

OFFPRINTED FROM

The Foundations of
Cognitive Science

Edited by

JOÃO BRANQUINHO

CLARENDON PRESS · OXFORD

2001

4 The View from Here: The Nonsymbolic Structure of Spatial Representation

Ilya Farber, Will Peterman, and Patricia Smith Churchland

1. Introduction

One of the great challenges facing the cognitive and neural sciences right now is to figure out how nervous systems understand and represent space. Besides being intrinsically complex and fascinating, spatial representation is an important entry point to a range of large-scale questions of enduring philosophical interest. For starters, there is clearly a close relationship between the representation of space and the representation of one's own body, and if Antonio Damasio (1994) is correct body representation in turn underpins other forms of self-representation. Related questions involve the nature of coherence and unity in perceptual experience and motor control, and the processes whereby one modality can "trump" another for the sake of coherence. These lead to questions about modularity and how best to characterize such specialization as is seen in brains of mature animals, especially given interactive effects in brain development and in perception (Ballard 1991, Aloimonos 1993), and recent results concerning plasticity in the developing and adult brain (Elman *et al.* 1996; Finlay, Hersman and Darlington 1998). Finally, the study of spatial representation provides a fresh perspective on traditional debates about the nature and structure of mental representation, debates that have tended to focus exclusively on examples from the visual and propositional domains.

In this essay we will apply data from the neurobiology of spatial representation to one of the most enduring questions in cognitive science: is mental representation fundamentally symbolic?¹ Our view is that the psychology and neurobiology of spatial representation, to the degree that they are understood, do not sit well with a symbolic interpretation. In particular, recent results

¹ For the purposes of this essay, not much will turn on the precise definition of 'symbolic'. For clarity, however, let us stipulate that we understand the term to refer to systems which represent the state of the world in terms of the formal relationships among discrete logical tokens, and which represent changes in the world via syntactic manipulation of such tokens. Pure symbol systems, such as Turing machines and the formal languages of logic and mathematics, are those whose logical operations proceed without any reference to—or interference from—whatever physical structure they're implemented in. While PCs and natural languages lack this purity, symbolicists still regard them as *essentially* symbolic systems inasmuch as they can be compactly and profitably described as close approximations to ideal symbolic systems.

from neuroscience—at the cellular as well as the network and systems levels—suggest that the character of spatial representation is intimately dependent on the structure and physical organization of the underlying neural systems.² In contradiction to Steven Pinker’s claim that “information and computation reside in patterns of data and in relations of logic that are *independent of the physical medium* that carries them” (1997: 34; emphasis added), the data surveyed below suggest that knowledge and reasoning about space seem highly *dependent* on the physical organization of the medium that carries them. If this is true, it presents proponents of the symbolic approach with some painful choices, which we will sharpen in our concluding discussion.

2. Spatial Problem-Solving

Consider a contrast between what seems simply “given” and obvious, and what seems to require inferences or computation or “thought”:

1. In a visual scene, one “directly” sees that the hammer is *on* the table, that the cat is *under* the bench, or that the flower is *to the right of* the tomato. Shifting one’s gaze from the flower to the tomato is trivial; moving one’s hand to grasp the flower or the tomato is trivial.
2. Children playing “hide and seek” in a park have to gauge which, amongst the available objects, will best screen their bodies from the view of the seeker. This is a nontrivial task, which will require time spent looking back and forth to judge lines of sight between the seeker and the various objects.

Philosophers schooled in a language-of-thought approach will tend to assume that only the second task involves mental representations and thought processes in any real or important sense, and that the processes enabling the two achievements are fundamentally different. In particular, it may be assumed that the second alone involves reasoning, reasoning which runs something like this: “I want to hide from Fred. He is going to see me unless I hide behind something. That bush is too short, that one is too narrow, that one is too sparse and transparent, so I will crouch behind this one.”

We suggest that, appearances notwithstanding, the two kinds of operations are fundamentally related, and that understanding how the brain solves the “how should I hide myself?” problem will depend on understanding how the brain solves the problem of getting a representation of object-centered space. Both achievements depend on the brain’s handling of spatial relations, undoubtedly a fundamental feature of brain organization in all animals. As

² Our essay thus connects to a tradition in psychology (Farah 1990; Stiles and Thal 1993; Kosslyn 1994) and linguistics (Elman and Zipser 1988; Kuhl 1991; Lakoff 1987) that seeks to explain at least some mental functions in terms of more embodiment-sensitive concepts such as mappings and dynamics.

we will show, the available evidence implies that the relevant representational systems are not language-like, and cannot readily be accounted for under a “symbol manipulation according to algorithms” paradigm.

Many animals routinely solve the hiding problem. Shirley Strum reports (in conversation) an example involving a female baboon who spies an out-troop male, potentially dangerous. When baboons are frightened, their tails reflexively rise up. In the example Strum reports, the female, partially shielded by a bush, reaches around to her back and pulls her tail back down out of the male’s view.

Clark’s nutcrackers hide nuts and, in later retrieval forays over the winter, remember which locations have been emptied and which still contain nuts. Grizzly bears have been observed hiding by ensuring that their considerable girth is well out of view of a human passing along a trail, though remaining quite visible to other creatures of lesser concern. This is not just the reflexive dive-for-a-dark-place behavior typical of a startled cockroach. It requires spatial understanding and knowledge of the relation of one’s own body to other bodies in space.

Spectacularly, ravens can solve, sometimes in *one trial*, the problem of how to get a piece of meat hanging from twine tied to a trapeze (Heinrich 1993). The twine is set to a length that requires about seven iterations of this routine: “pull up a length with the beak, step on it, pull up another length with the beak, step on it, . . .” This task does not resemble any commonly performed by ravens in the wild, though ravens do step on a carcass to hold it down while ripping off chunks of meat. Crows also rip meat off a carcass in this way, but unlike ravens, they *cannot* solve the trapeze problem. The trapeze task requires the animal to perform a sequence of maneuvers, the first six of which have no immediate reward or sign of success except the increased proximity of the meat. This implies that the raven understands quite a lot about the spatial nature of the problem, and how its body needs to interact with the set-up to produce the right sort of change in spatial configuration. The point is, success in solving this problem is not achieved by trial and error, and it is not achieved by chaining in conditioning (in the way that pigeons can be trained, step by operant step, to perform a complex act). Success at the trapeze task relies on spatial representations.

