad range of species as belonging to the same type. These practices continue.

Table 1. Areas Ferrier Found to Respond to Mild Stimulation

1. Opposite hind limb is advanced as in walking
2. Flexion with outward rotation of the thigh, rotation inwards of the leg, and flexion of the toes
3. Movements comparable to 1 and 2, plus movements of the tail
4. Opposite arm is adducted, extended, and retracted, the hand pronated
5. Extension forward of the opposite arm (as if reaching or touching something in front).
- a,b,c,d Clenching of the fist
6. Flexion and supination of the forearm
7. Retraction and elevation of the angle of the mouth
8. Elevation of the ala of the nose and upper lip
9. Opening of the mouth, with protrusion of the tongue
10. Opening of the mouth, with retraction of the tongue.
11. Retraction of the angle of the mouth.
12. Eyes open widely, pupils dilate, and head and eyes turn to the opposite side.
- 13,13' Eyes move to the opposite side
14. Pricking of the opposite ear, head and eyes turn to the opposite side, pupils dilate widely.
15. Torsion of the lip and semiclosure of the nostril on the same side

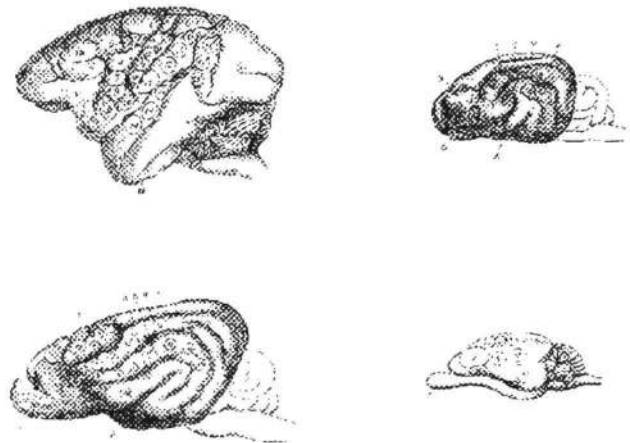


Figure 3: Areas on the cortex of monkey (upper left), dog (lower left), cat (upper right), and rabbit (lower right) where David Ferrier was able to elicit responses to mild stimulation. the common numbering system is shown in Table 1.

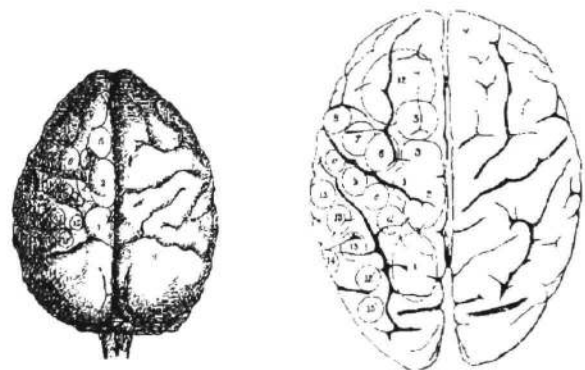


Figure 4: Ferrier's projection of areas responsive to electrical stimulation on the macaque cortex (left) onto the human cortex (right).

Maps of cortical areas have become more refined as neuroscientists have developed additional tools, such as connectivity analysis, to identify brain areas. Still, maps of, for example, visual processing areas in the brain—developed by Ungerleider and Mishkin (1982) and van Essen and Gallant (1994)—are based principally on studies of macaque monkeys.

Oddly, when they consider theories of mind-brain relations, philosophers seem to forget that the overwhelming majority of studies have been on non-human brains. Experimentally induced lesions and cell recording are two of the principal tools for unraveling the functional significance of different areas, but for obvious ethical reasons these are largely restricted to non-human animals. Although the ultimate objective is to understand the structure and function of the human brain, neuroscientists depend upon indirect, comparative procedures to apply the information from studies with non-human animals to the study of the human brain. For example, they determine the location of areas such as V4 and MT in the human brain by using neuroimaging techniques to find where tasks that would drive cells in these areas in macaques result in increased blood flow in humans (Zeki et al., 1991).

