

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/5668509>

# Neurophilosophy: The early years and new directions

Article in *Functional neurology* · October 2007

Source: PubMed

---

CITATIONS

22

READS

386

1 author:



[Pat Churchland](#)

University of California, San Diego

127 PUBLICATIONS 5,952 CITATIONS

SEE PROFILE

# Neurophilosophy: the early years and new directions

**Patricia Smith Churchland**

Department of Philosophy, University of California  
San Diego, La Jolla, CA, USA

Corresponding author: Prof. Patricia Smith Churchland  
Department of Philosophy, University of California,  
San Diego, La Jolla, CA 92093-0119, USA  
E-mail: pschurchland@ucsd.edu

Accepted for publication:

## Summary

**Neurophilosophy embraces the hypothesis that what we call “the mind” is in fact a level of brain activity. A corollary of this hypothesis states that we can learn much about the reality of mental function by studying the brain at all levels of organization. Until fairly recently, many philosophers preferred to believe that important domains of mental function could never be addressed using the tools of empirical science. Nevertheless, co-evolutionary progress by psychology and the neurosciences on many topics, including consciousness, free will and the nature of knowledge, have meant that such convictions need to be updated. Some large-scale mind-brain problems have not yet been solved, and do require significant theoretical innovation. In particular the problem of how to understand the true nature of representations remains unsolved.**

*KEY WORDS: consciousness, folk intuitions, homeostatic emotions, information theory, neurophilosophy, representations.*

## The Early Years

Neurophilosophy emerged as a subfield once the neurosciences had developed to the point at which progress could be made on philosophy's large-scale questions – what is knowledge, how do we make decisions, what is the nature of consciousness, thought, and ideas? Since Hippocrates (1) (460 BC-370 BC), neurological observations of functional deficits following brain damage have inspired very general ideas about the brain basis of behavior. Detailed hypotheses concerning *mechanisms*, however, had to await the advanced technology that permitted investigation of neurons and their organization (2-5). Lesion studies did lay essential foundations, but understanding neurons – their anatomy, how they communicate with one another and with muscles and glands, how they change over time in development and in learning – proved indispensable to making further progress on how the mind/brain works. Advances in neurophysiology depended on very sophisticated technology for

several clear reasons: *first*, neurons and synapses are micron-scale, spatially extended, dynamic structures; *second*, their signaling events happen at the millisecond timescale; and *third*, neurons are densely packed (about 100,000 neurons per mm<sup>3</sup> of cortex) and prone to fast degeneration. In contrast to early astronomy, for example, where revolutionary progress was made with simple lens telescopes, progress in neurophysiology depended on an understanding of the physics of electricity, and the invention of sophisticated tools such as microelectrodes and amplifiers.

Natural philosophy, from ancient times until the nineteenth century, encompassed all natural phenomena, including the nature of fire, life, the stars, tides and of course, the mind. Moral philosophy, on the other hand, targeted social phenomena – politics, ethics, and how to live a good life. By the end of the nineteenth century, many inquiries had become experimental and systematic, and thus hived off the hub of natural philosophy to become the separate disciplines now familiar to us – astronomy, chemistry, physics and biology. With the pioneering studies of the mind by Helmholtz (6) (1821-1894) and Wundt (7) (1832-1920), psychology began to emerge as a discipline in its own right. Anatomical studies of the physical properties of neurons, such as those undertaken by the great masters, Cajal (8) and Golgi (9), boosted the prospects, in the long term, of explaining psychological phenomena as brain phenomena. Innovative physiological studies in the mid-nineteenth century by Hitzig and Fritsch (10) identified a motor area of cortex in the dog by stimulation with electrical current to produce muscle contraction. Their results added credence to the then-controversial hypothesis that nervous tissue worked by electrical signals rather than by “animal spirits”, and also to the hypothesis that regions of the brain were specialized for specific functions.

## Is the Mind a Natural Phenomenon?

Philosophical inquiry into the nature of mental functions has followed two largely parallel lines of thought. One line, typified by Plato (11), Descartes (12) and, most recently, Chalmers (13), embraces the idea that the mind cannot be a physical thing, but must be ontologically distinct from the physical brain. Something like this is also espoused by Searle, though he prefers to characterize the ontologically special thing in terms of properties (subjectivity). The naturalist tradition, typified by Hippocrates (1), Aristotle (14), Hume (15), Quine (16), Crick (17), and living physicalists such as P.M. Churchland (18-21), Dennett (22-25), Bechtel (2-4), and myself (5, 26), favors the hypothesis that mental functions map onto to a certain kind of physical organization. In other words, that they are brain activities.

For all dualists, a pressing and unsolved problem concerned how the non-physical soul and the physical brain could possibly interact, given their lack of shared properties. Descartes (12) argued that the soul must be immaterial, mainly on the grounds that reasoning, in his view, is uniquely non-mechanical. His conclusion was based on the belief that reasoning allowed for novelty, whereas mechanisms were restricted to finite capacities. His second argument depended on the idea that free choice is free of the causal constraints typical of mechanisms. A real choice, in Descartes' view, could not be the outcome of causal processes. On the question of choice, Searle (27) defends a position rather similar to that of Descartes, in that he takes subjectivity (having experiences) to be inexplicable in terms of objectivity. In addition, Searle argues that nature allows for causal "gaps" in the human nervous system activity and that, by virtue of these gaps, the self can choose which reason to act on, and thus exercise free will.

According to Descartes' own formulation of dualism, the interactions between body and soul are limited to input of sensory signals, and output of motor commands: the soul, not the brain, is responsible for thinking, feeling, deciding, perceiving, and consciousness. Descartes was painfully aware that his hypothesis lacked any account of the interaction of a completely non-physical thing (the soul) with a physical thing (the body). With the advent of modern physics, the insuperable difficulty for the Cartesian conception remains its complete inability to explain the means whereby the non-physical soul (or non-physical properties) can have effects on the physical world without violating the law of conservation of mass-energy. Precisely the same problem plagues contemporary ideas concerning ontologically special properties believed to inhere in conscious experience.