One further example illustrates quite directly that spatial representation plays a role in food finding. Consider a rat put in a T maze where the left arm is always baited. The rat goes in at the bottom of the T, turns left, and gets the food (see Figure 4.1). Suppose now that a barrier is removed, converting the T maze to a cross maze, with the rat entering via the top arm. To get to the bait from this new starting point, the rat must turn right. If the rat has a genuinely spatial representation and understanding of its position in the maze, it will adjust for its new entrance path by turning right; if it has merely been conditioned to a “left turn” response, it will turn left as in the training condition. *Rats turn right*. Overtraining or hippocampal lesions can prevent

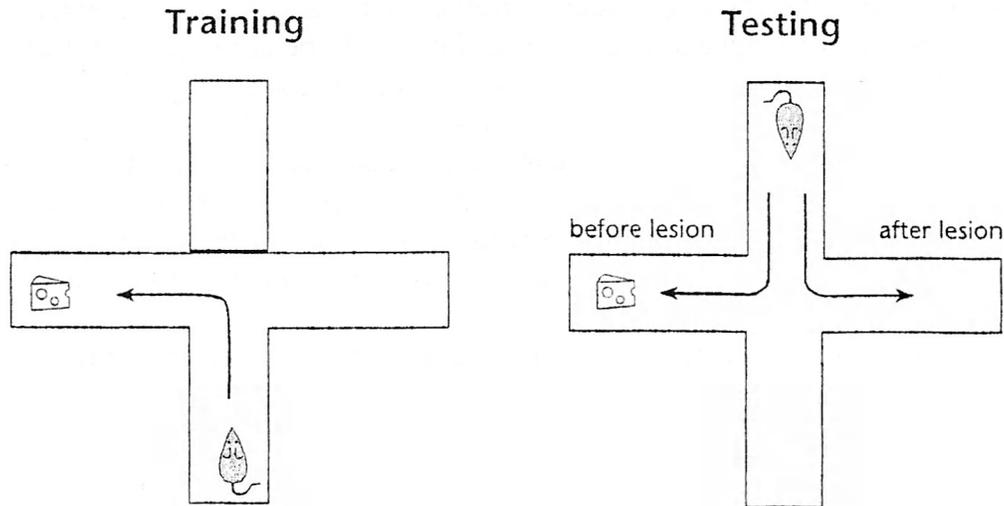


Figure 4.1 Spatial reasoning in the rat. In the training condition (left), the rat always starts in the same spot and learns to find food in another fixed spot. In the test condition (right), a block is removed and the rat is placed in the newly opened passage. Normally trained rats will turn right, correcting for their reversed spatial orientation with respect to the training location. Overtrained rats, or rats with hippocampal lesions, will turn left.

this compensation, causing the rats to turn left and miss the bait. (Packard and Teather 1998*a*, 1998*b*).

It is fairly safe to assume that the mechanisms baboons, bears, rats, ravens, and nutcrackers deploy in solving these tasks do not much resemble language-like propositional reasoning. Those who favor linguistic/symbolic models of human reasoning are thus left with a dilemma: either they must claim that humans and animals solve the same spatial problems in fundamentally different ways, or they must surrender spatial reasoning to the realm of the nonsymbolic and then search for some way to distinguish it in principle from other representational tasks of equivalent complexity. Neither path seems particularly promising.

For those (Fodor 1975, 1995; Pylyshyn 1984; Pinker 1997) who favor the “language of thought” hypothesis, symbols are valued because only symbols can be governed by formal rules, such as *modus ponens*. *Modus ponens* and its algorithmic kin don’t work without formal symbols, and without such formal rules, the argument goes, we cannot explain problem solving, reasoning, and belief-desire structures generally. *With* formal rules and symbols, we can explain such profound things as the tendency of someone who believes “if P then Q” and “P” to also believe “Q”.

We have no objection to the claim that trained humans occasionally perform *modus ponens*; but the very motivation for the symbolic project requires its proponents to claim much more. Two basic tenets of computationalism are that the fundamental architecture of thought is symbolic, and that

normativity only makes sense in the context of symbolic representation (in other words, that you can't be *right* or *wrong* about something unless you're using symbols). Together, these principles gave rise to the standard methodology of "good old-fashioned" cognitive science: *start* with the basic resources of a symbolic computational system and then figure out how the system could be rigged up to exhibit the powers and properties exhibited by humans.

We find it more rewarding to look at the various problems involved in explaining mental processes from the other end. First, certain behavioral data imply that at least some forms of thinking and problem solving *are not* symbolic. Second, data from neuroscience and neural modeling reveal non-symbolic computational strategies whereby nervous systems can and do solve spatial problems such as reaching to a target, finding hidden food sources, and so on. Together, these findings provide a strong motivation for investigating the actual structure and dynamics of real neural systems—for *learning* from nature, rather than imposing our prior theoretical convictions on it.

Although considerable progress has been made in the last three or so decades, we still do not have anything like a comprehensive theory of the mind/brain, and there remain many unexplained functions: event storage and retrieval, attentional processes, cross-modal integration, pattern recognition, accurate prediction of events and their durations, flexibility in representation, the use of uncertain and incomplete knowledge, language learning, analogical inference, conscious awareness, skill learning, navigation, and so on *and on*. Nevertheless, such progress as *has* been made on these issues does not, in general, derive from the symbols-and-formal-rules paradigm. Hence Pinker's claim that "human thought and behavior, no matter how subtle and flexible, *could* be the product of a very complicated program" (1997: 37; emphasis added), however rousing a bit of boosterism for the symbolic approach, seems largely irrelevant to the project of understanding the way human thought and behavior *actually* work.

3. The Neural Basis of Spatial Representation

In primitive animals such as the leech, spatial representations are essentially limited to location on the body surface. Receptors in the skin project to segmental ganglia which also house motor neurons that enable the leech to bend away from noxious stimuli. Anterior segments in the leech can also detect specific chemicals, and the leech can swim up a chemical gradient to a food source or down a chemical gradient to get away from something nasty. For these humble if nontrivial accomplishments, it need not have a representation of where, in external space and/or relative to itself, the targets are. In these sorts of animals, the spatial world and its body world are pretty much one. For the leech, "my body hurts here" is equivalent to "the world hurts here". As Damasio might put it, for such organisms, there is *only* subjectivity.

In fancier animals, evolution has stumbled on the strategy of exploiting a range of sensory signals from distinct modalities to get ever more accurate information, which can in turn support ever more sophisticated movement. Eyes, ears and vestibular systems, wings, and movable heads allowed animals to intercept moving prey detected at a distance, and sometimes even in the dark. Hearing and seeing are especially helpful in this regard, and the more so when auditory and visual signals *share constancies* such as intensity, source, and duration. But other signal-processing devices evolved too, such as electroreception (in fish) and infrared detection (in snakes), both of which register events at a distance. When there are independently movable parts, including movable signal detectors such as eyes, pinnae, whiskers, and antennae as well as movable heads and limbs, the brain has available a very rich story of “my body in the world”. With mammals, and vertebrates in general, we see the emergence of a more complex, richer representation of the subject in an *objective* space—a space of things “out there”.

What do we know about how neurons (in vertebrates) represent space? A vast literature is relevant here, but to keep within the limits of this essay we will restrict our focus to mammals. There are three areas in which substantial progress has been made in understanding how the mammalian brain represents objects in space. The first involves the hippocampus and has its roots in John O’Keefe’s discovery in the 1970s of “place cells” in the hippocampus of the rat.³ This line of research explores the role of cells in the hippocampus in representing and navigating through space. It turns out that a given place cell may code for a different region of extra-body space in different environments, and that the “maps” are not topographical. The findings so far suggest that these spatial coordinates are allocentric rather than egocentric (in other words, based on an external frame of reference rather than on a self-centered one). Some recent results indicate a relation between spatial learning, place cell responsivity, and rehearsal of place-finding in dreaming (Wilson and McNaughten 1994).