Why do the differences between the brains of different organisms, which so exercise philosophers, *not* impede neurobiological research? Part of the reason seems to be that neurobiologists often employ criteria for type identities in brains that are more coarse-grained than most philosophers have envisaged. Of course, the philosophical practice of comparing *mental* states across species is also rather coarse-grained. But no one should begrudge that. When pondering hypotheses that identify the psychological structures and processes of minds with the biological structures and processes of brains, surely one crucial issue is insuring that we compare analyses with compatible grains. Accordingly, neurobiologists do treat psychological processes (albeit not those of folk psychology, but ones that figure in information processing accounts of psychological function) as comparable across species, but they largely elude the problem of multiple realization by working with analyses from the two pertinent levels that have at least roughly similar grains (Bechtel & Mundale, 1999).² Ascertaining the compatibility of "grain" between research at two different levels is one of the most basic examples of the co-evolution of sciences.

²There are many occasions when neurobiologists employ a much finer grain. A major issue in recent years has been the plasticity of cortex, often demonstrated by the rewiring of sensory processing areas that occurs in response to altered sensory input. Even in the context of comparative studies, there are times when neurobiologists are concerned with micro-details (e.g., in measuring allometry or analyzing how connectivity changes between species). When neurobiologists move to this grain size for brain areas, though, they usually change the grain-size of their behavioral measures as well and attend to *differences* in behavior between animals or across species.

When Putnam (1967) employed for his example of common psychological states hunger in humans and octopi, his grain for type identifying psychological states was not especially fine. Such a broad extension of psychological types poses problems for the functional identification of psychological states, since the links to other mental states and to behaviors that are central to functional analyses differ profoundly between such radically different species. Still, given that evolution tends to conserve and extend existing mechanisms rather than create new ones, researchers could well end up type identifying even the neural mechanisms involved in hunger in the octopus and human, which would substantially defuse Putnam's intuitively plausible example. This is not to rule out the possibility of radically different ways of performing similar functions emerging in evolution. However, when researchers discover multiple mechanisms for performing similar functions, such as alternative pathways for processing visual input in invertebrates and vertebrates, it provides an impetus for psychologists to search for functional (behavioral) differences that motivate the differentiation of types at the psychological level as well. Acknowledging the possibility of different mechanisms performing similar functions does not preclude maintaining type distinctions that preserve one-to-one mapping between neural and psychological types. With just such lessons in mind, HIT looks to the comparative practices of neurobiology to dodge the multiple realizability objection.

HIT versus the correlation objection: Hypothesized identities as discovery strategies

Champions of the correlation objection, i.e., the objection that identity theorists can never establish the actual identity of neural and psychological states but only their regular correlation, assume (correctly) that identity theorists bear the burden of evidence in this debate. In a perversely Humean spirit, though, they set the bar impossibly high, requiring identity theorists to establish each identity claim's truth—in effect—beyond a shadow of a doubt. Discredited in the philosophy of science, verificationism, oddly, enjoys new life in the philosophy of mind.

Neurobiological practice provides direction for answering this objection too. Scientists often propose identities during the early stages of their inquiries. These hypothetical identities are not the conclusions of scientific research but the premises. They serve as heuristics for guiding scientific discovery. (McCauley, 1981) Instead of appealing to Leibniz's law of the identity of indiscernibles as a metaphysical principle for settling things *a priori*, they opportunistically exploit its converse, the indiscernability of identicals, to guide subsequent empirical research. This formulation of Leibniz's law entails that what we learn about an entity or process under one description must apply to it under its other descriptions. Scientists propose these identities, in part, precisely because the two accounts do not mirror one another perfectly. They use each to guide discovery in the other.

This involves employing what we learn through psychological research to guide the discovery and elaboration of neural

mechanisms and what we learn about neural mechanisms to develop more sophisticated psychological models. We will sketch the case of visual processing, which has involved a set of related hypothetical identities that have linked neural and psychological investigation for over a hundred years in an ongoing story of progressive theoretical revision at both levels of analysis. Researchers revised their initial identification of cortical visual processing with processing in V1 as they recognized, with the help of increasingly sophisticated neurobiological accounts, that a much larger part of cortex subserves vision; these revisions in the neurobiological account are now inspiring revisions in the psychological account of vision. For practitioners in these fields at the end of a century of research, both the comparable complexity and the general compatibility of models from the psychological and neural sides of the divide render philosophers' disquiet about the incompleteness of the evidence a needlessly fastidious extravagance. Few researchers would contest identifying visual processing with processing in the areas denoted in the figure of the flattened cortex of the macaque by van Essen and Gallant (figure 5).