Detailed studies of focal lesions in humans have revealed very specific dependencies between high-level capacities, such as speech, socially appropriate behavior, and facial recognition on the one hand, and specific areas in the brain on the other. As a result, Descartes' idea that the brain was a simple reflex machine connected to a non-mechanical creative soul came to seem deeply implausible. These dependencies suggested that the brain itself was actually performing high-level functions, not merely acting as a sensory-motor interface, as Descartes had proposed. The unique role Descartes assigned to the soul shrank as understanding of the role of the neural networks in the brain expanded, and the idea of decision making occurring in a "causal vacuum" became increasingly implausible as properties of neurons were systematically explored. Additionally, developments in comparative biology, evolutionary biology and genetics suggested that human brains are actually very similar to those of our evolutionary ancestors. Our brains may be larger than some, but they are structurally very similar to the brains of all mammals. In this context, it is worth recalling that humans have about 28,000 genes, and we differ from mice in only about 300 genes. Taken together, these developments made it seem highly probable that it is the *brain* that thinks, feels, decides, sleeps, and dreams. The vertebrate genome, from teleost fish to humans, is remarkably conserved, and this unexpected conservation, along with profound similarities in vertebrate nervous system anatomy and chemistry, suggests that how animal brains do these

things is probably not radically different from how human brains do them. Many contemporary philosophers, while explicitly denying Cartesian sympathies, nevertheless do not accept the hypotheses that mental functions are brain functions. Instead, they believe conscious awareness and subjectivity to be uniquely non-neural phenomena. Some suggest that neuronal activity might *cause* conscious events, but that conscious events themselves cannot be identical to any neural activity whatsoever. Despite some significant differences between them, Searle (28), Nagel (29), Chalmers (13) and Jackson (30,31) share the conviction that conscious processes can never be explained in neural terms. What are their reasons?

In their opinion, common sense tells us that the dissimilarities between your observing my neurons firing and my seeing blue are so blindingly obvious and so self-evidently profound that they imply an *ontological* difference (an unbridgeable difference in terms of the nature of the *stuff*) between the two. This argument depended on the claim that the nature of what I know when I see blue is entirely different from the nature of what I know when I observe neurons firing, and that this is obviously an *ontological* difference (a difference in the nature of the reality). In other words, the epistemological difference (difference in how we know things) is such as to imply an ontological difference (difference in the kinds of stuff there really are).

Nagel (29) famously expressed this by saying that he – and we – could never know what it is like to be a bat, even though he – and we – might understand bat neurons and bat brains. Ergo consciousness cannot be neurally explained. Jackson (30,31) argued that a neuroscientist could (in principle) know everything about how the brain worked, but if he had never experienced the color blue, he would still not know what it was like to see blue; ergo consciousness cannot be neurally explained. Chalmers (13) favored this thought experiment: a zombie could exist who was exactly like me in every neural and behavioral respect, including in the capacity to use sentences such as "that is a beautiful pink flamingo" and "I have a sore throat", but who actually had no conscious experiences whatever; again, ergo consciousness cannot be neurally explained. Searle (28) claimed that the reduction of subjectivity to objectivity was obviously absurd and hence could be known to be false; ergo consciousness cannot be neurally explained. This view was especially odd in Searle's case, as he explicitly rejected dualism, thus raising the issue of inconsistency.

Neurophilosophers responded to each of these arguments (22-25,32). One observation was that all these views actually make predictions about what science can and cannot discover on the basis of what we know now, all believing that, by looking intensely at the problem, they can determine what we *cannot* know and explain. Worse, they share a weakness, namely that they rely upon an argument based on ignorance. That is, they all take as a premise a claim about what we do *not* know, or cannot imagine or cannot conceive of, in order to reach the conclusion that we *do* know that consciousness cannot amount to neuronal activity and must be ontologically unique. Additionally, despite the parade of examples from the history of science where the "intuitively obvious" and the "self-evidently true" were shown to be factually false (light really *is* electromagnetic radi-

ation, the Earth *does* move, space is *not* Euclidean), these contemporary philosophers adhere to the folk intuition that brain activity and mental experiences are too different to permit a neural explanation of mental events (32).

As I see it, in general intuitive judgments are insufficiently reliable to launch definitive predictions about whether in the future an explanation of some phenomenon will be forthcoming. History is full of wonderful examples of intuitions about the nature of reality rent asunder by cold hard facts. Thus it seems intuitively very obvious that 'down' is an absolute direction (point to your feet), that two events either happen at the same time or they do not, without relation to any reference frame), or that 'wanting' cannot be dissociated from 'liking' (33,34). Yet science shows that our intuitions have erred on all these matters, and in some cases, psychology can give us explanations of *why* we have the intuitions we do, compelling and in error though they may be. Additionally, it is sobering to realize that neuroscience has made non-trivial progress on many issues related to conscious experiences: the neural differences between being awake, being in deep sleep, and being in REM; on the cellular mechanisms of top-down attention; on what happens during anesthesia, or during coma; on neural events occurring during recollection of past episodes; on the differences between a visual stimulus one is, or is not, aware of. Clinging to common-sense intuitions as evidence for the "forever unexplainable" status of conscious events is, from a scientific perspective, evidently a weak strategy in the long run, and as real neuroscientific progress is made on the problem, the strategy is correspondingly diminished (35-38).

Although the foregoing are general comments, detailed responses to the specific claims of ontological uniqueness of conscious events were forthcoming. More specifically, to Jackson, it was pointed out that experiencing the color blue presumably involves a specific pattern of activity in, say, V4, among other places (32). Now a discursive understanding of the hypothesis that specifies that V4 activity is required could not *itself* be expected to activate V4, though it might activate some other region, such as the dorsolateral prefrontal cortex. Thus, as Eddie Hubbard once remarked in conversation, assuming that discursive neurobiological knowledge of the conditions for experiencing blue is sufficient to activate V4 neurons so that we do visually see blue is tantamount to assuming that a student who has understood everything there is to know about Newton's laws of gravity will fall down. Or, one might say, if he does not fall down, then the laws of gravity are ontologically unique. To take another analogy, it is like assuming that if an intern learns everything there is to know about pregnancy, either this knowledge will itself make her pregnant or pregnancy is ontologically unique. This argument is effective against Jackson precisely because the crux of his idea is the supposition that if discursive understanding of the neurobiology of color perception does not itself produce color perception, then color perceptions are ontologically special.

Many contemporary dualists also shared a rhetorically convenient misunderstanding. Seen most starkly in Searle (28), this misunderstanding was that if a science reduces a macro phenomenon to a micro phenomenon, then the macro phenomenon is not real or "goes away".

Using this construal of "reduction" it was then argued that because it is pretty obvious that conscious experience is real, it cannot, therefore, be reduced to activity of neurons. The heart of the misunderstanding concerns the idiosyncratic notion of reduction, where it tends to be assumed that in science reductions make things disappear. This assumption is just confused.

Famously, physics reduced visible light to electromagnetic radiation, but no one believes that light therefore ceased to be real or became scientifically unworthy. Temperature was reduced to mean molecular kinetic energy, but temperature did not disappear. Some beliefs about the nature of light and temperature did change, but the important point is this: reduction of a phenomenon traditionally means only that we have an *explanation* of the phenomenon. And explanations of events do not normally make them go away. Of course sometimes scientific progress does require us to rethink what we believed about things. Thus the explanation of Huntington's disease in terms of a single gene and its effects in the brain means that we no longer think of Huntington's patients as possessed by devils (5,26,39). Given the aforementioned confusion about "reduction", the expression "the reduction of A by B" might usefully be replaced by the expression "the explanation of A by B" to remove the possibility of getting sidetracked by equivocation (40,41).