The second pertains to discoveries by Fuster and by Goldman-Rakic of cells in the prefrontal cortex of the monkey that hold spatial information in working memory. Other research (Fogassi *et al.* 1992, Colby and Duhamel 1993, Graziano and Gross 1993) revealed multi-modal cells in prefrontal cortex—visuo-somatosensory cells—which encode the position of objects in *head-centered* coordinates (see below). The third research line concerns structures in the posterior parietal cortex of the monkey (areas 5 and 7) that provide the substrate for transforming signals from retinotopic coordinates to body-centered or eye-centered or object-centered coordinates. (See Figure 4.2 for the locations of these structures in the macaque monkey.)

The three brain areas of particular interest here (hippocampus, prefrontal, and posterior parietal) are also highly interconnected, so it may not be foolish

³ For a more recent account, see Wilson and McNaughten 1993.

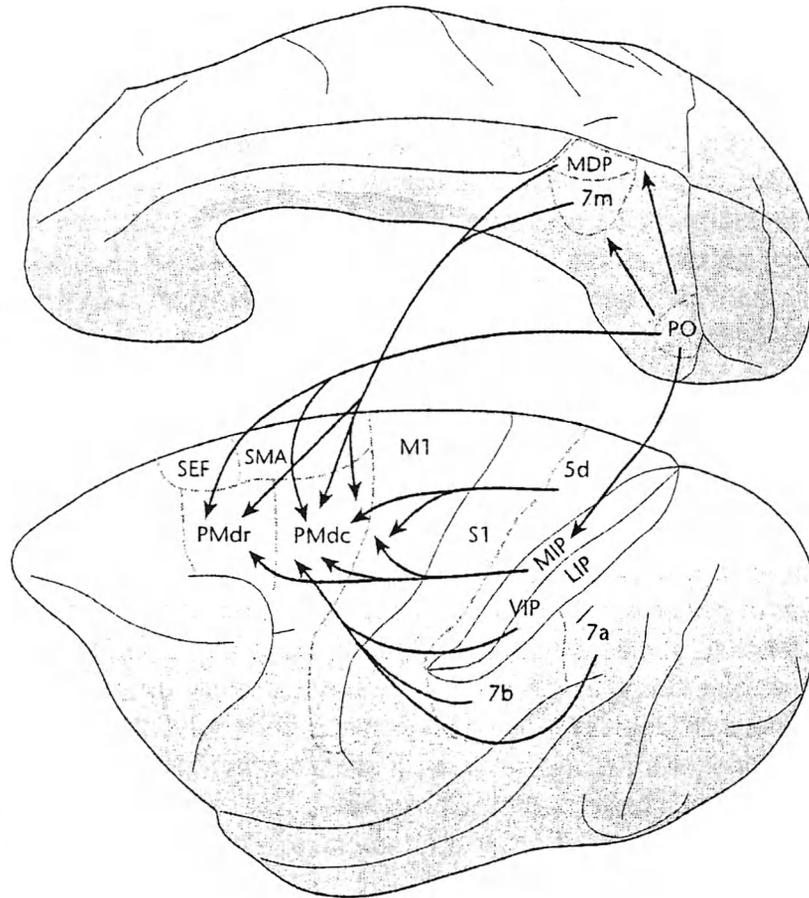


Figure 4.2 Connections to and from posterior parietal cortex, shown in medial (top) and lateral (bottom) views of the left hemisphere of a macaque. Arrows are shown as unidirectional, but most corticocortical projections are reciprocal. Areas included in the posterior parietal cortex (discussed in Section 3, below) are 5a, 7a, 7b and 7m, the medial dorsal parietal area (MDP), and the medial, lateral, and ventral intraparietal areas (MIP, LIP, VIP). Other abbreviations: PO, parieto-occipital visual area; S1, somatosensory cortex; M1, primary motor cortex; PMd, dorsal premotor cortex; SMA, supplementary motor area; SEF, supplementary eye field. From Wise *et al.* (1997), used by permission of the authors.

to hope for a consilient, interlocking theory in the long run.⁴ Research at the cellular level on prefrontal and posterior parietal has been done mainly on the macaque monkey, though human neuropsychological data suggest it is reasonable to expect a roughly comparable story for humans.

Our focus will be on the posterior parietal cortex. This area appears to provide the fundamental “objects out there external to my body” organization critical for primate sensorimotor representation and control. The

⁴ Nor indeed are these the only areas that play a role in spatial knowledge. To do justice to the matter, one should also discuss the superior colliculus (Groh and Sparks 1996), ventral stream (Milner and Goodale 1995), and probably the cerebellum and the basal ganglia (see esp. Jeannerod 1988; and Milner and Goodale 1995).

hippocampus and prefrontal areas likely use these basic parietal representations for additional purposes (e.g. remembering the locations of goodies, planning movements, generating images of movements, etc.).

Understanding of where things are in three-dimensional space does not just arise magically, of course, nor is it just *given*, whatever that might mean. It depends crucially on the structural organization of various receptor sheets, and on how sensory signals are integrated and represented. A range of results from basic neurobiological research, behavioral research, and neural modeling come together in a rather compelling idea developed by Pouget and Sejnowski (1997a, 1997b, 1997c). The Pouget and Sejnowski hypothesis grounds an approach to explaining how the primate brain integrates diverse sensory signals and generates an “objective” representation, that is, a representation of where things are in the space relative to one’s independently movable body parts.

The crux of the idea developed by Pouget and Sejnowski is that posterior parietal generates *basis functions* which can then be “pushed through” assorted filters to yield “go-to” locations in the corresponding motor reference frames. The basis functions in question are most simply thought of as representations of “my-body-my-view” space. They will be partly responsible for our ability to derive spatial knowledge from perception, as well as our ability to act in ways that incorporate spatial information of which we are not directly aware (as seen most vividly in deficit cases where patients may be able to point to objects that they aren’t aware of seeing or catch a ball without being aware of any visual motion).

These representational structures can be deployed by different motor structures for distinct motor chores, such as moving the eyes, hands, pinnæ, legs, or head. The representations are not exactly or merely perceptual, not exactly motor, nor exactly egocentric or allocentric. They combine information from multiple sources in a way that is suited to multiple applications, but cannot neatly be described in terms of a single map or reference frame. Incidentally, this is one of many examples where the representational business of a neuronal pool does not correspond to any folk psychological representation.

Here, as in other places in computational neurobiology, the mathematical workhorse is the parameter space, with mappings across the dimensions of the parameter space yielding coordinate transformations, and volumes or trajectories in parameter space serving as representations. To show why the Pouget and Sejnowski hypothesis is promising for the problem at hand, we must first introduce relevant data concerning parietal cortex, and then explain what is computationally advantageous as well as neurally apropos about their hypothesis.

In the early stages of the visual system (e.g. V1, V2), the location of the visual signal is specified in retinal coordinates. To move the eyes and head to look at a heard or felt object, or to reach with an arm or tongue or foot for a

seen object, the brain needs to know where to go in the appropriate coordinate system; retinal coordinates will not suffice. It needs to know, *inter alia*, where the eyeball is with respect to the head, where the head is with respect to the shoulder and trunk. Coordinate transformations are needed to specify where the eyeball should go to foveate, or what, in *joint* coordinates, the arm's position *is* and what it *should be* so that it makes contact with the target (see Figure 4.3). More generally, sensory coordinates have to be transformed into motor coordinates in order to connect with a sensorily specified target.