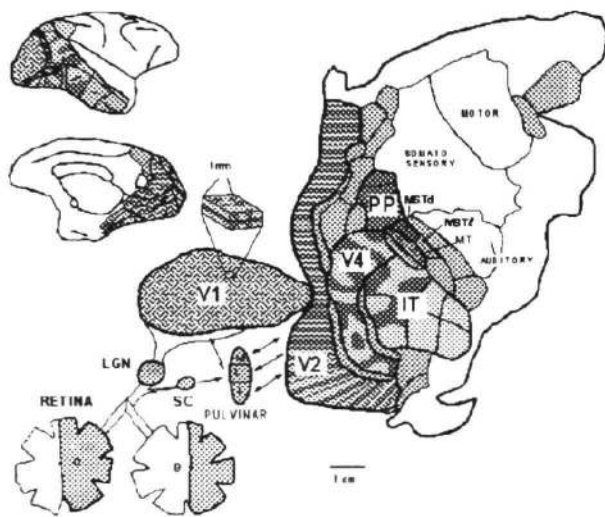


Figure 5: Van Essen and Gallant's (1994) map of distinct processing areas which figure in visual processing in a flattened macaque cortex.

Research on the neural mechanisms of vision began in the last half of the nineteenth century with efforts to locate a visual center in the brain. Based upon neuroanatomical studies indicating that the optic tract, after projecting to a part of the thalamus known as the lateral geniculate nucleus (LGN), subsequently projects to the occipital lobe (Meynert, 1870) and upon clinical evidence concerning visual deficits following stroke and other damage to the occipital lobe, most researchers identified it as *the* locus of visual processing. Ferrier, however, dissented, arguing on the basis of lesion studies in monkeys and his stimulation studies that the angular gyrus in the parietal cortex was the locus of vision. One critical piece of evidence that suggested that Ferrier was

wrong was the discovery that the organization of the occipital lobe reflected topographical layout of the visual field. Early evidence for this came from Salomen Henschen's (1893) attempt to map lesions in the occipital cortex and corresponding deficits in the visual field in humans, but the map he offered reversed the mapping contemporary scientists accept. During the Russo-Japanese War Tatsuji Inouye and during World War I Gordon Holmes and William Tindall Lister developed the modern account of the topographical arrangement of occipital cortex from their studies of wounded soldiers (Glickstein, 1988). Using single cell recording in cat and monkey Talbot and Marshall (1941) corroborated their proposals (figure 6).

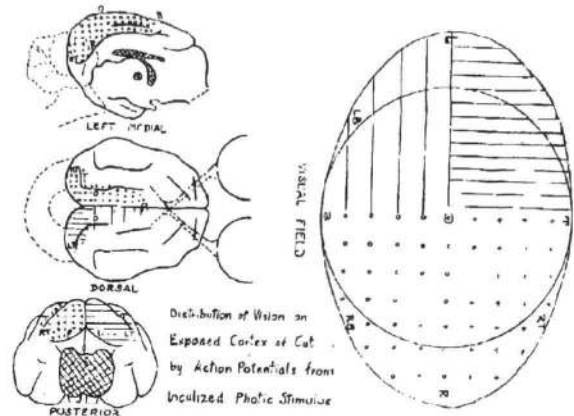


Figure 6: Talbot and Marshall's (1941) mapping of the visual field onto the primary visual cortex of the cat.

Discovering this organization in the occipital cortex supported the hypothesis that it was the location for visual processing in the brain, but it left most of the questions about how the brain processes visual information unanswered. Stephen Kuffler's (1953) research on the retina and LGN had revealed the distinctive center-surround response of cells in those areas (i.e., some cells would respond to a stimulus when it was in the center of their visual field but be suppressed when it was in their surround, while others would respond to a stimulus in the surround but not the center). Two researchers in his laboratory, David Hubel and Torsten Wiesel, set out to find similar response patterns in the occipital cortex of cats and monkeys but discovered that cells there were responsive to bars instead (Hubel & Wiesel, 1962, 1968). What they termed *simple cells* responded to bars of specific orientation at specific locations, while what they termed *complex cells* responded to bars of specific orientation at any location in a cell's receptive field and might show selective responses to bars moving in some particular direction. While Hubel and Wiesel's demonstration of specific visual function in V1 provided further support for its identification as a visual area and important details about the character of visual processing, it also showed that visual processing could not be identified with V1 alone. They ended their 1968 paper with a prophetic remark:

Specialized as the cells of 17 are, compared with rods and cones, they must, nevertheless, still represent a very elementary stage in the handling of complex forms, occupied as they are with a relatively simple region-by-region analysis of retinal contours. How this information is used at later stages in the visual path is far from clear, and represents one of the most tantalizing problems for the future. (Hubel and Wiesel, 1968, p. 242)

Although Karl Lashley (1950) had strongly resisted proposing specialized visual processing areas outside of V1, Alan Cowey (1964), relying on single cell recording, demonstrated that V2 also contained a systematic map of the topographical organization of the visual field. In 1965 Hubel and Wiesel (1965) showed that yet a third visual area, V3, preserved the topographical organization. Semir Zeki (1969) offered further evidence of the systematic nature of the maps by showing that small lesions in V1 resulted in deterioration of cells in corresponding parts of V2 and V3. In 1971 he repeated the approach by making lesions in V2 and V3 and tracing their effects into areas on the anterior bank of the lunate sulcus which he labeled V4 and V4a. Turning to cell recording, Zeki established that cells in V4 responded to the wave length of stimuli, while cells on the posterior bank of the superior temporal sulcus (an area he labeled V5 but others have designated MT) responded to motions of stimuli in specific directions (Zeki, 1973, 1974).

Various research from the 1950s to the early 1970s identified specific responses to visual stimuli in areas of the temporal cortex and of the posterior parietal cortex. Within the former, areas TE and TEO in the inferotemporal cortex responded to specific shapes (Gross, Rocha-Miranda, & Bender, 1972). In posterior parietal cortex cells responded differentially to the locations of stimuli (Goldberg & Robinson, 1980). In a relatively brief period, such research defeated the hypothesis identifying visual processing exclusively with processing in V1. Rather than undercutting the strategy of hypothesizing identities, though, determining visual function in these other areas led to more identity claims that were even more detailed and that identified various aspects of visual processing with neural processes in additional brain areas.

In 1982 Mishkin and Ungerleider proposed that visual processing in cortex followed two pathways beyond V1, a ventral pathway into inferior temporal cortex, which processed information about the identity of stimuli, and a dorsal pathway into parietal cortex that processed information about the location of stimuli (figure 7). Livingstone and Hubel (1984) extended this proposal back to the retina. Some (Milner & Goodale, 1995) have challenged the precise characterization of the processing in the two pathways, but most research (van Essen & Gallant, 1994) supports the general conception of two partially segregated processing streams.

The discovery of multiple brain areas that seem to be processing different visual information has proved the principal guide to detailed characterization of visual processing in the brain, not top-down analyses in psychology or artificial intelligence (e.g., motivated by Marr, 1982). Subsequently,

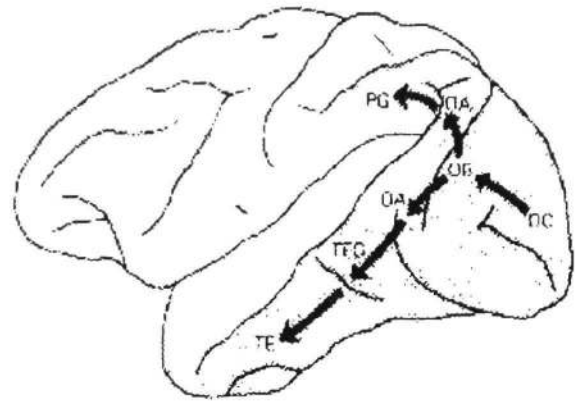


Figure 7: Ungerleider and Mishkin's (1982) proposal of two pathways in visual processing.

though, these neurobiological accounts have played a heuristic role in developing higher-level analyses of vision. Ulric Neisser (1989) was one of the first cognitive theorists to draw upon the two pathway account proposed by Mishkin and Ungerleider. Neisser construed the dorsal pathway as embodying Gibson's notion that we directly see the layout of the environment and the ventral pathway as responsible for more inferential cognitive processing (see also Milner and Goodale, 1995). Also, in the 1980s two groups of connectionist modelers developed modularized networks that separately performed *what* and *where* analyses of images on a simple retina in order to determine the computational advantages of separate pathways (Jacobs, Jordan, & Barto, 1991; Rueckl, Cave, & Kosslyn, 1989). Finally, psychologists have recently developed behavioral measures (e.g., speed of processing) capable of demonstrating the difference in pathways in normal behaving humans (Hale, 1996).