### Neurophilosophy: Facts can Trump Folk Intuitions

The main idea of neurophilosophy can be expressed quite simply: if you want to understand the nature of the mind, you need to understand the nature of the brain. By now, this idea seems uncontroversial in neuroscience and psychology, and a slowly growing number of younger philosophers are beginning to find it fruitful. In general, however, philosophers in the mainstream largely, or completely, ignore developments in the brain sciences, e.g. Block (42,43). As recently as 2003, Fodor, certainly prominent and every inch a mainstream philosopher, unapologetically advised: "If you want to understand the mind, study the mind... not the brain, and certainly not the genes" (44). As a sociological observation, I think that many students who see promise in neurophilosophy tend to pursue studies in neuroscience or psychology, and this means that those who go to graduate school in philosophy are more likely to prefer the Fodor paradigm of studying the mind.

### Neurophilosophy embraces three additional ideas

*First*, if you want to understand the nature of the brain, you need to study many levels of organization, from molecules, to synapses, neurons, micro networks, macro networks and systems. Within the many neurosciences, this is not controversial, and the many outstanding textbooks, journals and anthologies reflect this (45). A corollary of this point about levels is that psychological research provides essential characterizations of the phenomena that need to be explained. Because psychologists may fear that developments in neuroscience mean they are apt to be marginalized, this fact should be emphasized (46).

Second, because figuring out the mechanisms whereby structure yields function is very difficult and because very new ways of conceiving the problems are needed, it is essential to compare brains and behavioral capacities across many species, including other primates of course, but also birds (especially, perhaps, corvids), rats, voles, dogs, and also invertebrates, such as the leech, the fruit fly and the pit viper. Hypotheses about brain organization are sometimes provoked by observations in the much simpler nervous system of invertebrates. For example, in the leech, swimming and crawling are two very different behaviors, which suggests the existence of command neurons in different modular networks. This suggestion turns out not to be how nature solved the problem. As it happens, the very same populations of neurons are used in both behaviors, and what differs is the timing and orchestration of their activity. This result raises the question of whether in the mammalian brain, different functions – perhaps declarative memory and planning the future, for example – may likewise deploy the same populations of neurons with differing timing and orchestration of their activities (47). Briefly, I shall list just three other results that inspire new ways of thinking about any problem about the brain. Ag-

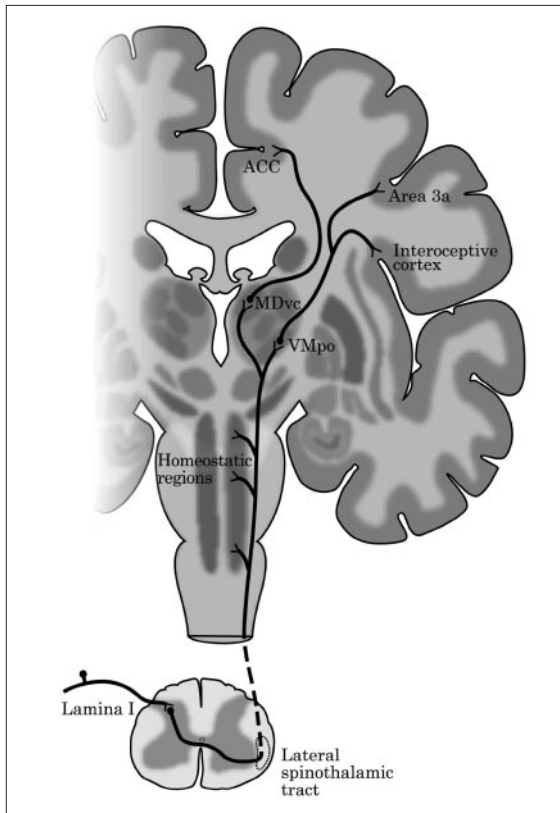


Figure 1 - The pathway, unique to primates, from lamina 1 of the spinal cord, connecting to brainstem nuclei, and projecting to the VMpo nucleus of the thalamus which in turn projects to the dorsal posterior insular cortex (interoceptive cortex) and area 3a. This pathway provides a cortical image of all homeostatic activity. The older pathway typical of all subprimates goes through the DMve of the thalamus, which then projects to the anterior cingulate cortex. Reproduced, with permission, from Craig AD (52).

gression levels in fruit flies can be manipulated by small changes in a rather large class of genes, none of which is an “aggression” gene (48). In the pit viper, though the image recorded by neurons in the infrared pit is very fuzzy, in the tectum the image undergoes an upgrading of precision, probably relying on stored information collected by the visual system during daylight (49). Ravens can solve a seven-step problem in one trial and, as they cache food, show that they are sensitive to the point of view of potential thieves (50). There are many more results that are both unsettling and inspiring in the way these are, but this sample will suffice for now.

Third, we need to understand the evolution of nervous systems and we need detailed anatomical and physiological comparisons between species. Two major findings, both anatomical and comparative, have greatly deepened how I think about human capacities for awareness and for social behavior, and how these two functions may be interconnected. The first is owed to A.D. Craig (51,52), and the second to John Allman. I shall discuss them in that order.

Craig has shown that there is an important but little studied anatomical pathway, emerging from the lamina I region of the spinal cord and connecting to a wide range of brainstem nuclei, and, uniquely in primates, projecting to the VMpo and VMB nuclei of the thalamus. These thalamic nuclei in turn project to the dorsal posterior insular region of the cortex (Fig.s 1 and 2). In the insula, there appears to be an integration of signals from diverse sources. In most subprimates, the lamina I signals are highly integrated in the brainstem and project to medial thalamic nuclei, and then on to the orbitofrontal cortex and anterior cingulate cortex. Primates also have this pathway, but they in addition, they have the pathway to the dorsal posterior insula via the VM nuclei. What is the function of this pathway? Drawing on physiological, clinical, and imaging data, as well as the

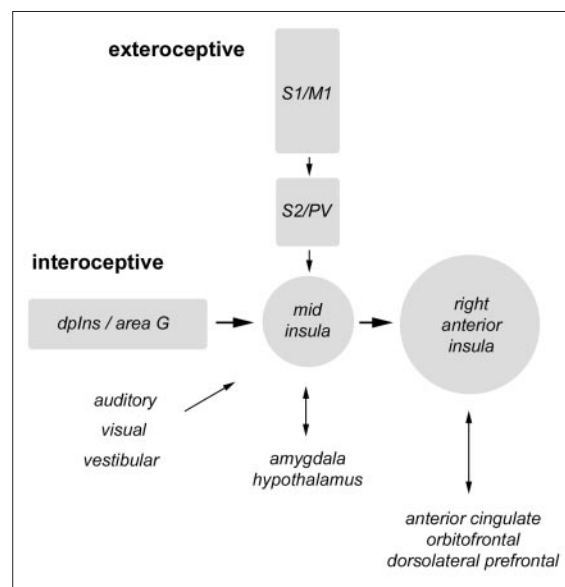


Figure 2 - Schematic representation showing the convergence of signals on the insular cortex. Reproduced, with permission, from Craig AD (52).



anatomical data, Craig proposes that neurons in this pathway represent the state of the body, including the emotional, motivational and interoceptive states, but also temperature, and various kinds of pain (muscle, gut, skin, burn, and excessive cold). As Craig puts it, the result of this integration in the insula is a representation of "the material me". That is, the insula represents how things are with me and whether I need to do something to protect my physical integrity, which includes my long-term integrity and my social functionality (51-53). The story is, moreover, even more interesting because within this pathway, the sympathetic (so-called "fight and flight") system projects almost exclusively to the right insula, and the parasympathetic ("rest" and "digest") system to the left insula.

