Appears in: *Psychological Review*, 2002, *v. 109*, pp. 156–185.

Frames of reference in unilateral neglect and visual perception: A computational perspective

Michael C. Mozer
Department of Computer Science
and Institute of Cognitive Science
University of Colorado
Boulder, CO 80309–0430
`mozer@cs.colorado.edu`

Abstract

A key question motivating research in perception and attention is how the brain represents visual information. One aspect of this representation is the reference frame with respect to which visual features are encoded. The reference frame specifies the origin of coordinate axes, the up-down, left-right, and front-back directions, and the relative scale of each axis. Reference frames can be prescribed by the viewer's gaze, head, or trunk, by intrinsic characteristics of an object, or by the environment.

To determine the frames of reference involved in human vision and attention, neurological patients with unilateral neglect have been extensively studied. Neglect patients often fail to orient toward, explore, and respond to stimuli on the left. The interesting question is: with respect to what frame of reference is neglect of the left manifested? When a neglect patient shows a deficit in attentional allocation that depends not merely on the location of an object with respect to the viewer but on the extent, shape, or movement of the object itself, the inference is often made that attentional allocation must be operating in an object-based frame of reference. Via simulations of an existing connectionist model of spatial attention, we argue that *this inference is not logically necessary: object-based attentional effects can be obtained without object-based frames of reference*.

The model, called the AM, has previously accounted for a variety of attentional effects in normals (Mozer, 1991; Mozer & Sitton, 1998) as well as reading deficits in neglect dyslexia (Mozer & Behrmann, 1992) and line bisection performance in neglect (Mozer, Halligan, & Marshall, 1997). The AM operates entirely in an egocentric reference frame; visual features and locations are encoded with respect to eye fixation. The AM selects or *activates* a region in a viewer-based representation of space for preferential processing based on three principles: (1) Locations in which visual features appear should be activated; (2) locations adjacent to activated locations should also be activated; and (3) locations whose activity grows the slowest should be suppressed.

From these three principles emerge an explanation of key neuropsychological data concerning object-based attention in neglect, including the following. Behrmann and Tipper (1994) observed that neglect remains with the left side of an object when the object is inverted. However, this reversal of neglect is not observed for displays in which the left and right sides of the object are disconnected (Tipper & Behrmann, 1996) or for additional, fixed objects in the display (Behrmann & Tipper, 1999). Pavlovskaya, Glass, Soroker, Blum, and Groswasser (1997) improved accuracy of object identification by cuing patients to a location left of the center of the object. Arguin and Bub (1993) found reaction times longer for a stimulus located on the left side of an object than the right side, controlling for retinal position of the stimulus. In two different paradigms, Driver and Halligan (1991), Driver, Baylis, Goodrich, and Rafal (1994), and Driver and Baylis (unpublished) observed neglect operating with respect to the principal axis of an object. The authors of these studies interpreted their data as evidence that the allocation of attention is subserved by object-based frames of reference. The AM model provides an alternative explanation of these phenomena that does not require object-based representations in visual attention and perception. By calling into question the common assumption that object-based reference frames are required for ordinary perception, the AM eliminates some of the strongest support for *allocentric* or *structural-description* theories of object recognition.

Reference frames in visual perception

A key question motivating research in perception and attention is how the brain represents visual information. One aspect of this representation is the reference frame with respect to which visual features are encoded. The reference frame specifies the center location, the up-down, left-right, and front-back directions, and the relative scale of each axis. Figure 1 shows two different reference frames. In terms of the reference frame centered on the telephone, the buttons would be described as being on *top* of the base and in *front* of the handle, whereas in terms of the reference frame centered on the viewer, the buttons would be described as being in *back* of the base and to the *left* of the handle. Reference frames can be prescribed by the viewer, objects, or the environment. *Viewer-based* frames are determined by the gaze, head orientation, and/or torso position of the viewer. *Object-based* frames are determined by intrinsic characteristics of an object, such as axes of symmetry or elongation, or knowledge of the object's standard orientation. *Environment-based* frames are based on landmarks in the environment, such as walls in a room, or other absolutes, such as gravity or compass directions.[1] A frame of reference makes certain information in the stimulus explicit and therefore readily available for utilization in information processing, and hides other information.

Object recognition

In the visual system, the features or parts of an object must be encoded with respect to one or more reference frames. The appropriateness of one reference frame or another depends fundamentally on the mechanisms posited to operate in object recognition. Thus, in considering the reference frames likely to be utilized in the neocortex, it is worth reviewing theories of object recognition, particularly with regard to the contrasting claims they make as to the frames of reference involved in perception.

The goal of object recognition is to determine the identity or category of an object in a visual scene from the retinal input. In naturalistic scenes, object recognition is a computational challenge because the object may appear in various poses and contexts—i.e., in arbitrary positions, orientations, and distances with respect to the viewer and to other objects. Consequently, theories of object recognition must overcome the effect of viewpoint on the appearance of an object. *Viewpoint-invariant recognition* (also known as *object constancy*) is achieved when an object is identified as being the same regardless of its pose.

Object recognition involves matching representations of objects stored in memory to representations extracted from the visual field. The debate in the literature concerns the nature of the representation extracted from the visual field. Theories of object recognition can be broadly divided into two classes, those that propose a level of object representation that is largely viewer independent, and those that propose that all representations utilized in the course of recognition are tied to the appearance of the object in the visual field. We will refer to these two classes as *allocentric* and *egocentric*, respectively, to denote the fact that the former class invokes allocentric frames of reference, whereas the latter class invokes purely egocentric frames of references. (Tarr, 1999, and Hummel and Stankiewicz, 1998, term the two classes as *structural description* and *view-based* theories, for reasons that will become clear shortly. We prefer the terms allocentric and egocentric to highlight the involvement of reference frames.)

An early and highly influential allocentric theory of object recognition was outlined by Marr and Nishihara (1978; see also Marr, 1982, and Pinker, 1984). According to this theory, an internal description of an object's structure is constructed from observed visual features, essentially by transforming the viewer-based retinal input to an *object-based representation* (Figure 2a)—a representation of the relationship of object features or parts to one another, independent of the viewer, using object-based frames of reference. This transformation solves the problem of viewpoint-invariant recognition, because every view of an object maps to the same object-based representation (ignoring the issue of occluded features).

Since Marr and Nishihara's seminal work, a variety of allocentric theories have been proposed. Some focus on rigid and two-dimensional objects—either implicitly or explicitly—in which case the object-based representation can be constructed using only image plane transformations and recognition can be achieved in a straightforward manner by matching the object-based representation to stored templates of familiar objects (Hinton, 1981; Humphreys & Heinke, 1998; Olshausen, Anderson, & van Essen, 1993; Zemel, Mozer, & Hinton, 1988). More recent allocentric theo-

---

[1] Alternative terminology for these three reference frames abounds in the literature. Retinotopic, head-centered and body-centered are specific instances of viewer based; stimulus based is equivalent to object based; and gravitational and scene-based are instances of environment based. Further, egocentric is often used as a synonym for viewer based, and allocentric as a synonym for object or environment based.

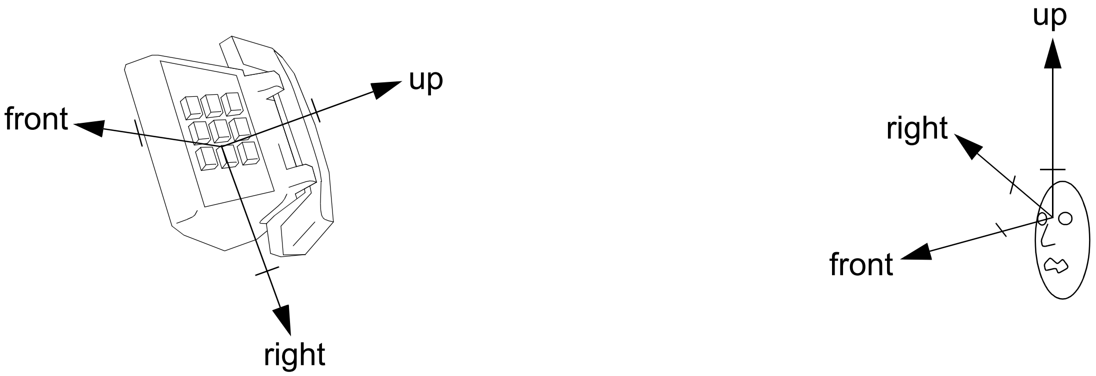


FIGURE 1. Two reference frames that can describe the telephone, one of which is intrinsic to the object and the other is based on the viewer's gaze. The reference frame prescribes the center location, the up-down, left-right, and front-back directions, and the scale of each axis (indicated by the mark along each axis).

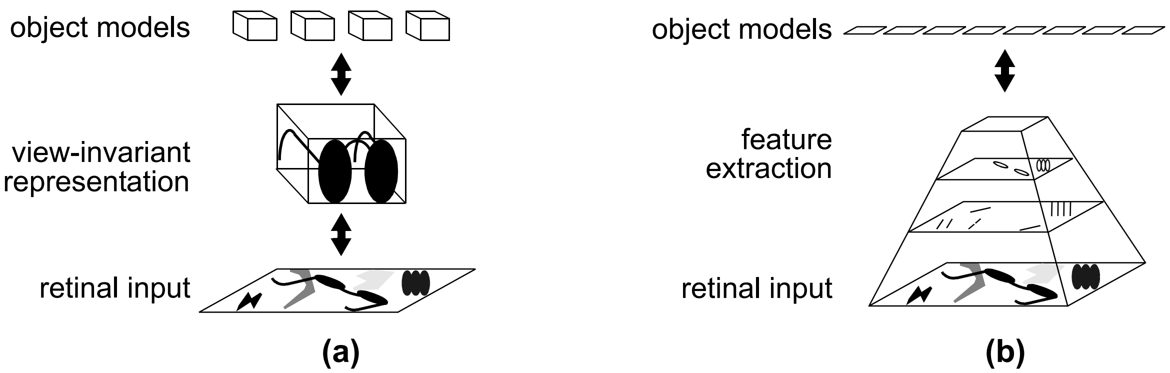


FIGURE 2. A sketch of two classes of theories of object recognition. (a) In an *allocentric* theory, visual features in the retinal image corresponding to an object (the sunglasses) are detected and transformed into a view-invariant representation that captures the three-dimensional structure of the object, and the relationships of its parts to one another. Recognition is then performed by matching stored object templates—one per object—to the view-invariant representation. (b) In an *egocentric* theory, many transformations of the features in the visual input are considered in parallel, and typically recognition is achieved via a multistage hierarchical process that constructs increasingly complex featural representations with increasing viewpoint invariance. The pyramid structure indicates the transformation from low-order to high-order features while simultaneously factoring out some position, scale, and possibly orientation specificity. Recognition is performed by matching the representation at the top of the pyramid to stored object templates, where multiple templates of the same object are required to achieve complete viewpoint invariance.

ries in psychology tackle the task of recognizing the sorts of objects encountered in the real world—complex, articulated objects with three-dimensional structure (Biederman, 1987; Hummel & Biederman, 1992; Hummel & Stankiewicz, 1996, 1998). In the style of Marr and Nishihara, these modern theories operate on structural descriptions that decompose an object into its parts. Structural descriptions can be hierarchical, decomposing parts into parts of parts. The parts are described in terms of a relatively small catalog of primitives, each characterized by view invariant attributes. Parts are linked by a small vocabulary of spatial relationships that describe how one part is situated with respect to another.