Integration of somatosensory “body-knowledge”, including proprioceptive and vestibular knowledge such as “where this body part is in relation to other body parts”, with visual-auditory “where things are in relation to my body” knowledge, allows for general representations of “me in external space”. As Damasio has argued, however, the spatial aspects of body-representation are

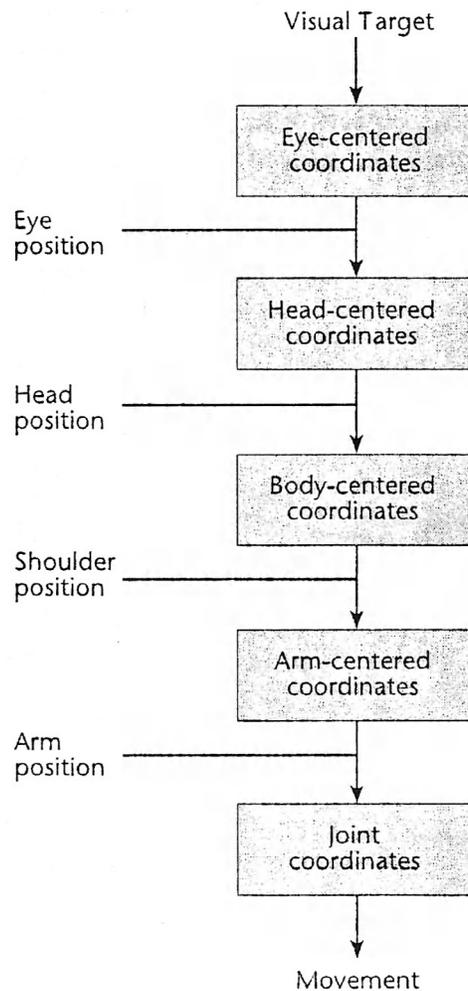


Figure 4.3 Coordinate transformations required to specify an arm movement toward a visual target. From Pouget and Sejnowski (1997c), used by permission of the authors.

only one *part* of the self-representation story, since other aspects involving various dimensions of feeling and homeostasis will figure in what it is to have a “me” representation.

Neurobiological studies of areas 7a and 7b in parietal cortex have provided important clues as to how coordinate transformations are accomplished. What follows is a simplified description of what is known, but it should suffice to convey the main points. As background, note that the visual system in monkeys (and probably primates generally) has a number of important early (pre-cortical) components, including major projections to the superior colliculus and the pulvinar of the thalamus (see Figure 4.4). The pathway we consider here goes from the lateral geniculate nucleus (LGN) of the thalamus to cortical area V1. Thereafter, there is a broad divergence into two broad pathways commonly referred to as the dorsal and ventral streams, with some significant cross-talk between the two. There appears to be divergent functional specialization, the details of which are currently under investigation. To a first approximation, the ventral stream seems to be specialized for categorization, shape, and color, while the dorsal stream shows greater sensitivity to tasks such as stereopsis, motion perception, and spatial location. Area 7 is located in the more anterior region of the dorsal stream. The ventral stream

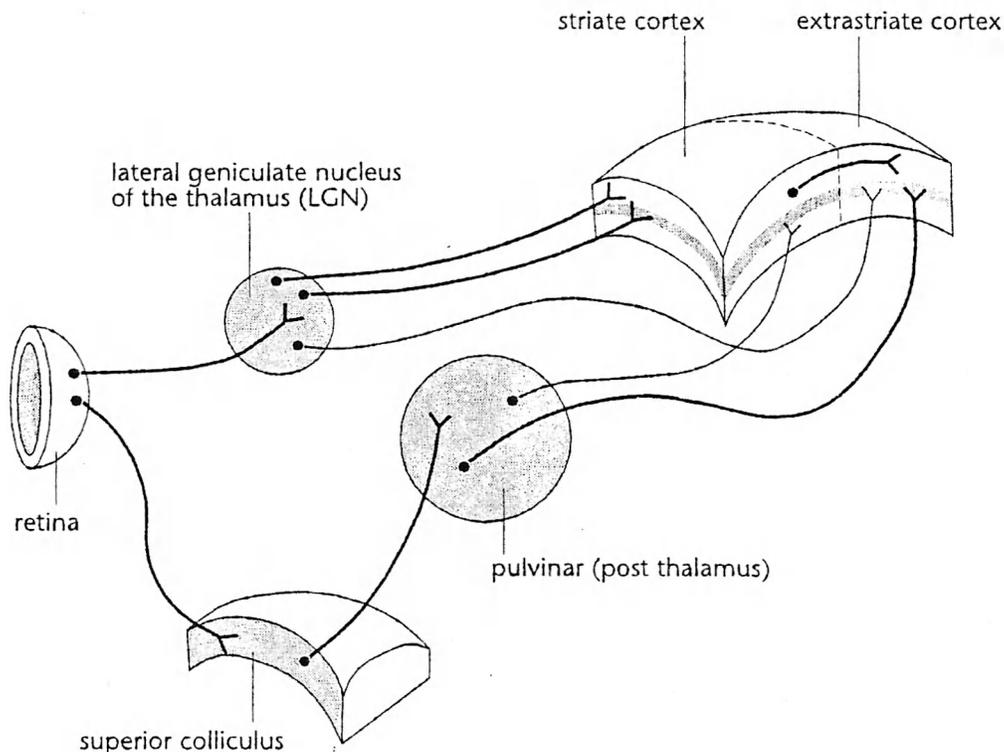


Figure 4.4 Retinocollicular and retinogeniculate visual pathways in mammals. Heavy and thin lines represent major and minor projections. Modified from Butler and Hodos (1996).

appears to play an important role in the perception of visual space as well (Milner and Goodale 1995), and this should be borne in mind despite our focus here on the dorsal stream.

Monkeys with bilateral lesions in area 7 show poor reaching to a target (ataxia), mis-shaping of hands to fit the shape of the target, and slowness of movement. They also show defective eye movements, principally in foveating, and they have other impaired spatial abilities. They are poor at finding the home cage when released, poor at route finding to a food source, and poor in judging spatial relationships among objects (e.g. “the food source is the box located nearer to the can”). Another region of parietal cortex is area 5, which contains cells that fire maximally to a signal when the arm is reaching, and others that fire selectively to the expectation of a stimulus. Because a great deal of research has probed the visual properties of this area, one tends to think of these regions as essentially visual. Recent data reveal, however, that they are much more than that. Response modifiability by many factors including auditory, somatosensory, and vestibular signals, attention, intention, expectation, preparation, and execution, clearly indicates these neurons are more than just sensory. (See Andersen, Essick, and Siegel 1985; Andersen *et al.* 1990; Andersen 1995; Mazzone and Andersen 1995; Wise *et al.* 1997.)