Within the insula, the representations of the "material me" become increasingly complex from posterior regions that reflect the complexity of the bodily state, to anterior regions, that integrate sensory signals about the environment and their significance for "me", thereby yielding a more comprehensive picture of "me-in-my-world". Craig's anatomical data fit very well with the wide range of data from human subjects, including studies showing awareness deficits (hemineglect, anosagnosia) following right insular damage. More generally, his data refine and extend Damasio's somatic marker hypothesis that emphasizes, as James (54) did, the connection between emotions and the state of the body, as represented by the autonomic system (55). Craig uses the expression "homeostatic emotions" to cover the diverse feelings that are both motivational, urging the animal to do something (e.g. approach, avoid, seek warmth, seek water), and qualitative (e.g. the pleasurable feeling of warmth when chilly, the unpleasant feeling of warmth when hot, and the fearful feeling when perceiving a potential threat).

The lateralization of autonomic projections to the insular cortex fits with the clinical and imaging observations that well-being-of-me activities are more associated with the right insula, whereas social affiliative behaviors, including communicative behaviors, are more associated with the left regions, including areas 44 and 45. But what is the evolutionary significance of this separation? Why did it emerge in hominids? Integrating results from psychological as well as neurological investigations, Craig suggests that "... our capacity for awareness of our emotional behaviors (that is, our subjective feelings) evolved because it enabled an enormous enhancement in the efficiency and complexity of emotional communication" (52). Since there is, almost certainly, both sophisticated self-representation and significant communication in mammals and birds, what precisely is Craig's point here? [See also (56)].

At the risk of misunderstanding Craig, I propose that his crucial idea is this: a more regimented separation of self-representation and other-representation allows a greater degree of representational distance between *self* and *other*, and thus allows for greater flexibility in social behavior as one contrasts "what *that person* wants and believes" with "what *I* want and believe". Crudely, if mirror neurons help me understand your goals and feelings via simulation in my brain, then I am apt to feel your fear when I should not, and I am apt to share your goal or imitate your behavior when I might do better to go my own way. The greater my capacity both

to understand what you are feeling and wanting, while retaining the capacity to keep separate what *I* feel and want, the greater my success, in both the social and the non-social domains. One simplified way of conceiving how this might be achieved by brains is via an anatomical separation between self-regarding and other-regarding business. Surprisingly perhaps, the capacity to distinguish saliently within one brain *your* need from *my* need may also entail a certain suppression of the effects of mirror neuron stimulation). In other words, mirror neurons enable me to predict your behavior because I represent you as feeling this or that in part via *my* feeling this or that and predicting my behavior. But I had better do what *I* need to do, not what *you* need to do. Predicting what you will do is important, but if the mechanism for achieving that prediction cannot keep what you are feeling distinct from what I am feeling, then I shall be in a muddle.

In this hypothesis, aspects of the representation of *the material me* got cast as the *immaterial me* by Descartes' reflections on the contrast between my pain (which I know I have, because I experience it) and your pain (which I infer you have, given your behavior, but which I do not experience).

I have tried to explain a possible relation between subjective awareness and lateralized insular organization, but Craig's own explanation, albeit brief, adds weight to the idea that insular organization in hominids is related to awareness because it enhances *communication*. How does this communicative part of the hypothesis go? I am uncertain how precisely to answer this question, but the conventional wisdom is that, in a general way, social complexity increases with increasing cognitive resources. Perhaps, however, the heightened contrast within the brain between the internal running story of *my* needs and internal states, and the internal running story of *others'* needs and internal states mediated, *inter alia*, by mirror neurons, allows greater scope for deception, repression of predictable responses, context-sensitive bluffing, and so forth. Perhaps more importantly, it yields greater scope for joint planning of complicated collective undertakings, such as attacks on other groups or defense against attacks from other groups, hunting of large beasts, and so forth. Preparation for business of this kind is likely enhanced by the capacity to envisage and plan for the long-term motives and goals of others, and the capacity to share the plans with group members.

The second anatomical finding I referred to above concerns an unusual class of bipolar cortical cells that Allman and colleagues have labeled, *Von Economo neurons* (VENs) (57). These cells differ from the typical cortical pyramidal neurons by virtue of their spindle shape, large size, and bipolar morphology (a single long apical dendrite, and long basal dendrite, with little "branchiness") (57). VENs appear to be restricted to the anterior cingulate cortex (ACC), the frontal insula (FI), where they are found in layer Vb, and the claustrum.

Von Economo neurons have attracted attention for several reasons, one of which is that both the ACC and the FI are known to play a role in decision making, planning, emotion and social evaluation (the claustrum is more of a mystery). The other reason is that though the population of VENs is dense in human brains, it appears to be sparse in chimpanzees and orangutans, and VENs are not found in any monkey species studied so far. Togeth-

er with Craig's anatomical data (see above), these findings seem to point towards a recent evolutionary modification related to very fast goal evaluation and decision making, especially in the context of self-regarding and socially intricate actions. A more precise characterization of their functional role is needed, and awaits further research.

### New Directions: Coding and Representing

Many important questions in systems neuroscience remain unanswered (58), and though not all are especially pertinent to traditional philosophical questions, some certainly are. One central problem is this: what is it for neurons, individually or collectively, to represent? What, for the brain, are representations?

Speaking informally, it is generally agreed that neurons represent states of the outer world, as well as states of the body *and* other states of the nervous system, including states in the brain stem and limbic system. In this sense, a representational state of the brain is important for what it signifies beyond itself – for what it points to, one might say. And in this sense, it carries meaning. The kidney, by contrast, is essentially a filter, and does not represent anything. It is because neuronal activity can be about something in the external world, such as the presence of a predator, or something in the internal world, such as low glucose levels, that body movements can be informed rather than random. Representations permit predictions, and predictions serve survival and the passing on of one's genes.