For example, the Biederman (1987) theory invokes 36 primitives, called *geons*. Geons are detected by nonaccidental properties in an image. The attributes of a geon include whether its edges are straight or curved, whether its size is constant, expanded, or contracted, and whether it is symmetric about one or more axes. The relationships among geons are captured by qualitative properties such as whether one geon is larger than another, whether one geon is above, below or to the side of another, and whether the geons are joined end-to-end or end-to-side. (The model of Hummel & Stankiewicz, 1998, can accommodate quantitative as well as qualitative spatial relationships.) Because in principle any shape can be described by the finite set of primitives and relationships, the Biederman theory allows for the representation of novel shapes. Geons and their relationships can be represented by a graph structure, and recognition of an unknown shape can be performed by matching the structure of the unknown shape to the structure of a stored object. Modern theories such as Biederman's make better contact with psychological data than the work of Marr and Nishihara, but share the essential characteristic of a level of representation that is nearly or completely allocentric.

In contrast, *egocentric* theories of object recognition suppose that representations in the visual system are tied to the appearance of objects as viewed. Consider a naive scheme in which two-dimensional images of viewed objects are stored during learning, and these stored templates are matched pixel-by-pixel to an image containing an object to be recognized. Although this scheme seems hopelessly simplistic, minor variants of the scheme achieve a surprising degree of generalization to novel poses of three-dimensional objects by interpolation between stored poses (Poggio & Edelman, 1990; Poggio & Shelton, 1999; Romano, 1993; Siebert & Waxman, 1990; Ullman, 1989; Weinshall, Edelman, & Bülthoff, 1990).

Egocentric theories that focus on explaining human vision go beyond the simple two-dimensional template matching idea by taking into account the coarse-scale anatomy of visual cortex (Fukushima & Miyake, 1982; Hubel & Wiesel, 1979; Le Cun, Boser, Denker, Henderson, Howard, Hubbard, & Jackel, 1989; Mozer, 1991; Perrett & Oram, 1998; Reisenhuber & Poggio, 1999; Sandon & Uhr, 1988; Wallis & Rolls, 1997). Specifically, two properties of visual cortex are generally deemed relevant. First, visual cortex is hierarchically organized, with simple, low-order, view-specific feature detectors at the earliest stage of vision, and increasingly complex, higher-order, and view-invariant detectors at subsequent stages (depicted in Figure 2b by the pyramid structure). Second, information is processed from many locations in the visual field simultaneously and at many scales and orientations. Rather than being forced to choose a single reference frame, parallelism of the visual system allows multiple transformations of detectors at each stage in the hierarchy. In this framework, the focus of processing is on extracting features that reliably indicate the presence of an object, not on constructing a view-invariant representation. Nonetheless, a partially view-invariant representation may come to dominate later stages of processing in the service of recognition.

One fundamental distinction between allocentric and egocentric theories of object recognition is that allocentric theories suppose a stage of processing at which objects achieve a view-invariant representation, whereas egocentric theories suppose that objects are encoded in a view-dependent manner at all stages. In examining the theories more closely, however, this dichotomy is somewhat artificial: one might imagine allocentric theories in which representations are not fully view invariant, and egocentric theories allow for representations that are not fully view dependent. Thus, the two classes of theories might better be understood as lying on a continuum of view invariance. Perrett, Oram, Hietanen, and Benson (1999) describe allocentric theories as an extreme class of model in which any view of an object can be transformed to match a *single* stored description, whereas egocentric theories require *multiple* characteristic views, and achieve a limited degree of transformation.

Allocentric versus egocentric theories

In recent years, a heated debate has ensued between proponents of modern allocentric and egocentric theories (Biederman & Gerhardstein, 1995; Biederman & Kalocsai, 1997 Edelman, 1997; Edelman & Duvdevani-Bar, 1997; Perrett et al., 1999; Tarr, 1999; Tarr & Bülthoff, 1995). Arguing the strengths and weaknesses of the theories is beyond the scope of this article, but we summarize some of the key evidence that has been fodder for the debate. We sort the evidence into three categories: neuroscientific, behavioral, and computational.

Neurophysiological studies have not identified a stage in visual information processing at which cells fire in a completely view invariant manner. Rather, even at the latest purely visual stages of processing in monkeys, such as area TE of inferotemporal cortex, cells remain somewhat sensitive to viewing angle and articulation poses for objects, faces, and geometrical shapes (Logothetis & Pauls, 1995; Perrett & Oram, 1998; Tanaka, 1993, 1996). Indeed, for faces, cells are tuned to characteristic views, consistent with egocentric theories (Perrett, Oram, Hiertanen, & Benson, 1999).

The behavioral evidence concerning the viewpoint invariance of object representations and recognition performance is somewhat more ambiguous. For example, Ellis, Allport, Humphreys, and Collins (1989) asked subjects whether two pictured objects had the same name. A benefit was found if the two familiar objects were the same, even from different viewpoints, suggesting the use of allocentric representations. However, a benefit was also found if the objects were presented from the same viewing angle and were of the same size.

Psychophysical support is found for view-specific object representations (Vetter, Hurlbert, & Poggio, 1995; Tarr, 1995), as well as for theories positing multiple specific views with interpolation between (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992). Tarr and Pinker (1990; see also McMullen & Farah, 1991) observed the use of object-based frames only in special cases of recognition, but suggest that the ordinary visual reference frame is tied to egocentric coordinates.

In studies using novel objects, results are generally consistent with allocentric theories. For example, viewpoint-invariant recognition is achieved for unfamiliar, depth rotated, compositional objects (Biederman & Gerhardstein, 1993; Biederman & Bar, 1999; see further discussion by Biederman & Gerhardstein, 1995, and Tarr & Bülthoff, 1995). Some studies have found that perception of unfamiliar 3D objects is aided by prior exposure to objects at a particular viewpoint and is disrupted by rotation in depth (Bülthoff, Edelman, & Tarr, 1995), but such data may not be evidence against allocentric theories. The Biederman (1987) theory, for example, does not predict invariance for all objects and views. Recently, however, costs in recognition time and accuracy have been found for viewpoint changes in recognizing single geons of the sort postulated by Biederman's theory (Tarr, Williams, Hayward, & Gauthier, 1998).

Complementing the evidence for egocentric and allocentric theories from experimental studies are arguments about the computational complexity and feasibility of object recognition under the two different classes of theories. A major issue faced by egocentric theories is explaining how they can generalize to novel objects and novel viewing conditions (Tarr, 1999; although see Reisenhuber & Poggio, 1999, for a promising model in this regard), and how they can be used to achieve broad category-level judgements (Hummel & Stankiewicz, 1996). Allocentric theories, in turn, must be proven to be computationally sufficient to extract three-dimensional structure from images in a primarily bottom-up fashion.[2]

In summary, although the neurobiological, psychological, and computational evidence somewhat favors egocentric theories, it provides mixed support for both allocentric and egocentric theories. Due to the increasing sophistication of theories in both camps, the debate has become increasingly intricate and complex. Nonetheless, we can strip away much of the subtlety of the various theories, such as the exact nature of structural descriptions in allocentric theories, by focusing on a simple, almost trivial, case: rigid two-dimensional shapes. Although modern allocentric theories accommodate articulated three-dimensional objects via structural descriptions, the theories reduce to a simpler state of affairs for rigid two-dimensional objects: the view-invariant representation can be established via a single object-based frame of reference. Consequently, evidence for the use of object-based frames of reference in the recognition of rigid two-dimensional objects would certainly bolster the case for allocentric theories, and would, on grounds of parsimony, be debilitating for egocentric theories.

A rich source of data diagnostic of the neurobiological and psychological reality of object-based frames of reference is obtained in the neuropsychological literature via the study of patients with *unilateral neglect*. The interest in neglect has been intense, in large part because neglect appears to provide compelling support for the existence of object-based frames of reference, and therefore, for allocentric theories. In the following section, we describe neglect and its relation to theories of object recognition.

---

[2] Interactive approaches in which object-based representations and object identity are determined in parallel through a constraint satisfaction search (e.g., Humphreys & Heinke, 1998) have proven difficult to scale up, due to the massive combinatorial search that has many local optima (Hinton & Lang, 1985; O'Reilly, Munakata, & McClelland, in press, Chapter 6).

Unilateral neglect

Damage to parietal cortex can cause patients to fail to orient toward, explore, and respond to stimuli on the contralesional side of space (Farah, 1990; Heilman, Watson, & Valenstein, 1993). This disorder, known as *unilateral neglect*, can compromise visual, auditory, tactile, and olfactory modalities and may involve personal, extrapersonal, and imaginal space (Halligan & Marshall, 1993). Unilateral neglect is more frequent, longer lasting, and severe following lesions to the right hemisphere than to the left. Consequently, all descriptions in this paper will refer to right-hemisphere damage and neglect of stimuli on the left. The interesting question surrounding unilateral visual neglect is: With respect to what reference frame is left neglect manifested?