Area 7 is multimodal, and contains cells responsive individually to either visual, auditory, somatosensory, chemical, vestibular, or proprioceptive signals. Interestingly, auditory cells in this region appear to be mapped in retinotopic coordinates. A few cells are themselves multimodal: they may respond to both visual and auditory signals, or to somatosensory and visual signals, or to chemical and somatosensory signals (Wise *et al.* 1997).

Richard Andersen and his colleagues (1985) discovered that certain ostensibly visual cells show modified responsivity as a function of eyeball position. That is, if a given cell’s receptive field was, say, at $+15^\circ$, -14° , then its firing to a signal at that retinal position would be modified depending on where the eyeball was in the head. Crudely, this means that such cells “know” where the external object is *in head-centered space*. The discovery is important because it means that some cells in area 7 are the beneficiaries of a transformation that uses retinal coordinates and eyeball position coordinates to yield information about where external targets are in head coordinates. The representation is thus *somewhat* perceptual, *somewhat* “perceived-thing-relative-to-my-body”. Such computational results can then be used by the brain for foveation or reaching. What computation enables coordinate transformations from retinal position and eyeball position to position in head space or body space? How, roughly speaking, can you get *objective* space out of *subjective* spatial relations?

In a classic example of how neural modeling can yield useful ideas for neuroscience, Zipser and Andersen (1988) asked whether an artificial neural net could perform the task of finding the head-centered position of visual objects from their retinal position and the position of the eyeball. Using those

two types of position information as input, they trained the net by back-propagation to specify positions in head-centered coordinates. It was a major finding that the two kind of inputs were indeed sufficient for a network to generate a head-centric representation. Exactly what function was computed by the hidden units, however, was not yet clear. Pouget and Sejnowski then investigated the hidden units in the context of a wide range of neurobiological data, including single-cell data. What emerged was a plausible but rather surprising answer to the question of how neurons accomplish the task.

Except for organisms whose sensory and motor systems are *very* simple, the transformation from sensory to motor coordinates is nonlinear. This includes transformations from retinal coordinates to eye position coordinates and joint coordinates and tongue coordinates. As Pouget and Sejnowski note, the retinal receptive fields are Gaussian and the brain does not have access to the retinal position R as a series of numbers—the horizontal and vertical components—but to a set of nonlinear functions of R . Girosi, Jones, and Poggio (1995) had explored the idea that basis functions are an efficient way for the nervous system to approximate nonlinear functions. Gaussians and sigmoids are a subclass of basis functions, as is the product of a Gaussian and a sigmoid.

The Pouget and Sejnowski story goes as follows: eye position units are sigmoidal, and retinal position units are Gaussian. Hidden units (interneurons in area 7) compute the product of the two, which is a basis function (see Figure 4.5). This example involves only two dimensions, but the idea can—and, for neurobiological reasons, *should*—accommodate dimensions for somatosensory, proprioceptive, vestibular, and other signals. Thus, each hidden unit provides a set of basis functions. Computationally this is convenient, because the same basis functions can then be used by different regions to compute the movement appropriate for the appropriate reference frame—reaching or saccadic eye movement or head-turning to an auditory or visual stimulus or whatever. In short, on the Pouget and Sejnowski hypothesis, the area 7 cells studied by Andersen and colleagues are disposed to represent “where perceived objects are in my body space”. (See also Andersen *et al.* 1997.)

This result is surprising because it means that rather than actually generating representations for many different reference frames, the brain may use the more down-scale method of generating one set of basis functions and manipulating them as needed. When pushed through relatively simple distinct filters, they provide information suitable to the motor pathway selected (eye muscles, head muscles, trunk muscles, etc.). Depending on the sensorimotor goals, emphasis may be given to cells that are driven primarily by particular sorts of inputs—so a system involved in directed reaching might read off from area 7 a “world” of somatosensory-to-visual coordinations, while a system related to postural adjustments might read off a very different “world” defined by somatosensory, vestibular, and proprioceptive axes (see Figure 4.6).

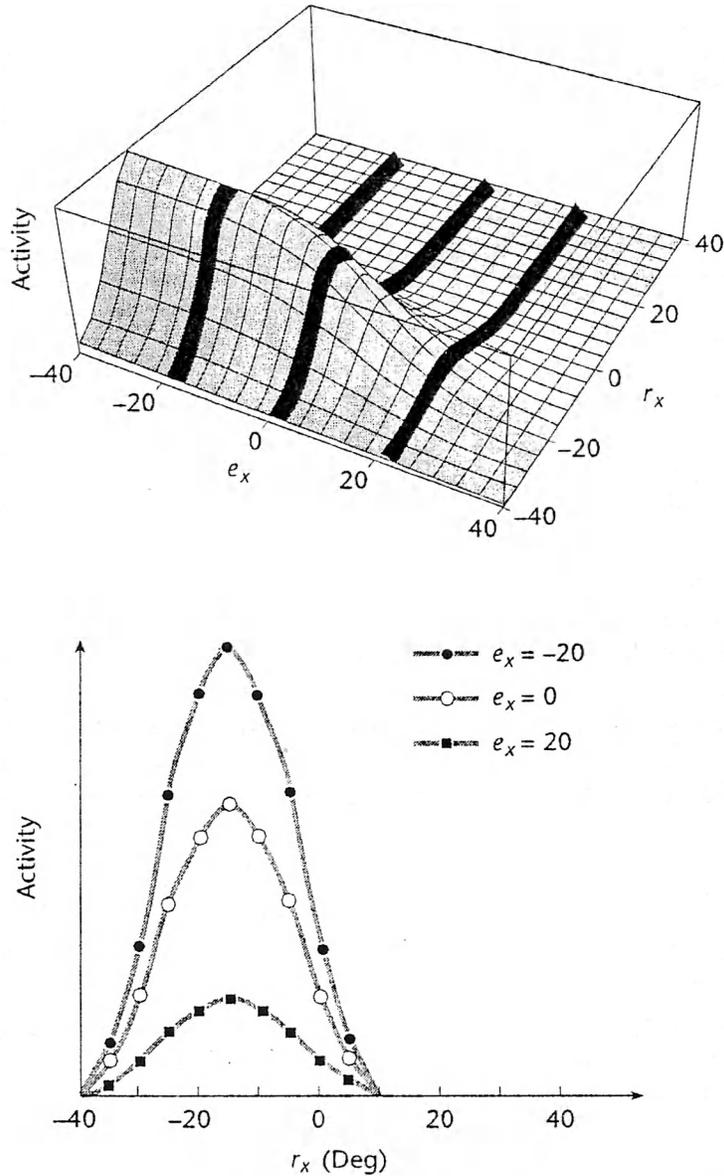


Figure 4.5 Basis function obtained by multiplying a Gaussian of retinal location with a sigmoid of eye position (*top*). When sampled at three different gaze angles (thick lines on top graph), the visual receptive field (*bottom*) shows the same modulation as found in parietal cortex. From Pouget and Sejnowski (1997c), used by permission of the authors.