The "pointing-beyond-itself" feature of the brain must ultimately be understandable in terms of neural mechanisms. Although Cartesians tend to believe that only an ontologically special state could "point beyond itself", and that such a property involves an "intrinsic intentionality" that is impossible to explain neuroscientifically, the poverty of evidence to support the Cartesian hypothesis has sapped its appeal. Therefore, the explanation for how neurons represent is most likely to be naturalistic, not involving something "ontologically unique". A basic problem, therefore, is to understand the complex physical arrangements among neurons, and between neurons and the world, such that certain properties of those arrangements constitute the physical basis for neurons' "pointing beyond", for their being *about* something (such as the approaching cougar). Where do things stand in neurobiological theory concerning neural representations?

The powerful mathematical resources of information theory as used by communications engineers (59-61) have often seemed the most promising tools for characterizing precisely the relationship between neuronal responses and effective stimuli. What theorists hoped was that these resources would allow them rigorously to address what is coded in a single neuron's response to a stimulus, how reliable the neuron's code is, and how to estimate the quantity of information transferred. In what follows, I shall briefly outline the approach, and then explain why I am skeptical that it can grow into a theory of representations in general, and thus why additional resources are needed.

Mutual information is a measure of the degree of statistical dependence between two random variables, such

as between thunder and lightning, or between smoking and lung cancer. The concept is rooted in probability theory, and is useful because it will measure any deviation from independence, whether owed to linear correlation or to some non-linear dependency. If the probability of a response  $R_j$  occurring, given that stimulus condition  $S_k$  obtains (i.e.  $P(R_j | S_k)$ ) equals the probability of  $R_j$  occurring all by itself, then the mutual information is 0; i.e. the occurrence of  $R_j$  does not carry any information about whether  $S_k$  occurred or not. When  $R_j$ 's occurrence is dependent on  $S_k$ 's occurrence, then we can talk about  $R_j$  as carrying information about  $S_k$  (as opposed to some other stimulus,  $S_i$ ). Thus, if  $R$  will not occur unless  $S$  occurs, this means that if we know that  $R$  did occur, we learn something about whether  $S$  occurred.

As applied to neurons and what they code, the idea is to determine the dependence of a neuron's spiking responses on the presence of a specific stimulus by observing the neuron's behavior in a range of stimulus conditions. The observed data allow you to calculate the probability of a particular response  $R_j$  given a particular stimulus  $S_k$ , as compared to the probability of  $R_j$  happening *all by itself* (the unconditional probability of  $R_j$ ). This means we can talk about  $R_j$  coding for  $S_k$ . What I have presented is a highly simplified characterization the information theory (IT) approach, and it is important to know that it has been well developed in many ways in order to address a large range of complexities (59,62).

A slightly different but related approach suggests that we take the "brain's eye view", calculating the probability of the occurrence of  $S_k$ , given that the neuron is responding with  $R_j$ . In other words, we can say a neuron's response  $R_j$  carries information about stimulus  $S_k$  if and only if  $S_k$  is predictable from  $R_j$  (where predictable can be defined rigorously in mathematical terms). The reconstructive or "brain's eye view" approach has many variations, and many names: "optimal estimation", "Bayesian classification", "Bayesian estimation" and "ideal observer analysis" (IOA) (59,62). For convenience, I shall focus on IOA. Though there are significant differences between IT and IOA, for the purposes of this discussion these differences do not matter.

The background hope is that just as neuroscientists have probed further and further in from the sensorimotor periphery of nervous systems, so IT and IOA may be applied at deeper and deeper levels, ultimately to yield a comprehensive theory of representation. As I discuss below, a profound and entirely unsolved problem is whether this strategy is suitably extendable to representations in general. In particular, I shall draw attention to the fact that some important representational functions are stimulus independent to varying degrees.

### Representing Relevant Things

Hopfield (63) has made the point that we cannot understand biological computation unless we understand how the brain ignores information irrelevant to its needs, but invests in behaviorally relevant information. The IT/IOA conception of information completely lacks the resources to deal with relevance-to-need, and Turing-style computation lacks the means to accommodate needs and motivations. Brains, by contrast, are need-driven and relevance sensitive, and a vast amount of the brain's activity is not stimulus-driven at all.

### Representing Absent Things

Psychological experiments with humans and other animals show that in fact brains represent absent objects – not just currently existing stimuli (64). Absent things represented include goals – one’s own as well as those of others. They include distant spatial locations, past events, and future events. Rats can represent regions of a maze not yet visited, on the current excursion (though visited the previous day) and birds, such as Clark’s nutcrackers, can represent scores of locations where cached nuts still remain after other sites have been emptied. Scrub jays represent what can be seen from another jay’s perspective. In sleep, the rat brain repeats a rewarding route in a maze, suggesting that the activity of the neurons represents certain spatial locations in the absence of the stimulus. A monkey may have a goal of obtaining a piece of fruit, perhaps representing the goal by a motor image of hand-grasping-fruit, and may plan steps to get it. A chimpanzee may have a goal of lurching on termites, and methodically prepares a suitable dipping stick to implant down the termitary. A troop of chimpanzees with the collective goal of catching a tasty monkey cooperate to surround the monkey and block its escape, thus representing the strategies of other chimps as part of the means to achieve the goal. Consider the orangutan, Ken, San Diego Zoo’s legendary escape artist. He once stashed a small object in his mouth during the day and after dark used it to pick the cage lock successfully to escape. His search for a suitable trifle to function as a pick, and his hiding it from the keepers by putting it in his mouth, suggests his brain represented a goal (to escape from the enclosure) and a means of achieving the goal (waiting until the zoo staff had gone and then picking the lock). Goal representations and predictive representations are paradigmatic examples in which the standard causal relationships between external stimulus and neuronal representation characterized by the IT/IOA model break down.

A host of related problems arise when we consider certain sensory illusions, particularly those where the nervous system constructs a feature, representing it as part of the perceived object in order to make sense of the sensory signals. Subjective contours – contours that do not exist in the stimulus but are perceived – have been well-studied by psychologists (Fig. 3). Offset pairs of

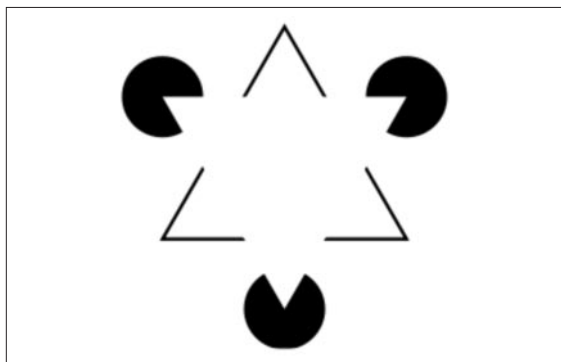


Figure 3 - Kaniza figure in which a white triangle appears to be imposed on black circles. Wikipedia, ILLUSORY CONTOUR [link accessed July 25, 2007]: [http://en.wikipedia.org/wiki/Illusory\\_contour](http://en.wikipedia.org/wiki/Illusory_contour).