In ordinary viewing situations, viewer-based and object-based frames of reference are often confounded, as when viewers, seated upright, fixate on an upright object immediately in front of them. Consequently, clever behavioral experiments have been designed to dissociate various reference frames and determine the contribution of each to neglect. In multiple experiments, patients show a deficit in attentional allocation that depends not merely on the location of an object with respect to the viewer, but on the extent, shape, or movement of the object itself. From this finding of *object-based neglect*, the inference is often made that attentional allocation must be operating in an object-based frame of reference, and consequently, object-based frames of reference are central to visual information processing. The observation of object-based neglect suggests the existence of view-invariant representations, and therefore supports allocentric as opposed to egocentric theories.

The point of this work is to show that this inference is not logically necessary: *Object-based attentional effects can be obtained without object-based reference frames*. Consequently, the bulk of the neglect data that has been mustered as strong support for allocentric theories of object recognition is equally consistent with egocentric theories. We argue this point via a computational model that utilizes only viewer-based frames, yet can account for data from a broad variety of experimental studies that were interpreted as supporting object-based frames. Through simulations of the computational model, it becomes evident that the data is trickier to interpret than one might at first imagine.

In the next section, we present the model and explain key principles of the model that will allow us to account for data. Then, we show simulation results for several different studies. And we conclude with a discussion of other data in the literature that has been used as evidence for and against the neurobiological reality of object-based frames.

A connectionist model of visual perception and attention

MORSEL (Mozer, 1991; Mozer & Sitton, 1998) is a connectionist model of visual perception and attention. The model has previously been used to explain a large corpus of experimental data, including perceptual errors that arise when several shapes appear simultaneously in the visual field, facilitatory effects of context and redundant information, visual search performance, attentional cueing effects, reading deficits in neglect dyslexia (Mozer & Behrmann, 1992), and line bisection performance in neglect (Mozer, Halligan, & Marshall, 1997). MORSEL (Figure 3) includes a *recognition network* that can identify multiple shapes in parallel and in arbitrary locations of the visual field, but has capacity limitations. It is an egocentric theory, of the sort depicted Figure 2b. MORSEL also includes an *attentional mechanism* that determines where in the visual field to focus processing resources.

Visual input presented to MORSEL is encoded by a set of feature detectors arrayed on a topographic map. The detectors are of five *primitive feature types*: oriented line segments at 0°, 45°, 90°, 135°, and line-segment terminators (ends of line segments). Figure 4 shows a sample input to MORSEL, four letters centered on corners of a square, where the representation of each letter occupies a 3×3 region of the topographic map. The upper panel presents the superimposed features, and the bottom panels separate the topographic map by feature type. In these separate maps, a dark symbol indicates activity of the detector for the given feature at the particular location, a light symbol indicates inactivity. Activity from the topographic map innervates both the recognition network and the attentional mechanism.

In our earlier modeling work, we stipulated that the topographic map is in a viewer-based reference frame, meaning that the input representation changes as the viewer moves through the world. However, our earlier work did not require us to commit as to the precise nature of the viewer-based frame, whether it be retinotopic, head centered, or body centered. Because the experimental paradigms that we simulate in this work confound eye, head, and body position, the various viewer-based frames are equivalent, and no specific commitment is required now either.

MORSEL is primarily a model of psychological structures, not neurobiological structures. One might treat MORSEL's primitive visual features as corresponding to primary visual cortex, the AM as corresponding to parietal

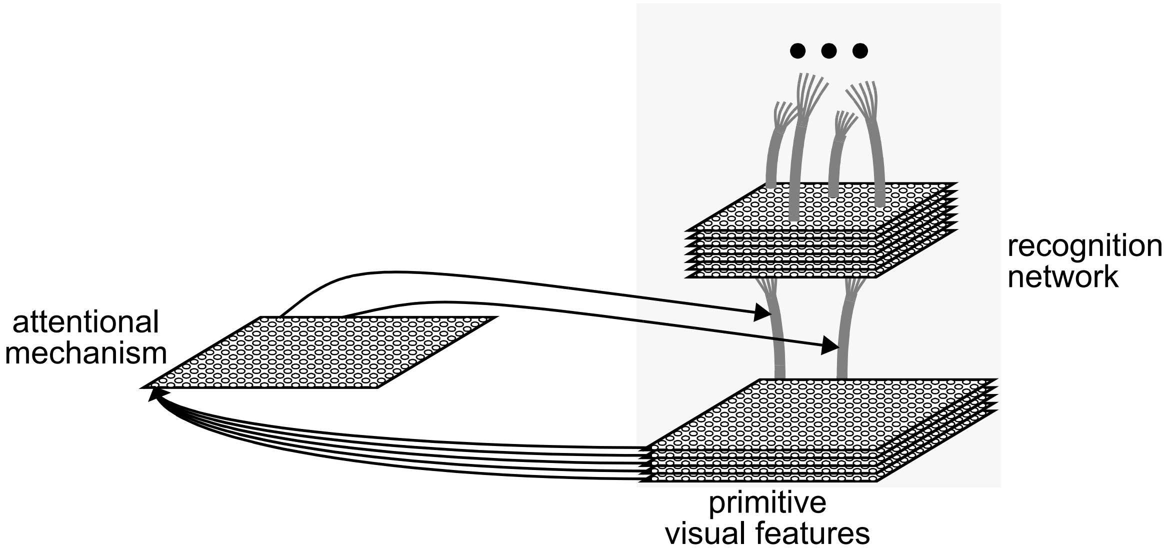


FIGURE 3. Key components of MORSEL (Mozer, 1991). MORSEL includes a recognition network, the first stages of which are depicted against a grey background, and an attentional mechanism.

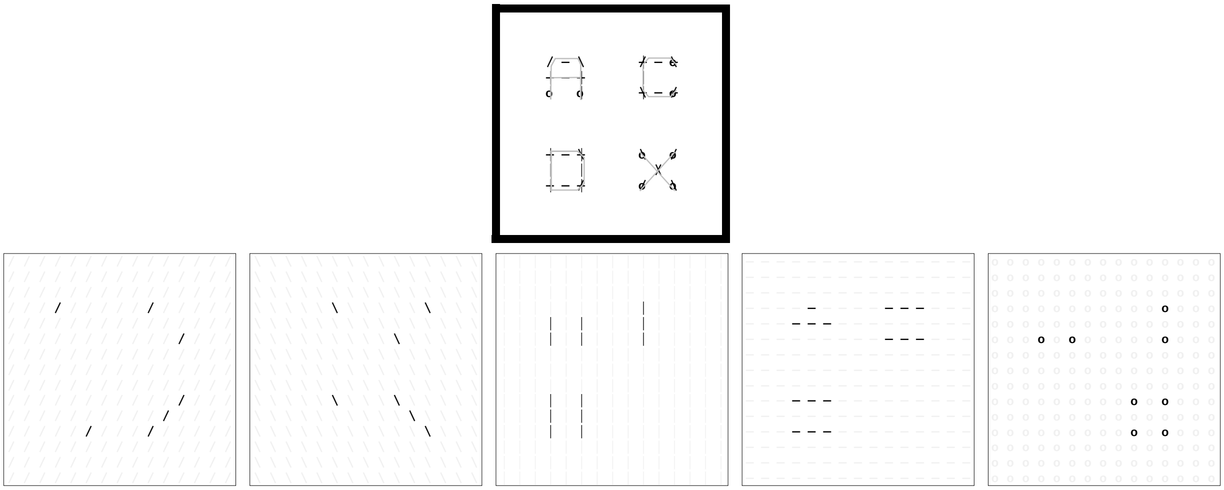


FIGURE 4. Top panel: Sample input to MORSEL, consisting of the letters A, C, D, and X, encoded in terms of five primitive features—line segments at four orientations and segment terminators (circles). Bottom panels: Feature map activity corresponding to the sample input. A dark symbol indicates the activity of the detector for a particular feature in a particular location; a light symbol indicates inactivity.

cortex, and the recognition network as being situated in the temporal pathway. Beyond this loose fit to neurobiology, we do not commit to a neurobiological instantiation at present. We instead treat MORSEL as a psychological-level theory which describes functional processing in neocortex. Consequently, we characterize processing units in the model in terms of their functional properties, not neurobiology. For example, the left visual field is represented by units on the left side of the primitive feature maps, even though those units would correspond to V1 neurons in the right cerebral hemisphere.

MORSEL is also not intended as a model of human development. The recognition network is trained to reproduce adult competence, but MORSEL makes no claims as to the nature of developmental processes that give rise to adult competence in visual perception. The connectivity of the attentional mechanism is determined by principles we describe in the next section. The connectivity is fixed in all simulations; connectionist learning procedures are not involved.

MORSEL is a comprehensive model consisting not only of the recognition network and attentional mechanism depicted in Figure 3, but several other elements that we sidestepped because of their irrelevance to the present work. Simulating the entire model can be a problem, because it is difficult to identify which components or properties of the model are responsible for producing a certain behavior. Consequently, our strategy has been to simulate only the critical components of the model, and to make simple assumptions concerning the operation of other components. We adopt this strategy in the present work, and use only the attentional mechanism to account for data from unilateral neglect, much as we did in Mozer, Halligan, and Marshall (1997).

The attentional mechanism

The attentional mechanism, or *AM* for short, is a set of processing units in one-to-one correspondence with the locations in the topographic map. Activity in an AM unit indicates the salience of the corresponding location, and serves to gate the flow of activity from feature detectors at that location in the topographic map into the recognition network (indicated in Figure 3 by the connections from the AM into the recognition network); the more active an AM unit is, the more likely that features in the corresponding location of the topographic map will be detected and analyzed by the recognition network. However, the AM serves only to bias processing: features from unattended locations are not absolutely inhibited, but have a lower probability of being detected by the recognition network.

Each unit in the AM receives bottom-up or *exogenous* input from the detectors in the corresponding location of the topographic map (indicated in Figure 3 by the connections from the primitive features to the AM). Each unit in the AM can also receive top-down or *endogenous* input from higher centers in the model, but this aspect of the model is barely utilized in the present research. Given the exogenous and endogenous input, cooperative and competitive dynamics within the AM cause a subset of locations to be activated.