What introspection presents as the “oneness” of spatial perception is undoubtedly illusory to some degree. Various versions of “where perceived objects are in my body space” can dissociate (largely without introspective notice) as a function of precisely which perceptual modalities are involved. The effect has been demonstrated in a variety of experiments. Some examples include: (1) the ventriloquist phenomenon, whereby sound is perceived as coming from an object that merely moves in synchrony with it; (2) changes in

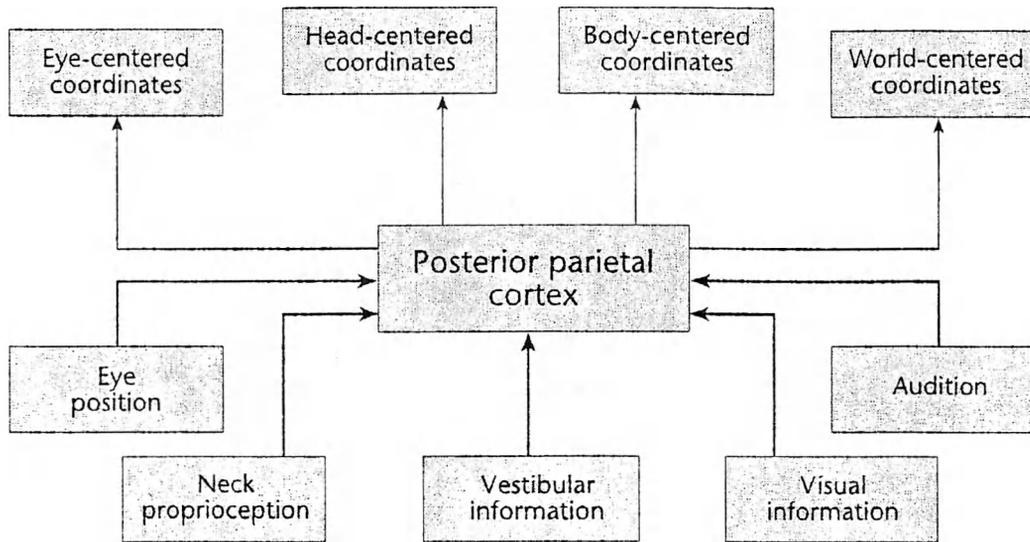


Figure 4.6 The role of posterior parietal cortex in the transformation of retinotopic visual information into higher-order reference frames. Eye position, head position (determined from neck proprioception and vestibular sources) and gaze position (determined from visual sources) are used to modify retinotopic signals. Posterior parietal cortex is thus positioned to provide an intermediate stage in the conversion of visual and auditory information into eye-, head-, body-, and world-centered coordinate frames. From Andersen *et al.* (1997), used by permission of the authors.

visual perception brought about by vibrating the neck muscles (thus stimulating the vestibulum); (3) Ramachandran's "rubber hand" experiment in which the perceived location of a somatosensory stimulus is outside of the actual body;⁵ and (4) Stevens's production of illusory visual motion by paralyzing the eye muscles.⁶ Additionally, *within* vision, in normal subjects, there can also be dissociation between spatial coding as it is visually

⁵ In this set-up the right hand is shielded from view, and a rubber right hand is placed in view on a table, about where one might place one's right hand. The subject watches while the experimenter makes many touching, stroking, and tapping motions on the rubber hand. Out of sight, identical motions are made on the subject's real hand. In my (PSC) case, after about three minutes, I ceased to feel the touches as located in my right hand; they felt like they were really located in the rubber hand. The sensation was very robust. The other subject on that day, David Hubel, had a similar experience. Nevertheless it should be noted that not all subjects make the transfer so readily. Although my experience was not the result of a well-controlled experiment, Ramachandran has obtained similar results with subjects under controlled conditions.

⁶ See Matin, Stevens, and Picoult 1983. In this experiment, run on himself and his colleagues, Stevens immobilized the eye muscles via a procedure known as a retrobulbar block. When a light is flashed in the visual periphery (e.g. to the right), one intends to move the eyes to the right to look at it. Because the extraocular muscles are paralyzed, however, no eye movement can happen. This mismatch between intent and performance produces the vivid visual *experience* of the whole world making an abrupt jump to the right (as though the eyes really had moved and—since the world still looks the same—the world must have moved along with them). In fact, of course, nothing moved, neither eyes nor world. One simply *intended* the eyes to move. This is a wonderful example of feedback from the motor command trumping visual motion.

experienced and as it is used for grasping (Milner and Goodale 1995).⁷ As is well known, such dissociations also occur in patients with parietal lesions (see the discussion of hemineglect, below).

To the extent that the experienced “oneness of space” is not illusory, it is highly dependent on the fact that all of the signals are generated in *one* nervous system, inside *one* body that is *one* spatially linked source of signals. There is no single objective spatial representation (of the sort standardly presupposed by symbolic models of representation), but a distributed, multi-modal representation that fundamentally integrates perception and action, self and world. Where constancies appear across distinct modalities, it becomes possible and even inevitable to understand them as representing an enduring world beyond the body. Kant’s “transcendental unity of apperception” is turning out to be the fundamental integration of diverse sensory and motor coordinate systems, exquisitely configured *physically* to represent the world.⁸

4. Deficits of Spatial Reasoning

The previous sections examined, from the behavioral and neurological perspectives, the remarkable powers of spatial representation and reasoning demonstrated by humans and animals alike. To complete our survey of the empirical data, we turn now to an account of some *defects* in spatial reasoning found in humans with parietal lobe damage. It has recently become apparent that nonsymbolic models such as those discussed above can predict the fine structure of these deficits to a remarkable degree.

One intriguing pattern of breakdown in spatial reasoning occurs in *hemineglect*, a condition found in patients with unilateral lesions of right parietal cortex. These patients display a marked tendency to ignore the contralesional (i.e. left) side of their body-centered world. They tend to look only to the right, though some can move their eyes to the left if directly asked to do so. Asked to make a drawing or reproduce a figure, they will omit most or all of the left half; asked to “cancel” (cross out) all the lines on a page, they will fail to cancel all the lines on the left half of the page; asked to bisect a horizontal line, they will draw the transecting line off-center to the right.

In one ingenious experiment (Bisiach and Luzzatti 1978), hemineglect patients were asked to imagine they were standing in a well-known plaza in their city, and to describe what they could see from a given vantage point.

⁷ They use a visual illusion in which a disc visually appears to be larger than it is, but the grasp aperture in *reaching* for the disc is set to the correct size.

⁸ On the positive side, Kant’s characterization of space as a “form of intuition” might turn out to have a discernible neurophysiological basis, investigation of which could yield new insights into the fundamental structure of human perception. Such considerations also lend weight to Searle’s claims about the inadequacy of the functionalist paradigm.

Their descriptions omitted objects on the neglected side *relative to the imagined vantage point*: told to imagine they were standing at the north end of the plaza, they would omit the buildings to the east, and when subsequently instructed to repeat the task from the southern vantage point they would list the eastern buildings and omit from their description all the western buildings that they had listed just moments before. This shows that hemineglect is a deficit of spatial reasoning and/or representation at some fairly basic level, and not just a perceptual failure. Corroborating evidence comes from studies in which a neglect patient perceived rotated non-symmetric figures as normalized (that is, rotated back to their canonical positions) and neglected the *normalized* left half, something which could only happen if the loss of information occurred at a later stage than some fairly high-level aspects of visual processing.