subjective contours can even be used in the construction of a three-dimensional structure (Fig. 4). Moreover, physiologists have discovered that neurons that normally respond to a real line (a line actually in the stimulus) respond comparably when conditions are right for perception of a subjective contour. Apparent motion is likewise common, and likewise problematic for the IT/IOA approach. When the experimental set up allows for bistable apparent motion (Fig. 5), the motion seen is either vertical for both pairs, or horizontal for both pairs; there is never a mix of horizontal and vertical. This finding is important because it shows that the direction of apparent motion is coordinated across fairly large areas of the visual field. As with subjective contours, neuro-physiologists have discovered that the same neuron that

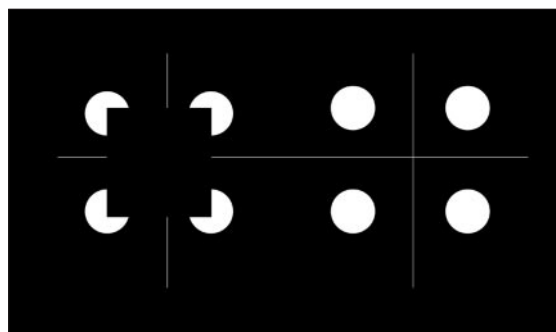


Figure 4 - This illusion shows that illusory motion can occur even when the object apparently in motion is actually an illusory square. The illusory square is created simply by cutting segments out of the white circles. In this display nothing is actually moving smoothly across the image; the representations on the left and right merely replace each other. But what is seen is a square moving smoothly back and forth across the white circles. (Download movie from the website.) Adapted from Ramachandran VS (75). [http://psy.ucsd.edu/chip/illusion\\_motion\\_capture.html](http://psy.ucsd.edu/chip/illusion_motion_capture.html)



Figure 5 - In this illusion, lines appear to move (illusory motion) and the display is constructed in such a way that they can seem to move up and down or back and forth. Although any line can appear to move either way, the surprising fact is that we see all the lines in the entire display moving in the same direction; when the direction for one shifts to the other orientation, they all shift. This implies that the visual scene is not constructed by piecemeal analysis but involves global constraints. Reproduced, with permission, from Ramachandran VS [link accessed July 25, 2007]: [http://psy.ucsd.edu/chip/illu\\_ambig\\_apprnt\\_mot\\_mult\\_stim.html](http://psy.ucsd.edu/chip/illu_ambig_apprnt_mot_mult_stim.html)



responds maximally to real horizontal motion will respond just as well to apparent horizontal motion.

According to the IT/IOA approach, what information is carried in the responses of these neurons when the horizontal motion is apparent but not present in the stimulus? The right IT/IOA answer is not “horizontal motion”, because the neurons respond maximally even when there is no actual horizontal motion. On the other hand, the possible answer “real horizontal motion or what the animal perceives as horizontal motion” is unavailable to the classical IT/IOA theory because the reference to what the animal perceives takes us out of the realm of the objective stimulus (on which their account depends), and into the realm of brain representation.

The gathering realization that normally the human brain exhibits a huge amount of stimulus independent activity (spontaneous activity) presents a different but related challenge (65,66). If any of the neural activity seen during the rest periods (no-assigned-task periods) is representational, if any of that neural activity consists of planning or recollecting or imagining, then no theory of representation that pins its explanation on present stimuli can be adequate.

### Representing Categories

Although it is easiest to see why absent objects are problematic for the classical approach to coding, general categories are also tricky. Consider, first, common general categories, such as *friend*, *kin*, *ripe*, *home*, *territory*, *lost* as well as categories for activities (*eating*, *hiding*, *mating*, *threatening*, *calling*), spatial relations (*under*, *over*, *at home*, *inside*, *outside*), and relative terms (*bigger*, *smaller*, *easier*). An instance of a general category can be a stimulus on a given occasion, but by definition, a category itself is not singular, but general – it is not the stimulus. A category can be true of many individuals, or apply to many individuals, and we may acquire the category via encounters with individuals. But an individual stimulus it is not. Relative terms are especially puzzling for the IT/IOA approach because a small elephant occupies more volume than a big mouse. So how can IT/IOA, defined in terms of singular items in a stimulus set, handle representation by neurons of general categories?

This turns out to be a complex problem because many categories are brain-dependent constructions sensitive to what the organism values. What counts as playing for a young cougar is not playing for a raven; what counts as edible is different for a panda and a raccoon. If a category is acquired by generalizing across similar items, the internal question is what, for a given animal, counts as similar. The logical point here is simple: things are similar *in relation to a concept*; they are not similar absolutely. Rats are similar to termites, *in relation* to the category, “household pests”, but not in relation to the category “furry”. How an animal carves up its world and categorizes stimuli depends quite a lot on what the animal cares about, what it has learned to value or fear or desire, on how the animal makes it living, and on what categories have predictive significance for the animal.

The second point involves the discovery by psychologists that, by and large, our everyday categories (e.g. vegetable, house, friend, face, trap, under) have a prototypical organization (67-69). For example, in the cate-

gory vegetable, carrots and potatoes are prototypical, radishes and mushrooms somewhat less so, and parsley not at all. A dog is a prototypical animal, but a scorpion or maggot is not. An eagle is a prototypical bird, but a penguin is not. Most categories, in short, are not defined in terms of necessary and sufficient conditions for category membership, but have instances that are more or less central, more or less peripheral. The catch here is that category boundaries are typically fuzzy and there is often no determinate answer to the question, “is X really a P?” For example, is a mushroom really a vegetable? Is Mount Soledad really a mountain? Is a lean-to really a dwelling? Is a snow cave? An experimenter armed with IT and seeking to determine what category a neuron represents may encounter few problems with a relatively simple category, such as “table salt”, but deeper difficulties emerge with “kin”, “inside” and “playing”. An even greater difficulty may arise for representation of the abstract goals of another. Rizzolatti and colleagues have shown that certain neurons in F5 of the monkey will respond whether the effector for getting the fruit is the hand, the mouth or the foot (70,71).

One rejoinder to this inventory of challenges to the IT/IOA approach to information and representation is to emphasize the point that the sense of ‘represents’ in these problems is not the same as the sense of ‘represents’ when IT/IOA says that a visual system neuron represents the presence of horizontal motion. That is, information (representation) is ambiguous, and can mean either information in the very narrow but well-defined IT sense or information in a broad but less-well-defined psychological sense. So far, so good. The rejoinder then argues two points: a) The IT/IOA approach does answer the question of how neurons code information, in the sense of *information* explicated by IT/IOA; b) The broader sense of representation can only get off the ground by building upon the IT/IOA foundation. Therefore, the inventory merely constitutes work for the future, not anomalies requiring a new approach.