Figure 5 shows an example of the AM in operation. Each panel depicts the state of the AM after various numbers of processing time steps or *iterations*, assuming a 15×15 topographic map. The area of a black square is proportional to the exogenous input at that location in the topographic map. The area of a white square is proportional to the AM activity. The white squares are superimposed on top of the black; consequently, the exogenous input is not visible at locations with AM activity. Initially, at iteration 0, the AM is reset and has no activity. Three distinct blobs of feature activity are evident on the input, but as processing proceeds, the AM selects the largest blob. Note that the input blobs do not indicate the type or precise arrangement of features, just the total activity in a region. Although the model appears to have formed a spotlight of attention, the dynamics of the model do not mandate the selection of a convex region, or even a single region. Typically, however, a single region is selected, and the selected region conforms to the shape of objects in the visual input, tapering off at object boundaries. Under certain circumstances, the model can select multiple regions, as one of our simulations will demonstrate.

The operation of the AM is based on three principles concerning the allocation of spatial attention, which most would view as noncontroversial:

(1) Attention is directed to locations in the visual field where objects appear, as well as to other task-relevant locations.

(2) Attention is directed to contiguous regions of the visual field.

(3) Attention has a selective function; it should choose some regions of the visual field over others.

These abstract principles concerning the direction of attention can be incorporated into a computational model like the AM by translating them into rules of activation, such as the following:

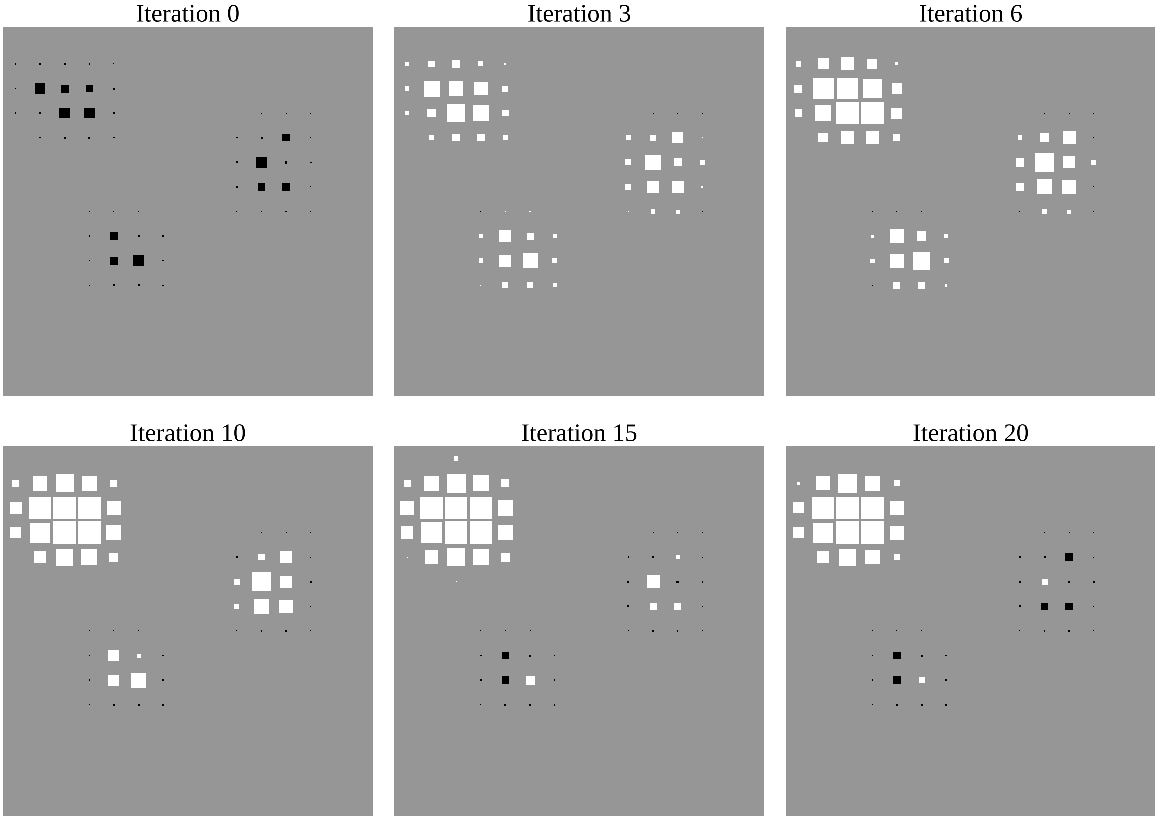

Iteration 0
Iteration 3
Iteration 6
Iteration 10
Iteration 15
Iteration 20

FIGURE 5. Example of the operation of the AM. Each panel depicts the state of the AM at a particular processing iteration, assuming a 15×15 topographic map. The area of a black square is proportional to the exogenous input at that location in the topographic map. The area of a white square is proportional to the AM activity. The white squares are superimposed on top of the black; consequently, the exogenous input is not visible at locations with AM activity. The exogenous input pattern indicates three objects, the largest one—the one producing the strongest input—is in the upper left portion of the field. By iteration 20, the AM has reached equilibrium and has selected the region surrounding the largest object.

(1) *Locations containing visual features should be activated*. This rule provides a *bias* on unit activity (i.e., all else being equal, the principle indicates whether a unit should be on or off). One can see this rule at work in Figure 5, where the initial activity of the AM (upper-middle frame) is based on the exogenous input (upper-left frame).

(2) *Locations adjacent to activated locations should also be activated*. This rule results in *cooperation* between neighboring units, and is manifested in Figure 5 by the increase in activity over time for the blob in the upper left portion of the field.

(3) *Locations whose activity is the weakest should be suppressed*. This rule results in *competition* between units, and is manifested in Figure 5 by the decrease in activity for the two lower blobs once the upper-left blob begins to dominate in activity.

These three rules qualitatively describe the operation of the model. The model can be characterized quantitatively through an update equation, which expresses the activity of a processing unit in the AM as a function of the input to the AM and the activities of other AM units. If we denote the activity of an AM unit at location (x, y) in the topographic map at a particular time t by $a_{xy}(t)$, then its new activity at the following time step is expressed as

$$a_{xy}(t+1) = f\left(a_{xy}(t) + exo_{xy} + \mu \sum_{\substack{i,j \in \\ \mathrm{NEIGH}_{xy}}} [a_{ij}(t) - a_{xy}(t)] - \theta[\bar{a}(t) - a_{xy}(t)] \right) , \tag{1}$$

where exo_{xy} is the exogenous input to the AM from features in the topographic map at location (x, y), f is a linear threshold function that caps activity at 0 and 1,

$$f(z) = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{if } 0 \le z \le 1 \\ 1 & \text{if } z > 1 \end{cases}$$

and NEIGH_{xy} is the set of eight locations adjacent to (x, y). The first term on the right side of Equation 1, $a_{xy}(t)$, causes a unit to sustain its activity over time. The second term, exo_{xy}, implements the bias rule. The third term implements the cooperation rule by causing an increase in activity when a unit is less active than its neighbors. Because it also causes a decrease in activity when a unit is more active than its neighbors, the third term can be viewed as encouraging a unit to take on the average value of its neighbors. Finally, the fourth term in Equation 1 implements the competition rule by causing a decrease in activity when a unit is less active than $\bar{a}(t)$, a measure of the average activity of AM, defined below. The parameters μ and θ are positive and weight the contribution to the activation dynamics of the cooperation and competition rules, respectively.

The most natural mechanism to perform selection is a winner-take-all competition rule, rather than the *comparison-to-average* competition rule we have proposed. Indeed, nearly every other connectionist model of attention uses a winner-take-all rule. A winner-take-all rule involves inhibition of a fixed magnitude between each pair of locations. With a winner-take-all rule, the total inhibition in the network rises and falls with the total activity in the network. Through this simple negative feedback loop, the network achieves homeostasis at a fixed level of activity, i.e., a distribution of attention of roughly fixed area. Because the size of the attentional spotlight has been shown to vary with task demands and stimulus properties (e.g., Halligan & Marshall, 1994; Hillis, Mordkoff, & Caramazza, 1999; LaBerge, 1983), we devised the comparison-to-average competition rule, which allows both small and large distributions of attention. If a region of the AM map has uniform activity, the comparison-to-average competition rule will support that level of activity, regardless of the size of the region. What matters is the homogeneity of activation within a region, and the activation of one region relative to another. If units in one region have more initial support than units in another region, and the support within a region is homogeneous, the competition rule will cause one region to be selected over the other. Contrary to the standard conception of a spotlight of attention, the comparison-to-average competition rule will allow two regions to be simultaneously active if they have nearly identical support, a property we later make use of to explain data.

The computation of the average activity, $\bar{a}$, requires some additional explanation. The fourth term in Equation 1 causes a unit's activity to be inhibited in proportion to $\bar{a}$. If $\bar{a}$ were simply the mean activity level of all AM units—i.e.,

$$\bar{a}(t) = \frac{1}{n}\sum_{x,y} a_{xy} ,$$

where *n* is the number of units in the AM—the level of inhibition would rise or fall as the total activity rises or falls, driving the total activity to remain roughly constant; consequently, the AM would tend to select a fixed-size region. The AM should be capable of attending to small or large regions, depending on the stimulus and task environment. To achieve this property, the inhibition between each pair of units is modulated by the number of active units, instead of what amounts to fixed inhibition between units. That is, $\bar{a}$ is defined as the mean activation considering only the active units, computed by replacing *n* with n_{ACT}, the number of active units:

$$n_{\mathrm{ACT}} = \lim_{\varepsilon \to 0} \sum_{x,y} \frac{a_{xy}}{\varepsilon + a_{xy}} .$$

As ε approaches zero, n_{ACT} becomes simply the number of units with positive activity levels.

In the original model, it turned out that to control the behavior of the AM, an additional *depreciation* factor, γ, was needed in the definition of $\bar{a}$:

$$\bar{a}(t) = \frac{\gamma}{n_{\mathrm{ACT}}}\sum_{x,y} a_{xy} ,$$

where $0 < \gamma \leq 1$. If $\gamma = 1$, a unit must have an activity level above the mean to remain on, but if $\gamma < 1$, the mean is depreciated and units whose activity is slightly below the mean will not be suppressed.