There is also a motor component to hemineglect. Neglect patients show little or no spontaneous use of limbs on the left side of the body, though some will reluctantly move the neglected limbs upon direct request. They will also neglect auditory stimuli from the left, sometimes failing to acknowledge others who are speaking to them from that side. This sort of polymodal, perceptuomotor deficit is what one might expect from a lesion of parietal cortex, since it receives and integrates inputs from multiple modalities and is known to be involved in the coordination of perception with action.

Pouget and Sejnowski (1997*a*, 1997*b*, 1997*c*) used hemineglect as a test of the predictive power of their basis function hypothesis (described in Section 3 above). They created a network model which has the oculomotor input/output structure and response properties that they attribute to area 7a, and then “lesioned” it by removing the units that correspond to the right side of the brain.⁹ This left the network with a disproportionately high number of neurons that were most responsive to rightward eye positions and/or right visual field stimuli. They then equipped it with a winner-take-all output selection mechanism, and tested it on stimuli similar to those used with neglect patients.

The network output exhibited striking similarities to the human behavioral results. In the line cancellation task, the network failed to cancel lines on the side opposite the lesion. More importantly, the line between the cancelled and non-cancelled areas was fairly sharp, even though the underlying representation had only a smooth gradient. The network also paralleled human behavior on the bisection task: it was successful before the lesion, and shifted the

⁹ The units in each hemisphere were organized into maps, with one axis representing sensitivity to horizontal retinal field position (vertical position was not considered) and the other representing sensitivity to eye position. The maps were constructed to have neuronal gradients, such that the right hemisphere had more neurons which were responsive to left retinal field and eye positions, and vice versa for the left hemisphere. Parietal cortex is known to have these sorts of gradients for retinal position; eye position gradients are observed in other areas, but it is not known whether they exist in parietal cortex.

transection point to the right after the lesion. In other experiments, it was shown to suffer object-centered as well as visual-field-centered neglect, and a version that received head position information (instead of eye position) exhibited a curious effect found in human patients, whereby performance on left-field tasks can be improved by turning the head to the right. Both of these phenomena have been problematic for existing theories of hemineglect, and Pouget and Sejnowski are able to explain how they arise naturally from an organization of response functions that can plausibly be attributed to human parietal cortex.¹⁰

While hemineglect is associated with unilateral damage to parietal cortex, bilateral damage can produce another, more rare condition known as Balint's syndrome. Balint's syndrome patients have little or no ability to locate, count, or track objects, and have poor or nonexistent depth and motion perception. In addition, they exhibit *simultagnosia*, an inability to perform perceptual tasks involving more than one object present in the visual field at the same time.

There is no established explanation of the cluster of deficits seen in Balint's syndrome. One line of theory holds that parietal cortex is centrally involved in the task of disengaging attention from one object so that it can be directed to another, and that patients with parietal damage are thus unable to shift attention between multiple objects (Posner *et al.* 1984; Farah 1990).

More recently, Lynn Robertson, Anne Treisman, and their colleagues have proposed that the central deficit in Balint's has to do with feature binding (Robertson *et al.* 1997). Parietal cortex is the terminus of the dorsal pathway or "*where system*", which extracts information about the spatial position, relations, and movement of visual objects. Based on Treisman's "Feature Integration Theory" (Treisman and Gelade 1980)¹¹ and on the group's study of a Balint's patient, Robertson and colleagues claim that parietal cortex establishes a spatial map which is required for the proper grouping of perceptual features (as detected by the ventral pathway or "*what system*") into separate, coherent objects.

Robertson and colleagues make a strong experimental case for the claim that Balint's patients suffer from a jumbling-together of the features of the objects in their visual field. They also provide intriguing anecdotal support for

¹⁰ The Pouget and Sejnowski theory is based upon linear combinations of continuous functions. It may be suggested that such a model can (in principle) be approximated to an arbitrary degree of accuracy by a Turing machine. Supposing this to be true, nothing is revealed about which model most accurately captures what the brain is really doing. The behavior of the solar system could also be approximated to an arbitrary degree of accuracy by a Turing machine, but insisting that planetary motion actually involves symbol manipulation according to syntactically specified rules is unrewarding. That Turing equivalence is irrelevant here is further illustrated by the fact that the Pouget and Sejnowski model could also be implemented by analog VLSI, which is about as nonsymbolic as you can get.

¹¹ A model which, it should be mentioned, was originally developed to explain the behavioral data on feature binding in normals.

this hypothesis. Their patient spontaneously hit upon the technique of looking at individual objects through a tube, which would be a useful way to eliminate the features of other objects (or, on the Posner/Farah accounts, distractors). Commenting on experimental stimuli, he said things like, “It looks like sometimes one letter is going into the other one. I get a double identity. It kind of coincides”, and “I can only see two or three [dots] out of the whole bunch.”

For our purposes, several things are interesting about the Robertson *et al.* account. First, it is another example of a model based on the known structure of the (healthy) brain which produces good predictions about the behavior of damaged brains. If one of the goals in modeling is to understand the ways in which humans actually perform tasks, then this sort of prediction is clearly an important test (and to date we have been unable to locate a *single* nontrivial prediction of this sort made—let alone made and confirmed—by a symbolic model of spatial representation). Second, though the Robertson *et al.* model does not specify an implementation structure, it is clear that any attempt to implement it as a strongly symbolic system would run directly into the computational explosion that has become famous in AI as the “frame problem” (discussed below), since it calls for the dynamic binding of large numbers of features on a constantly shifting visuospatial array. And finally, it highlights an important fact about human spatial representation, namely that it is fundamentally distributed. Between them, the dorsal and ventral pathways extract and represent all the visuospatial information needed by other areas of the brain, but at no point do they put it together into stable intersubstitutable entities of the sort required for symbolic representation and manipulation.

5. Spatial Reasoning and the Science of Mind

A central feature of symbolic approaches is that they demand a sharp distinction between the symbolic and non-symbolic powers of the brain. The most common form of the distinction is derived from *functionalism*, the philosophical hypothesis that mental states are defined by their roles in a functional (usually, computational) economy of other such states, independently of their physical instantiations.¹² This approach to the meaning (or “intentionality”) of mental states, combined with the computationalist approach to cognition that grew out of the work of Turing and of Newell and Simon, motivated a contrast between levels of description:

What psychology is about is the causal structure of the mind at the intentional level of description. It may very well be, for all I know, that at some neurological level it is a

¹² The other common form of the distinction is in terms of modularity. Since all the modules ultimately have to be implemented in wet stuff, though, the problems will be much the same.

system that obeys connectionist postulates. That is of no interest to psychology... the question of what the causal structure of the brain is like at the neurological level might be settled in a connectionist direction and leave entirely open what the causal structure of the brain is like at the intentional level. (Fodor 1995)

There are a number of problems with this approach. For one thing, functionalism has been shown to have severe problems with such basic human functions as sensation, analogical reasoning, and conceptual change, and ultimately it failed in its central goal of providing a functional/computational account of the meanings of words and mental states (which is why it was abandoned by its creator, Hilary Putnam). There remains no respectable argument, in philosophy of mind or in any of the cognitive sciences, for the *a priori* division of the mind/brain into autonomous levels of description. Of course, none of this rules out the *possibility* that our mind/brain just happens to be so divided; let us therefore take this Fodorian claim as an empirical hypothesis, and see how it fares.