The first part of the rejoinder just says that IT/IOA works for the cases it works for. That goes nowhere. The second point is what is interesting, and, I argue, probably wrong. The hypothesis that the broader sense of representation will emerge from the IT/IOA foundation may just possibly be correct. That it *is correct*, however, is entirely conjectural, and I have provided reasons to think it is not suitably extendable. Undoubtedly the IT/IOA approach is quite useful at this stage in neuroscience for certain very restricted questions such as the quantity of information processed in a given time. Thus the bandwidth of the human brain is alleged to be roughly 1015 bits per second – approximately equivalent to the existing bandwidth of Google. As a rough estimate, this is revealing and important. The real problem I am targeting is how to reformulate IT/IOA to capture the broader sense of information and representation.

My own suspicion is that for addressing the wider problem, IT/IOA is probably the wrong tool – like using a sieve to carry water. To be sure, neither of the roughly interchangeable pair of terms, *information* and *representation* (in the broader, not the IT/IOA sense) is well defined. We do not yet understand the fundamental principles of the brain-world relationship, especially in complex brains. To put it another way, we do not understand the relationship between the inner model of the world

that the brain constructs, and the world it is a model of (72). Because animals, including humans, can represent things that are not present – when they remember, or plan, or fantasize, or dream, or make mistakes –, the IT/IOA explanations, at least in the current form, will not suffice. So the deep problem is that the terms ‘information’ and ‘representation’, as used in the context of complex nervous systems, are words in search of an experimentally-anchored theory. Their lack of precision might be compared to that of the word “gene”, circa 1920, or “atom”, circa 1800. Lack of precision does not mean that they are worthless or meaningless, but only that they function within the framework of an immature science. Does anyone have *any* idea how representations in the broader sense might connect to the world, if not via the model offered by IT/IOA? Perhaps, a little bit. A long-tendered general idea is that representations (broad sense) connect to the world not singly, one-by-one, but as a semantic network (73). The rough idea is that representational networks fold onto the world in something roughly like the way a map folds onto the world, though representational networks are likely to be n-dimensional whereas maps are two-dimensional (74). The conceptual point is that a representation is what it is, partly owing to its place in the map – its relationships to other things in the representational schema –, and partly as a function of how the map as a whole captures the regularities of the world mapped. That is, a category such as *vegetable* or *under* is what it is partly owing to its semantic and evaluative relationships to other categories, and partly as a function of the way the representational framework as whole reflects properties in the real world. The evaluative part of the story is important here because normally the regularities that are mapped are those relevant to the animal’s thriving and surviving – those the brain is wired and willing to spend energy paying attention to and learning about. These are, to put it a different way, *animal-relevant* categories and regularities. What I am relating here does not really deserve to be called a theory of representation (in the broad sense of *representations*), but it is a significant a conceptual proposal about what directions might, in the fullness of time, prove fruitful. It will be evident that the issues raised here about how neural representations relate to the world cannot be answered by a brilliant experiment or two, though such breakthroughs are always important. What are needed now are empirically informed reflections about the problems, and perhaps even a Copernican shift in thinking about these problems, so that they may be approached in a completely new way.

## Conclusions

Neurophilosophers will undertake many projects and take many different approaches. My own preference has been to play a role in synthesizing across subfields, hoping that by immersing myself in as much of the relevant data as possible, the synthesis will be productive. Traditional philosophers steadfastly adhere to the conviction that problems concerning the nature of consciousness, choice, morality, and knowledge are really conceptual problems. What they typically mean is that reflection on intuitions and on what “we” mean by the expression, is a more fundamental business than science,

and sets the limits for science. While I am always happy to see ambiguities cleared up, the major problem with all these phenomena is not ambiguity but ignorance. We do not know enough. We do not know enough about how the brain works, about what capacities non-human animals have, about the relation between genes and epigenetic conditions and brain development. Nevertheless, many important pieces of the puzzle have been discovered, and these discoveries increase our understanding of the nature of the problems to be solved. Moreover, as we learn more about the phenomena, our meanings and our intuitions change. This is typical, since in science, definitional precision grows in tandem with factual discoveries. Thus “gene” is an expression more precisely describable now than in 1970, and more precisely describable in 1970 than in 1950.

## References

1. Encyclopedia Britannica, HIPPOCRATES, vol. V13, Online Encyclopedia [link accessed July 6, 2007]: [http://encyclopedia.jrank.org/HIG\\_HOR/HIPPOCRATES.html](http://encyclopedia.jrank.org/HIG_HOR/HIPPOCRATES.html)
2. Bechtel W. The challenge of characterizing operations in the mechanisms underlying behavior. *J Exp Anal Behav* 2005;84: 313-325
3. Bechtel W. *Mental mechanisms: Philosophical Perspectives on the Sciences of Cognition and the Brain*. London; Routledge (in press)
4. Craver CF, Bechtel W. Mechanism. In: Sarkar S, Pfeifer J eds *Philosophy of Science: An Encyclopedia*. New York; Routledge 2006:469-478
5. Churchland PS. *Neurophilosophy: Toward a Unified Science of the Mind-brain*. Cambridge, MA; MIT Press 1986
6. Cahan D. *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science*. Berkeley CA; University of California Press 1993
7. Stanford Encyclopedia of Philosophy, WILHELM MAXIMILIAN WUNDT [link accessed July 6, 2007]: <http://plato.stanford.edu/entries/wilhelm-wundt/>
8. Ramón y Cajal S. *Recuerdos de mi Vida*. Cambridge, MA; MIT Press 1937
9. Mazzarello P, Buchtel HA, Badiani A. *The Hidden Structure: A Scientific Biography of Camillo Golgi*. Oxford; Oxford University Press 1999
10. Fritsch G, Hitzig E. Ueber die elektrische Erregbarkeit des Grosshirns [On the electrical excitability of the cerebrum]. *Archiv für Anatomie, Physiologie und wissenschaftliche Medizin* 1870;37:300-332
11. Stanford Encyclopedia of Philosophy, PLATO [link accessed July 6, 2007]: <http://plato.stanford.edu/entries/plato/>
12. Stanford Encyclopedia of Philosophy, DESCARTES’ LIFE AND WORKS [link accessed July 6, 2007]: <http://plato.stanford.edu/entries/descartes-works/>
13. Chalmers D. *The Conscious Mind*. New York; Oxford University Press 1996
14. Stanford Encyclopedia of Philosophy, ARISTOTLE’S PSYCHOLOGY [link accessed July 6, 2007]: <http://plato.stanford.edu/entries/aristotle-psychology/>
15. Hume D. *A Treatise of Human Nature*. London; Penguin Books 1986 (first published 1739)
16. Quine WVO. *Ontological Relativity and Other Essays*. New York; Columbia University Press 1969
17. Crick F. *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York; Scribner 1995