To explain the activation function intuitively, consider the time course of activation as depicted in Figure 5. Initially, the activity of all AM units is reset to zero. When a stimulus display is presented, features are activated in the topographic map, which provides exogenous input to the AM (second term in Equation 1). Units with active neighbors will grow the fastest because of neighborhood support (third term). As the flow of activation progresses, high-support neighborhoods will have activity above the mean; they will therefore be pushed even higher, while low-support neighborhoods will experience the opposite tendency (fourth term).

Lesioning the AM to produce neglect

To model data from neglect dyslexia (Mozer & Behrmann, 1992) and line bisection (Mozer, Halligan, & Marshall, 1997), we proposed a particular form of lesion to the model—damaging the connections from the primitive feature maps to the AM. The damage is graded monotonically, most severe at the left extreme of the topographic map and least severe at the right (assuming a right hemisphere lesion, as we do throughout this article). Figure 6 depicts the damaged connections into the AM. The graded damage is important, because it results in a relative preference for the right; complete destruction of the connections in the left field and fully intact connections in the right field would yield a qualitatively different sort of behavior. The graded damage we propose is motivated by Kinsbourne's (1987, 1993) orientational bias account of neglect.

The damage is described in functional terms—i.e., how the damage affects the operation of the model. The model is neutral with regard to the neurobiological basis of this damage—i.e., how a unilateral brain lesion results in damage of this functional form. Additional assumptions will be required to specify the model at a neurobiological level. Other psychological-level theories of neglect face the same challenge. The neurobiological-level theory of Pouget and Sejnowski (1997) suggests that the gradient of attention following a unilateral brain lesion arises by way of gradients of representation in intact parietal cortex: the left hemisphere has a weak representation of the left side of space and a strong representation of the right side, and the right hemisphere is the mirror opposite. Consequently, damage to the right hemisphere leaves only the left hemisphere representation, which has a weak-to-strong gradient of representation from left-to-right. Note that even this neurobiological model makes an assumption of a gradient, although it is a gradient in the intact model, not a gradient resulting from damage. Such a gradient could easily be incorporated into the AM by splitting the model into two copies localized in the left and right parietal cortices. Each copy operates autonomously with weak cross-connections keep the two networks synchronized. The input to the left-hemisphere AM from the primitive feature maps would be weak-to-strong from left-to-right. The input to the right-hemisphere AM would be reversed: strong-to-weak from left-to-right. Damage to the right hemisphere would leave only the left-hemisphere AM, which has the input gradient we have proposed to model neglect. A more elegant account may be feasible: Shillcock and Cairns (1995) present a simple connectionist model in which a gradient of damage emerges from a hemispheric model in which no representational gradients are built in.

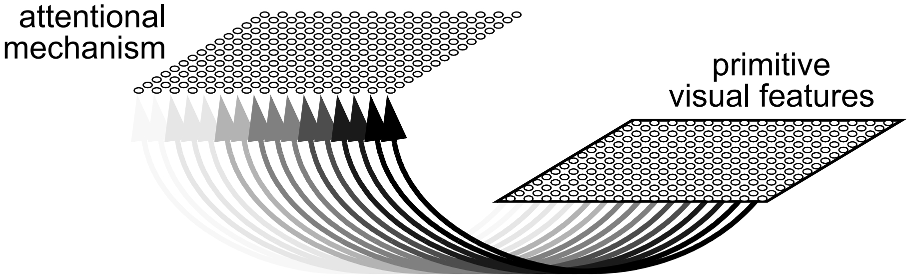


FIGURE 6. A sketch of the AM and some of its inputs from the primitive feature maps. Each feature detector connects to the homologous unit in the AM. In neglect, graded damage to these connections is hypothesized, resulting in feature detectors to be less effective in activating the AM. The damage is depicted by the fainter connections toward the left side of the field.

Our proposal for a graded lesion of the inputs to the AM can be contrasted with two alternatives. First, one might damage the visual recognition network itself. However, this would lead to blindness, and is inconsistent with the view of neglect as an attentional phenomenon and with the neuroanatomical lesion sites that give rise to neglect. Second, one might lesion the AM directly, either changing the activation dynamics or connectivity of the units such that damaged units integrated activity more slowly or had a weakened influence on the activity of other units. We conjecture that these types of lesions would yield a behavioral effect similar to the proposed lesion for the simulation studies reported in this article.

The damage depicted in Figure 6 affects the probability that primitive visual features are detected by the AM. To the extent that features in a given location fail to trigger attention, the AM will fail to focus attention at that location. Thus, the deficit is not "perceptual," in the sense that if somehow attention can be mustered, features will be analyzed normally by the recognition network. For example, the lesioned AM shows extinction: when a single stimulus is presented, it will be attended and recognized, but when two stimuli are presented side by side, the left stimulus is suppressed due to the simultaneous presence of the right stimulus (Mozer & Behrmann, 1992).

The nature of the attentional deficit is specified via a function relating the horizontal position of a feature on the topographic map to the probability that the feature will be transmitted to the corresponding location of the AM (Figure 7). The function is piecewise linear with a flat segment, followed by a segment with positive slope, followed by another flat segment. The left and right extremes of the curve represent the left and right edges of the topographic map, respectively. The probability that the AM will register a feature is low in the left field, and is monotonically nondecreasing further to the right.

The function is characterized by four parameters: (1) the minimum transmission probability (*anchor probability*), (2) the horizontal position in the topographic map at which the probability begins to rise (*anchor position*), (3) the slope of the rising segment (*gradient*), and (4) the probability of feature transmission on the right extreme of the topographic map (*saturation probability*). This parameterization allows a variety of transmission functions, including forms corresponding to normals (e.g., a minimum probability close to 1 and a gradient of 0), and homogeneous slope across the entire field (e.g., a shallow gradient and a saturation position at the far right edge), and a sharp discontinuity at the hemifield crossing (a very steep gradient and a saturation position just to the right of center). Presumably the exact nature of the function varies from patient to patient. Regardless of the specific form of damage, we emphasize that the damage is to a viewer-centered representation of space.

General simulation methodology

The AM as described is identical to the model used in our earlier simulation studies of neglect. However, the nature of the model's input was changed in one minor respect. In the earlier simulation studies, when a stimulus display was presented to the model, the exogenous input to the AM was determined probabilistically based on the stimulus display and the transmission-probability function (Figure 7). The exogenous input then remained constant as the AM settled. However, when the display itself is not static—as is the case in one simulation reported here—the exogenous input cannot be static. Consequently, in these simulations, we resampled the stochastic exogenous input at each time step of the simulation. This resampling had no systematic effect for static displays, but allowed us to simulate the AM for dynamic displays.

The AM has three parameters: μ, θ, and γ. In our earlier simulations utilizing the AM, μ was fixed at 1/8 and θ at 1/2. These values were used in the present research as well. The third parameter, γ, is dependent on the amount of activity in the stimulus display. In our earlier simulations, we devised a formula for setting γ based on the total exogenous input to the AM, exo_0, and a metaparameter γ' that modulates the fraction of the locations that provide exogenous input to the AM that should be selected by the AM:

$$\gamma = \min\left(0.75, \max\left(1.0, \frac{exo_0}{\gamma'}\right)\right) .$$

γ' was originally conceived as task- and stimulus-independent, and earlier simulations of the AM used a constant γ'. However, we discovered in the present work—which covers a much wider variety of stimulus displays than the previous simulations—that γ' needed to be set for each experimental task. It was the only free parameter of the model, and roughly corresponds to the degree of selectivity required to perform the task. Consequently, the setting of γ' depends on the density and distribution of features in a display. The adjustment was primarily performed to obtain sensible behavior from the AM, not to fit simulation data to human data. The model's behavior was qualitatively robust to the choice of γ'. However, if γ' was too large, the AM would fail to be selective, and if γ' was too small, all activity in

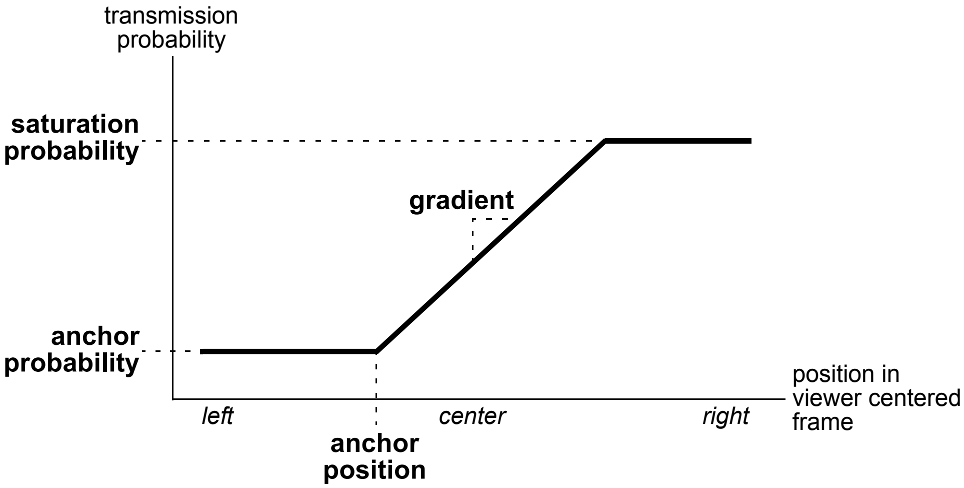


FIGURE 7. The transmission probability curve representing the damage to the model's attentional system. This function relates the position of a feature in the viewer-centered frame to the probability that the feature will be detected by the corresponding unit of the AM. The function is for a left neglect patient; the probability that the AM will register a feature is low in the left field, and is monotonically nondecreasing further to the right.

the AM would die out. We set γ' to 240 for simulations of Behrmann and Tipper (1994; Tipper & Behrmann, 1996) and Driver and Halligan (1991), 220 for simulations of Behrmann and Tipper (1999), 110 for simulations of Arguin and Bub, 650 for simulations of Pavlovskaya, Glass, Soroker, Blum, and Groswasser, and 70 for simulations of Driver, Baylis, Goodrich, and Rafal. From a rational analysis perspective (Anderson, 1990), one might think of γ' as an adjustable parameter of the cognitive architecture that is tuned to optimize performance.