Assume for the moment that it *is* worthwhile to draw a line between the intentional, symbolic level of description and the implementational, neural level. What should we say about spatial reasoning? On which side of the line does it belong? In those cases where they have been attempted, symbolic accounts of spatial processing have met with very limited success; even leaving aside questions of their empirical plausibility, such models are haunted by their own personal bugbear, the infamous “frame problem” (See Pylyshyn 1987 for reviews). Simplified, the frame problem says that deductivist representational systems must derive moment-to-moment properties of the represented world from a stored corpus of first-order axioms and a fairly sparse input. Unfortunately for these models, contingency is a fact of life in the real world, and it is essentially unbounded. Brute force prediction of chess moves is a huge, but finite problem. Brute force prediction of large-scale spatial interactions in a highly contingent world is beyond the means of any physical symbolic computer, present or future.¹³

Looking at actual spatial reasoners—ourselves—yields the same conclusion. There doesn't seem to be any scientifically useful level of description at which spatial reasoning is best described as an implementation-independent symbolic process. From single-cell response properties in area 7a, through visual and mental images and conscious control of movement, and all the way

¹³ Neurally inspired models may handle the problem of contingent response with less difficulty. For instance, it may be convenient to represent even deterministic contingent events as being probabilistic in nature. Probabilistic events present little difficulty for neural nets; even deterministic network models can encode relative probabilities with ease. If there is anything which neural nets are spectacularly good at, it is statistical analysis. Likewise, the problem of relative salience—that is, of predicting which among a vast list of features in a complex environment may determine change—has an elegant solution in the form of the delta rule and other connectionist learning rules. These rules do nothing *but* assign relevance to the different elements of a multidimensional input. Finally, of course, massively parallel architecture solves the problem of determining which elements need to be active at any given moment, by the straightforward expedient of activating them all.

up to mathematical problem-solving and social negotiation, there seems to be a smooth continuity of functional and explanatory dependence. These “levels of analysis” interact with each other, and the character of that interaction has a lot to do with how we see and feel, integrate across modalities, learn and recall, shift attention, and find our way around in space. If there is a line to be drawn, it clearly can’t be within or below the domain of spatial reasoning.

Nor does it seem appropriate to draw the line above it, excluding spatial reasoning from the realm of the intentional. When you solve a graphically presented geometry problem by inspection, you don’t feel disappointed at having failed to engage in intentional thought; quite the contrary. If the problem was a difficult one, you probably feel pleased with your intellect, and relieved at having solved the problem by cleverness instead of by “brute force” calculation. It is no accident that quarterbacks in (American) football, the players who must quickly evaluate complex spatial patterns composed of intentional agents, must pass an intelligence test before being considered for the major leagues.

It seems, then, that one can find no level of description at which intentional and symbolic thought will be coextensive. Spatial reasoning (along with social reasoning, creativity, and a host of other human functions) gets in the way: at least some intentional thought is spatial, and spatial thought is not best understood in symbolic terms. The functionalist levels-of-analysis hypothesis is simply incompatible with the data.

None of this is to deny that some kinds of thought *are* well described as symbolic. There is at least one form of symbolic activity—language—which all normal humans (and only humans) are astonishingly good at, and many others—algebra, formal logic, chess—which can be learned with greater or lesser degrees of difficulty. These activities display, in an approximate and finite form, many of the properties of genuine symbolic systems.¹⁴ At present there is very little understanding of the way in which the brain conducts these operations, and we regard this as a deep and pressing issue in cognitive science. Symbol manipulation is a strange and wonderful power; understanding how we do it would be a transformative advance in the science of mind.¹⁵

What we do deny is that the way to understand symbolic cognition—or for that matter, *any* form of cognition—is to divorce it from other forms of thought and from the body in which it occurs. In the face of the work discussed earlier, and a host of other studies conducted by Ramachandran, the Damasio, and others, it seems incontestable that studying the brain can provide deep insight into the nature of the mind. To be sure, this doesn’t mean that every theory in the cognitive sciences has to make mention of the brain; but it does mean that actively *denying* the relevance of the body and

¹⁴ Which should come as no surprise, since the nonhuman symbolic systems with which we are now so familiar—from the hypothetical Turing machine through the desktop computer—were created in the image of these human competencies.

¹⁵ For one recent attempt, see Deacon (1997).

brain—to *any* cognitive function, however “high” or abstract—can only hinder the progress of understanding.

Why, then, do so many symbolic computationalists insist on just such a denial? Why do they feel compelled to claim sole dominion over the task of explaining the human mind (or at least, the good parts), instead of acknowledging that the many levels of description and analysis should be richly interpenetrated and mutually informative?

The answer lies in the motivation which drove the development of mechanical symbol systems. As Clark (1993) puts it, a physical symbol system—however instantiated—must be “semantically well-behaved”. It must be so constituted that some aspects of its physical states can be mapped systematically onto semantic states, and that the transformations in its physical states map onto legal moves in semantic space. For example, in the case of a calculator, this means that each of its (normally reachable, electronic) states must have some mathematical meaning, and that the effect of pressing a button should be to bring about a new state that is mathematically related to the previous state.

What makes semantically well-behaved systems so useful is that they’re *truth-preserving*. If you start with a state that represents some true fact, then proper manipulations of the system will always result in further true facts, never in falsehoods. In short, these systems are reliable tools for doing deductive logic and mathematics. It is clear why one might hope that the human brain would turn out to be such a system: it would provide a natural explanation of our formal abilities; it would give us a reason to believe that the deep structure of our thought will be amenable to analysis using the tools and vocabulary that are available to us now; it would even imply heartening things about the educability and rationality of human beings. But there is a catch: to guarantee truth-preservation,¹⁶ a system has to be *purely* symbolic. If it overlaps or interacts (beyond the level of input/output relations) with some nonsymbolic system, then its state transitions will not be governed solely by their syntactic relations, and thus their semantic validity may be compromised.

The problem is, when you look at the actual human mind/brain, truth-preservation *in this formal sense* does not seem to be much of a priority. Both in neuroscience and in psychology, human beings turn out to be chock full of useful fictions. Human perception and reasoning are shot through with domain-specific heuristics and short-cuts, with confabulations, with undischarged assumptions (and if any of our readers find this characterization implausible, we recommend that they try their hand at teaching an introductory course in formal logic). Wherever we look in the brain, we find systems whose structure is very ill-suited to guarding against the

¹⁶ And other beloved properties such as compositionality, the ability to “nest” statements to arbitrary depth by substituting complex expressions for simple symbols.

possibility of falsehood but very well-suited to swift and flexible perception and action.

What we have tried to show is that there is a poor fit between the actual structure of the human mind/brain and the central motivating principles of the symbolic approach. This does not mean that symbolic characterizations of thought are useless; they still constitute the only general, systematic, intersubjective way of compactly describing high-level thought. What it does mean is that we should not accept any claims about the ubiquity or autonomy of symbolic processes. Thinking in terms of the brain's more pragmatic goals provides a better framework for asking questions, building models, and understanding the mental and physical structure of thought.