18. Churchland PM. *Scientific Realism and the Plasticity of Mind*. New York; Cambridge University Press 1979
19. Churchland PM. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA; MIT Press 1989
20. Churchland PM. *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge, MA; MIT Press 1995
21. Churchland PM. *Neurophilosophy at Work*. New York; Cambridge University Press 2007
22. Dennett DC. The unimagined preposterousness of zombies: commentary on T. Moody, O. Flanagan and T. Polger. *Journal of Consciousness Studies* 1995;2:322-326
23. Dennett DC. Review of Thomas Nagel's *Other Minds: Critical Essays 1969-1994*. *Journal of Philosophy* 1996;93:425-428
24. Dennett DC. Facing backwards on the problem of consciousness (commentary on Chalmers). *Journal of Consciousness Studies* 1996; 3 (special issue, part 2):4-6
25. Dennett DC. The zombic hunch: extinction of an intuition? In: O'Hear A ed *Philosophy at the New Millennium*. Royal Institute of Philosophy Supplements 2001;48:27-43
26. Churchland PS. *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA; MIT Press 2002
27. Searle JR. *Rationality in Action*. Cambridge, MA; MIT Press 2003
28. Searle JR. *The Rediscovery of Mind*. Cambridge, MA; MIT Press 1992
29. Nagel T. What is it like to be a bat? *The Philosophical Review* 1974;LXXXIII:435-450
30. Jackson F. Epiphenomenal qualia. *Philosophical Quarterly* 1982;32:127-136
31. Jackson F. What Mary didn't know. *Journal of Philosophy* 1986;83:291-295
32. Churchland PM, Churchland PS. *On the contrary: critical essays, 1987-1997*. Cambridge, MA; MIT Press 1998
33. Peciña S, Smith KS, Berridge KC. Hedonic hotspots in the brain. *Neuroscientist* 2006;12:1-12
34. Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology* 2007;191:391-431
35. Baars BJ. *A Cognitive Theory of Consciousness*. New York; Cambridge University Press 1988
36. Dehaene S, Naccache L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 2001;79:1-37. Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition special issue 'The Cognitive Neuroscience of Consciousness'* 2001;79:1-37
37. Llinás R, Ribary U. Consciousness and the brain. The thalamocortical dialogue in health and disease. *Ann N Y Acad Sci* 2001;929:166-175
38. Baars BJ, Gage NM. *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*. San Diego; Academic Press 2007
39. Bickle J. *Psychoneural Reduction: The New Wave*. Cambridge, MA; MIT Press 1998
40. Banks W, Farber I. Consciousness. In: Healy A, Proctor R eds, *Handbook of Psychology*, vol. 4: *Experimental Psychology*. New Jersey; Wiley 2003
41. Bechtel W, Abrahamsen AA. Explanation: a mechanistic alternative. *Studies in History and Philosophy of Biology and Biomedical Sciences* 2005;36:421-441
42. Block N. The harder problem of consciousness. *The Journal of Philosophy* 2002;XCIX:1-35
43. Block N. Review of Pat Churchland's *Brain-wise*. *Science* 2003;301:1328-1329
44. Fodor J. No person is a stream. *Times Literary Supplement*, 16 May 2003:1-2
45. Churchland PS, Sejnowski TJ. *The Computational Brain*. Cambridge, MA; MIT Press 1992
46. Kihlstrom JF. *Social Neuroscience: The Footprints of Phineas Gage*. Keynote address presented at *Neural Systems of Social Behavior*, University of Texas, Austin, May 2007: <http://socrates.berkeley.edu/~kihlstrm/SocialNeuroscience07.htm>
47. Vincent JL, Snyder AZ, Fox MD et al. Coherent spontaneous activity identifies a hippocampal-parietal memory network. *J Neurophysiol* 2006;96:3517-3531
48. Dierick HA, Greenspan RJ. Molecular analysis of flies selected for aggressive behavior. *Nat Genet* 2006;38:1023-1031
49. Sichert AB, Friedel P, van Hemmen JL. Snake's perspective on heat: reconstruction of input using an imperfect detection system. *Phys Rev Lett* 2006;97:068105
50. Heinrich B. *Mind of the Raven*. New York; Harper Collins 1999
51. Craig AD. How do you feel? Interoception: the sense of the physiological condition of the body. *Nat Rev Neurosci* 2002;3:655-666
52. Craig AD. Interoception and emotion. In: Lewis M, Haviland-Jones JM, Barrett LF eds *Handbook of Emotions*, Third Edition (ch. 16). New York; Guilford Publications 2007 (In press)
53. Damasio A. *Descartes' Error*. New York, NY; Grosset/Putnam 1994
54. James W. *The Principles of Psychology* 1890, with introduction by George A. Miller. Boston; Harvard University Press 1983
55. Damasio AR. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos Trans R Soc Lond B Biol Sci* 1996;351:1413-1420
56. Damasio A. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York, NY; Harcourt 1999
57. Watson KK, Jones TK, Allman JM. Dendritic architecture of the von Economo neurons. *Neuroscience* 2006;141:1107-1112
58. Van Hemmen JL, Sejnowski TJ eds *23 problems in systems neuroscience*. New York; Oxford University Press 2003
59. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W. *Spikes: exploring the neural code*. Cambridge, MA; MIT Press 1997
60. Borst A, Theunissen FE. Information theory and neural coding. *Nat Neurosci* 1999;2:947-957
61. Dayan P, Abbott LF. *Theoretical Neuroscience*. MIT Press 2001
62. Thomson EE, WB Kristan. Quantifying stimulus discriminability: A comparison of information theory and Ideal Observer Analysis. *Neural Computation* 2005;17:741-778
63. Hopfield J. Interview with Marcus Christen. *BioWorld* 2002; 2:2-4
64. Rizzolatti G. The mirror neuron system and its function in humans. *Anat Embryol (Berl)* 2005;210:419-421
65. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A* 2005;102:9673-9678
66. Gusnard DA, Raichle ME. Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci* 2001;2:685-694
67. Rosch EH. Natural categories. *Cognit Psychol* 1973;4:328-350

68. Rosch EH. Cognitive representation of semantic categories. *J Exp Psychol* 1975;104:573-605
69. Lakoff G. *Women, Fire and Dangerous Things*. Chicago; University of Chicago Press 1987
70. Umiltà MA, Kohler E, Gallese V, Fogassi L, Keysers C, Rizzolatti G. I know what you are doing. A neurophysiological study. *Neuron* 2001;31:155-165
71. Gallese V, Eagle MN, Migone P. Intentional attunement: mirror neurons and the underpinnings of interpersonal relations. *J Am Psychoanal Assoc* 2007;55:131-176
72. Churchland PS, Churchland PM. Neural worlds and real worlds. *Nat Rev Neurosci* 2002;3:903-907
73. Quine WVO. *Word and Object*. Cambridge, MA; MIT Press 1960
74. Churchland PM. Outer space and inner space: the new epistemology (Presidential Address). *Proceedings of the American Philosophical Association* 2002;76:25-48
75. Ramachandran VS. Capture of stereopsis and apparent motion by illusory contours. *Percept Psychophys* 1986;39:361-373