In Mozer, Halligan, and Marshall (1997), we simulated a range of lesions by varying the four parameters in the transmission-probability curve. For the present work, however, we chose a single lesion profile that had produced typical results in the earlier work. This profile had an anchor probability of .3 and a saturation probability at .9. The anchor position was at the left edge of the topographic map, and the gradient was chosen such that saturation was reached 5/6 of the way to the right edge of the topographic map. For the unlesioned model, the anchor and saturation probabilities were both .9. All simulations used a topographic map of dimensions 36×36, except for the Arguin and Bub simulation which required a 10×61 topographic map to allow for variation in the horizontal position of the stimuli.[3]

To simulate an experimental task, the experimental stimuli must be mapped to a pattern of exogenous input to the AM. As we have done in earlier work, the mapping was accomplished by laying a silhouette of the stimulus over the topographic map, and setting the exogenous input at all locations covered by the silhouette to .10 except along the stimulus contour, where the exogenous input is raised to .20 to reflect the contrast along the border. Further, as in the past, we assume a slight amount of blurring of the exogenous input: Each stimulus location provided input not only to the corresponding location of the AM but also the immediately adjacent locations, with a relative strength of 2%. This very slight spread is unlikely to affect processing, but we have preserved it to maintain consistency with the original simulations.

The experimental tasks simulated have as their dependent variable the response time to detect or identify a target. Rather than running the full MORSEL model and using the object-recognition network to determine detection or identification responses, we make a simple readout assumption that allows us to perform a simulation using only the AM. The assumption is that *the reaction time to detect or identify a target is inversely proportional to the attentional activation in locations that correspond to the target*. This assumption is justified by earlier simulations of MORSEL (Mozer, 1991), in which output activity of the recognition network was found to be monotonically related to the allocation of attention to locations of a target. Because the propagation of activity in MORSEL is temporally extended, we use not the instantaneous activation of the AM, but rather *the mean activity of the AM over the twenty iterations following target onset.* The results described in the following sections are not sensitive to the specific read out assumptions; results are qualitatively similar if the mean activity is computed over ten or forty iterations instead of twenty, or if the mean activity is mapped to response time by any monotonic transformation. Because trials will vary due to random effects of the transmission probability curve, we average activation across multiple trials in each experimental condition.

Simulations

Behrmann and Tipper (1994; Tipper and Behrmann, 1996)

When an experimental stimulus is presented upright and centered on the fixation point, viewer-centered and object-centered reference frames are confounded. To dissociate the two frames, Behrmann and Tipper (1994) rotated a display containing a *barbell*—two disks, one colored red and the other blue, connected by a solid bar. The barbell first appeared with, say, the red disk on the left and the blue disk on the right. It remained stationary for one second, allowing subjects to establish an object-based frame of reference. In the *moving* condition, the barbell then rotated 180° (Figure 8a) such that the blue disk ended up on the left and the red disk on the right—the two disks had exchange places (Figure 8b). Following the rotation, the red disk appears on the left with respect to the object-based frame, but on the right with respect to the viewer-based frame. The subjects' task was to detect a target appearing on either the red or the blue disk. A *static* condition, in which the barbell did not rotate, was used as a baseline (Figure 8b). Left-neglect subjects showed facilitation for targets appearing on the blue disk in the moving condition relative to the static condition, and showed inhibition for targets appearing on the red disk. Essentially, the laterality of neglect reversed with reversal of the barbell. Results were therefore consistent with object-based, not viewer-based, neglect.

---

[3] We conceived of the larger topographic map as representing the same visual field region as the smaller map, but at a higher resolution.

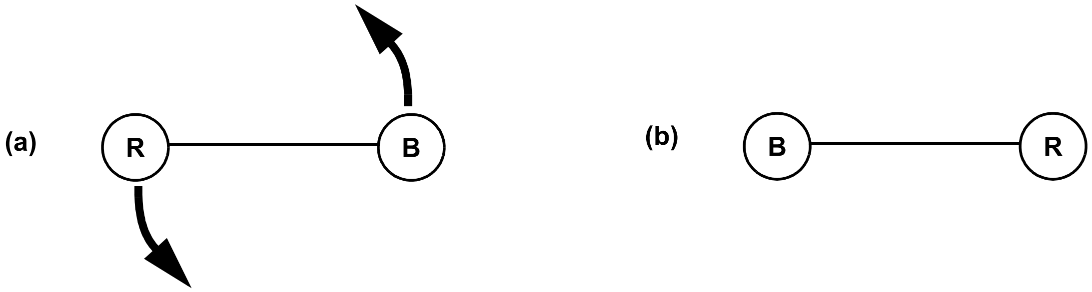


FIGURE 8. Barbell stimulus used in the Behrmann and Tipper experiment. The disk labeled “R” is colored red, the disk labeled “B” blue. In the moving condition, the initial display (panel a) was rotated 180°, resulting in the left and right disks exchanging places (panel b). In the static condition, no rotation occurred (panel b).

Tipper and Behrmann (1996) ruled out an explanation for this phenomenon in terms of overt tracking by eye movements. They also showed that the phenomenon appeared to depend on the disks being encoded as one object: in contrast to the condition depicted in Figure 8 in which the two disks are *connected*, when the bar between the disks is removed—the *disconnected* condition—the reversal of neglect no longer occurred when the disks rotate. This finding is what one would expect if neglect occurred in an object-based frame, because rotation of the display no longer corresponds to rotation of a single object.

To simulate the moving condition in the AM, a horizontal barbell was presented for 50 iterations, and then rotated 180° over the next 400 iterations, and was followed immediately by the target. To simulate the static condition, the horizontal barbell was presented for 200 iterations, followed immediately by the target. Encoding the rotating stimulus in a discrete array of cells is complicated due to quantization effects. Rather than attempting to hand-design an exogenous input patterns for the barbell at every angle θ, the exogenous input was automatically generated from the exogenous input for the horizontal barbell stimulus as follows. For each location (x, y), a new coordinate (x', y') was computed by θ° rotation. Because x' and y' are in general noninteger, the exogenous input at (x, y) could not be copied to (x', y') directly. Rather than rounding x' and y' to the nearest integers, the exogenous input at (x, y) was then split up according to the distance of (x', y') to the four integer grid locations surrounding it. This procedure minimized quantization effects that arose from the coarse representation of the topographic map.

As we explained earlier, we assume that the attentional activity in a region of space is related to the speed and accuracy of information processing in that region. In the Behrmann and Tipper experiment, the critical regions are those of the two disks. Read out from the model was performed by calculating the mean attentional activity directed toward each disk, averaged over all locations containing features of the disk and over the twenty iterations following the trial and over 200 trials, which will be referred to as the *read-out activity*. Greater read-out activity for a disk indicates a shorter response time to the target appearing in that disk.

Figure 9 shows a trial of the unlesioned AM in the moving barbell condition. The unlesioned model has a uniform transmission probability of .9 across the field, producing occasional missing features in the exogenous input, as can be seen in the upper-left panel. As the Figure shows, attention is rapidly deployed to the entire barbell, and remains with the barbell as it rotates. This result is not altogether trivial, as the model had never been tested on dynamic stimuli. The read-out activity was .99 for both the left and right disks. (With regard to the disks and targets, “left” and “right” will refer to viewer-centered locations.) The attentional state before rotation begins, at iteration 50, gives a good indication of the read-out activity in the static condition, which was also .99 for left and right disks. Thus, the unlesioned AM shows no difference in target detection time among conditions—moving versus static, left versus right target, and connected versus disconnected disks.

The lesioned AM shows quite different behavior (Figure 10). A relative degradation to the exogenous input on the left side of the barbell can be observed due to the transmission probability curve, causing the right half of the barbell to be selected initially. As the barbell begins to rotate, the focus of attention narrows further to just the disk, because the disk provides the greatest amount of exogenous input. As rotation continues, attentional activity lags slightly behind the exogenous input, due to limits on the time course of activation transmission, but catches up when the rotation is completed. Given the final distribution of attention in the moving condition, the model will be faster to respond to a target on the left than on the right. This reversal does not occur in the static condition, as suggested by the AM state at iteration 50. The trial depicted in Figure 10 is representative; it is consistent with the more quantitative measure of read-out activity (Table 1, connected condition) which indicates greater activity for the left disk in the moving versus the static condition, and less activity for the right disk.

When the disks are disconnected, attention jumps from the disk that started off on the left to the disk that ends up on the left (Figure 11). After the disks cross the midline, the disk rotating into the right field begins to receive more support from the exogenous input than the disk rotating into the left field. Eventually this exogenous support is sufficient to activate the right disk, and competition acts to suppress the left disk. This pattern is observed reliably, as indicated by the measure of read-out activity (Table 1, disconnected condition). The read-out activity shows nearly full activity to the right disk and none to the left disk, and no difference between moving and static conditions.

To summarize, the AM simulation replicates the primary findings of Behrmann and Tipper (1994; Tipper & Behrmann, 1996):

(1) For normals, no reliable differences are obtained across conditions.

(2) For patients shown connected disks, left-sided facilitation and right-sided inhibition is obtained in the moving condition relative to the static.

(3) For patients shown disconnected disks, left-sided facilitation and right-sided inhibition are not observed.