Conclusion

HIT (Heuristic Identity Theory) proposes that identity claims between psychological processes and neural mechanisms are advanced as heuristics that serve to guide further research. Emphasizing the thoroughly hypothetical character of identity claims in science, HIT focusses on the way that proposed identifications of psychological and neural processes and structures contribute to the integration and improvement of our neurobiological and psychological knowledge. Hypothesized identities advance research by suggesting new avenues for the empirical investigation of both mind and brain. The resulting empirical findings motivate scientists at both levels to tinker with their conceptions of the pertinent processes and structures. As even the brief discussion of visual processing demonstrates, these hypothetical identities evolve in response to on-going research. Explanatory and predictive successes are what justify these identity claim and what make additional theoretical and evidential resources available in future research.

In response to both the correlation objection and the multiple realizability objection, HIT stresses the importance of attending to the contributions psychophysical identity claims have made over time to progressive programs of research in neuroscience

and psychology. It is difficult to imagine that at the turn of the millennium any philosophers would regard these considerations as even secondary, let alone irrelevant, to evaluating the identity theory. We can think of no more reasonable grounds for adjudicating these matters.

References

- Bechtel, W. and Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science*, in press.
- Brodman, K. (1909/1994). *Vergleichende Lokalisationslehre der Grosshirnrinde* (L. J. Garvey, Trans.). Leipzig: J. A. Barth.
- Chalmers, D. (1996). *The conscious mind*. Oxford: Oxford University Press.
- Cowey, A. (1964). Projection of the retina on to striate and prestriate cortex in the squirrel monkey *Saimiri Sciureus*. *Journal of Neurophysiology*, 27(1964), 366-393.
- Ferrier, D. (1876). *The functions of the brain*. London: Smith, Elder, and Company.
- Glickstein, M. (1988). The discovery of the visual cortex. *Scientific American*, 259(3), 118-127.
- Goldberg, M. E., & Robinson, D. L. (1980). The significance of enhanced visual responses in posterior parietal cortex. *Behavioral and Brain Sciences*, 3, 503-505.
- Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 35, 96-111.
- Hale, S., Chen, J., Myerson, J. and Simon, A. (1996). Behavioral Evidence for Brain-based Ability Factors in Visuospatial Information Processing. Presentation at the Psychonomics Society.
- Henschen, S. E. (1893). On the visual path and centre. *Brain*, 16, 170-180.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 28, 229-289.
- 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 Physiology*, 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.
- Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science*, 15, 219-250.
- Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16, 37-68.
- Lashley, K. S. (1950). In search of an engram. *Symposium on Experimental Biology*, 4, 45-48.
- Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, 4, 309-356.
- Marr, D. C. (1982). *Vision: A computation investigation into the human representational system and processing of visual information*. San Francisco: Freeman.
- McCauley, R. N. (1981). Hypothetical identities and ontological economizing: Comments on Causey's program for the unity of science. *Philosophy of Science*, 48, 218-227.
- Meynert, T. (1870). Beiträge zur Kenntniss der centralen Projection der Sinnesoberflächen. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, Wien. Mathematisch-Naturwissenschaftliche Classe*, 60, 547-562.
- Milner, A. D., & Goodale, M. G. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Neisser, U. (1989). *Direct perception and recognition as distinct perceptual systems*. Paper presented to the Cognitive Science Society 1989.
- Putnam, H. (1967). Psychological Predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind and Religion*. Pittsburgh: University of Pittsburgh Press.
- Rueckl, J. G., Cave, K. R., & Kosslyn, S. M. (1989). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience*, 1, 171-186.
- Talbot, S. A., & Marshall, W. H. (1941). Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophthalmology*, 24, 1255-1263.
- Ungerleider, L. G. Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Ed.), *Analysis of Visual Behavior*. 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.
- Zeki, S. M. (1969). Representation of central visual fields in prestriate cortex monkey. *Brain Research*, 14, 271-291.
- Zeki, S. M. (1973). Colour coding of the rhesus monkey prestriate cortex. *Brain Research*, 422-427.
- Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology*, 236, 549-573.
- Zeki, S. M., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C., Frackowiak, R. S. J. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience*, 11, 641-649.