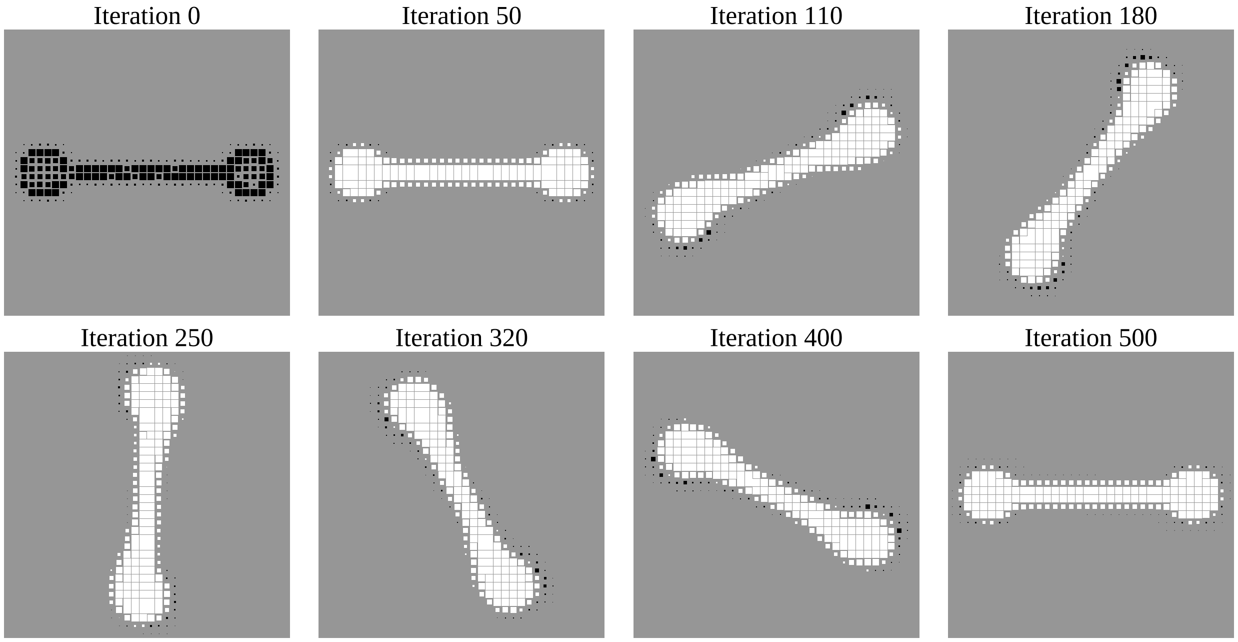


FIGURE 9. One trial of the unlesioned model on the Behrmann and Tipper rotating-barbell stimulus. Attentional activation (white squares) follows the exogenous input (black squares) as the barbell rotates.

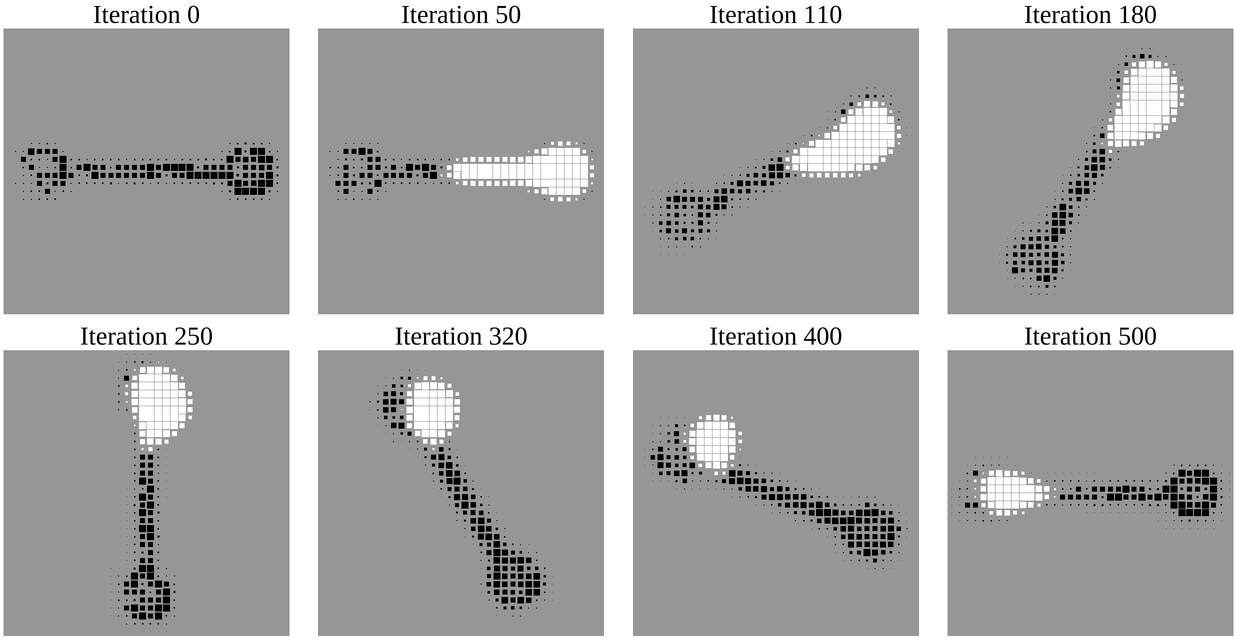


FIGURE 10. One trial of the lesioned model on the Behrmann and Tipper (1994) rotating-barbell stimulus.

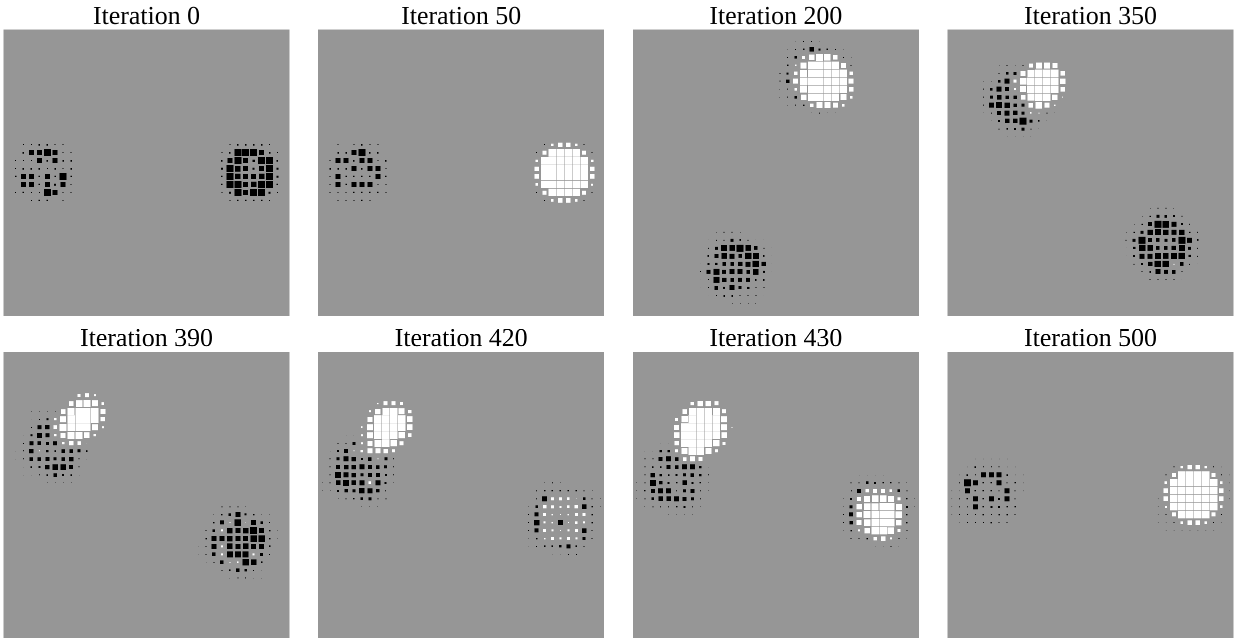


FIGURE 11. One trial of the lesioned model on the Tipper and Behrmann (1996) rotating disconnected disks.

TABLE 1. Read-out activity from the lesioned AM for the left and right disks in the experimental conditions of Tipper and Behrmann (1996)

condition		left disk	right disk
connected	moving	0.22	0.04
	static	0.00	0.99
disconnected	moving	0.00	0.93
	static	0.00	0.99

(4) For patients, there is a main effect of target side: left is slower than right.

The model's ability to replicate the pattern of data was not obvious without running a simulation, and in fact, its behavior for disconnected disks was unexpected. Nonetheless, the results emerged reliably from the simulation. In a situation such as this, our only recourse is to experiment with the model and determine what factors influence its behavior, with the goal of eventually extracting an intuitive explanation for its success.

Many factors did not affect the model's qualitative performance, suggesting that the result is robust. The specific design of the stimuli was unimportant; qualitative performance was robust to the shape of the disks, the pattern of exogenous input corresponding to the disks, or the size of the disks relative to the thickness of the bar (Figure 12). The connectedness of the two disks was not even essential to achieve the reversal effect (lower right panel of Figure 12), an observation to which we will return shortly. The model was robust to other factors, including: alternative parameters for the transmission probability curve (Figure 7) as long as a gradient was present, the rate of rotation of the stimulus, and the read-out formula. In fact, the reversal effect reported for rotating connected disks could be made even larger by increasing the size of the disks, reinforcing the exogenous input to the borders of the stimulus, increasing the rotation time, and/or reading out the asymptotic activity of the AM.

One factor that necessarily influences the model's qualitative performance is the strength of exogenous input corresponding to the bar. If the bar triggers a very weak exogenous input— for example, when the input corresponding to the bar is one pixel thick and those pixels are attenuated in strength relative to pixels of the disks—the exogenous input pattern for connected disks becomes quite like the pattern for disconnected disks, and the model treats the connected-disk condition like the disconnected-disk condition. What this circumstance corresponds to in terms of visual stimuli is unclear, because the exogenous input to the AM reflects the net activity of feature detectors, not raw pixels in the image. Even a thin, faint line in the image could trigger significant activity of edge detectors, resulting in a substantial exogenous input to the AM. (Indeed, since cells in visual cortex respond more strongly to edges than to solid regions, there is no reason to believe that bar thickness has a significant effect on the pattern of neural activity.)

To understand the simulation results, consider first the moving connected-disk trials. The model appears to track the right disk into the left field. Because attentional activity in the model corresponds to covert attention,[4] this tracking is not necessarily overt and is therefore consistent with the finding of Tipper and Behrmann (1996) that eye movements are not critical to the phenomenon. Tracking occurs because the attentional state has hysteresis (e.g., Sperling, 1970): the state at some iteration is a function of both the exogenous input and the state at the previous iteration. Attention would not ordinarily be drawn to a disk on the left given a competing disk on the right because the exogenous input to the left disk is weaker. Nonetheless, if attention is already focused on the disk on the left, even a weak exogenous input may be sufficient to maintain attention on the disk. Returning to the rules of activation of the model described earlier, the disk that has moved into the left field has support via the bias and cooperation rules, whereas the disk that has moved into the right field has support only via the bias rule.

However, the winner is not determined simply by the fact that one disk has the support of the cooperation rule and the other does not. Key to the model's behavior is the total *quantitative* support provided to each of the disks. If the total support is greater for the right disk, then attention will flip to the right. This flipping occurs on disconnected-disk trials. In the model, connectedness per se is not critical to the flipping of attention: Attention does not flip for a display in which the bar is broken in the center (Figure 12, lower right panel), which disrupts connectedness. Instead, the *neck* of the barbell—the region where the disk makes contact with the bar—seems the critical component. The neck provides a region of exogenous input adjacent to the disk, and by the cooperation rule, therefore provides a neighborhood that supports attentional activity. Figure 10 clearly shows that activation is centered on the neck as the disk rotates into the left field. Without the neck to "hook" activity in place, activity drops to the point that the left disk cannot fend off attack from the right disk. Although this account is not entirely satisfactory, in that we have not explained the phenomena in linguistically simple, qualitative terms, it is sometimes the best one can hope for in characterizing the behavior of a complex, dynamical system such as the AM. The explanation we have given for the AM's performance leads to empirical predictions, in particular concerning a stimulus display such as that in the lower right panel of Figure 12. We return to the issue of predictions of the model in the General Discussion.

The Behrmann and Tipper data seems strongly consistent with the hypothesis that neglect operates in object-based coordinates. The AM, however, provides an alternative explanation, because it has no object-based frame of reference, yet it can account for the data. The AM's account involves covert attentional tracking. Without simula-

---

[4] Attentional activity corresponds to covert attention by virtue of the fact that we do not model eye movements, and the stimuli do not change position in the visual field as attention shifts.

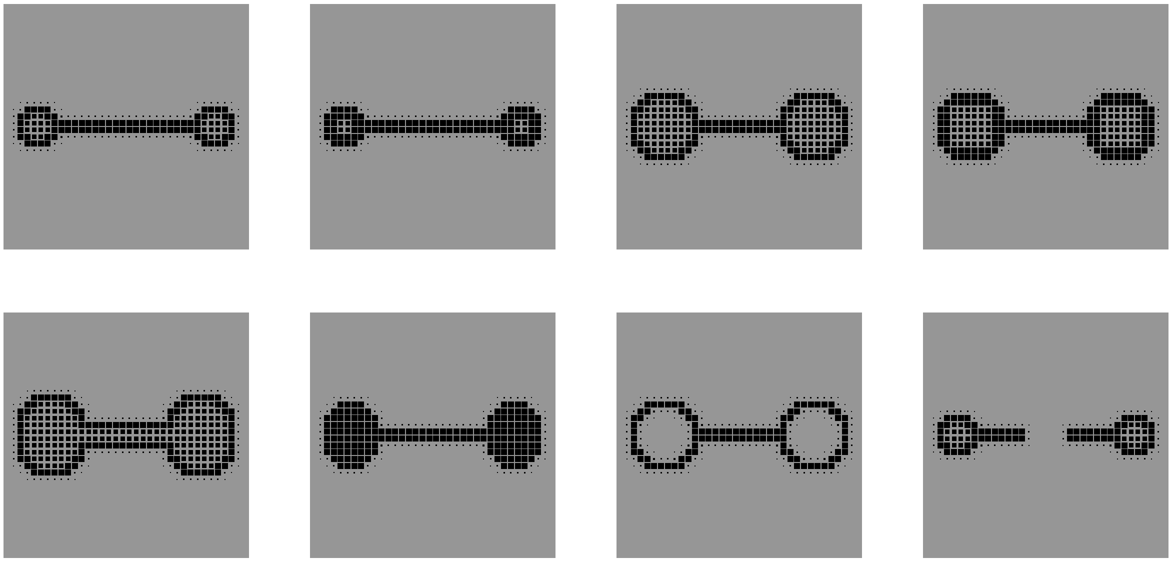


FIGURE 12. Exogenous input activity patterns for eight different barbell stimuli. Rotation of all eight versions achieved a reversal of neglect, suggesting that the model is robust to the exact shape of the barbell. The pattern in the upper left panel is the one used for all simulations reported in the paper, and corresponds most closely to the experimental stimuli used in neuropsychological studies in terms of the size of the disks relative to the length of the bar.

tions, the covert-tracking account is not compelling, because it would not appear to explain the lack of neglect reversal for disconnected displays. However, despite the absence of object-based representations, the AM does show a distinction between connected (single object) and disconnected (multiple object) displays, and hence increases the plausibility of the covert-tracking account.

Behrmann and Tipper (1999)

Recently, Behrmann and Tipper (1999) have explored an intriguing variation of the rotating-barbell experiment in which the display also contains two elements—squares—which remain stationary during the trial (Figure 13). Subjects were asked to detect a target that could appear either on one of the disks or on one of the squares. As in the earlier studies, facilitation is observed for targets appearing on the left (blue) disk in the moving condition relative to the static condition, and inhibition is observed for targets appearing on the right (red) disk, consistent with neglect in the object-based frame of the barbell. Simultaneously, however, neglect is observed in the viewer-based frame for the squares: target detection in the left square is slower than in the right square. The finding of neglect in both viewer- and object-based reference frames suggests that attention can select and access information encoded with respect to multiple reference frames.

To simulate this experiment, several changes were made to the previous simulation. First, the barbell rotated from 30° to 150°, instead of 0° to 180°. Maintaining the same rotation rate as in the previous simulation, the rotation took 267 iterations. Second, stationary squares were added to the display. Third, read-out activity was calculated at the square locations as well as the disk locations. Figure 14 presents a single trial of the lesioned AM. The model's behavior is somewhat surprising: Initially, attention is drawn to the right side of the display, which includes the right disk and right square. As the barbell begins to rotate, attention is stretched to span the disk and the square, but when the disk and square separate, the attentional blob connecting them is broken into two blobs. One might expect the smaller blob to be suppressed due to competition between the blobs, but the competition is weak, for the following reason. The competition rule causes a location to be inhibited to the extent that its activity is below the average activity of all active locations. Because locations in both blobs have comparable activity—in fact, they are near asymptote at the point when the blobs split—neither blob is significantly inhibited.

Quantitative measures of read-out activity (Table 2) are consistent with the example presented in Figure 14 and with the results of Behrmann and Tipper (1999). Facilitation—increased activity at the disk location—is observed for the left disk in the moving relative to the static condition. Inhibition—decreased activity at the disk location—is observed for the right disk. And in both moving and static conditions, facilitation is observed for the right square relative to the left. Thus, the AM can account for what appears to be neglect occurring simultaneously in multiple reference frames, although it encodes visual information in only a single reference frame, viewer based.

In the experiment just described, the target appeared on the squares half the time and on the disks half the time. In a second experiment, Behrmann and Tipper varied the target contingencies, such that for some subjects, 80% of targets were on the squares, and for other subjects, 80% of the targets were on the disks. They observed an accentuation of neglect for the shapes—disks or squares—that were probed more frequently, and interpreted this in terms of task demands modulating attention, and hence neglect, within location-based or object-based reference frames.

These results can also be accommodated within the framework of the AM. Suppose that the target contingencies modulate the AM's likelihood of attending the squares: The more frequently targets appear on the squares, the more likely locations of the squares are to draw attention. Because features are transmitted to the AM in an all-or-none fashion, a sensible way of increasing or decreasing the likelihood of attention is to modulate the transmission probability of features of the squares, much the same effect that we hypothesize neglect has. In the target-more-likely-on-squares condition, we multiply the transmission probability of features at the locations of the squares by 1.2; in the target-more-likely-on-disks condition, we multiply the transmission probability of features at the locations of the squares by 0.8.[5] Thus, neglect and target contingencies affect transmission probabilities independently. Results for these two conditions, shown in Table 3, clearly indicate that "object-based" neglect of the disks is accentuated when the target is more likely to appear on the disks, and "location-based" neglect of the squares is accentuated when the

---

[5] As we mentioned earlier, the AM receives not only exogenous input from the visual field but also endogenous input from higher brain centers. The endogenous input allows the AM to be guided in a task-dependent manner. Adapting to target contingencies is exactly the sort of guidance that higher centers could provide. In this experiment, because subjects fixated at the center of the display at the onset of each trial, object locations map to retinal locations, and top-down guidance can be provided by specifying likely retinal locations where targets would appear. In the case of displays that had more variability, it is conceivable that cognitive processing could map task constraints into biases on retinal locations via simple visual routines (Ullman, 1984).

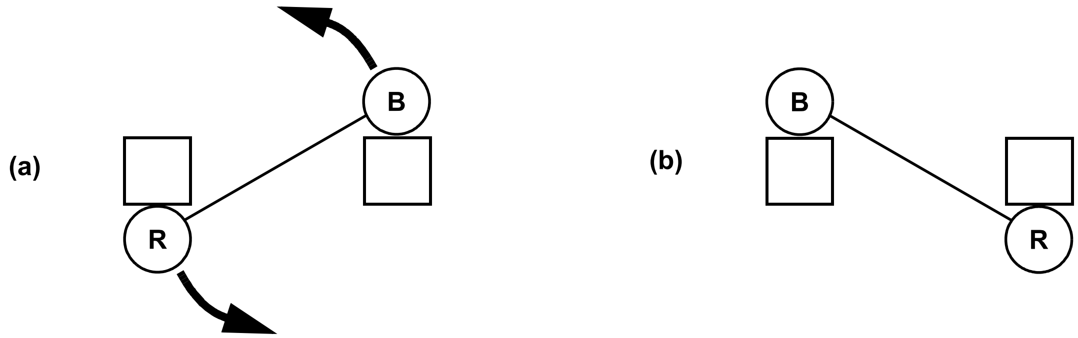


FIGURE 13. The multiple-object display studied by Behrmann and Tipper (1999). In the moving condition, the initial display (panel a) consists of two stationary squares and a barbell, which—as in the earlier studies—rotates such that its two disks exchange horizontal positions (panel b). Subjects were asked to detect a target that could appear on either disk or either square. In the static condition, no rotation occurred (panel b).

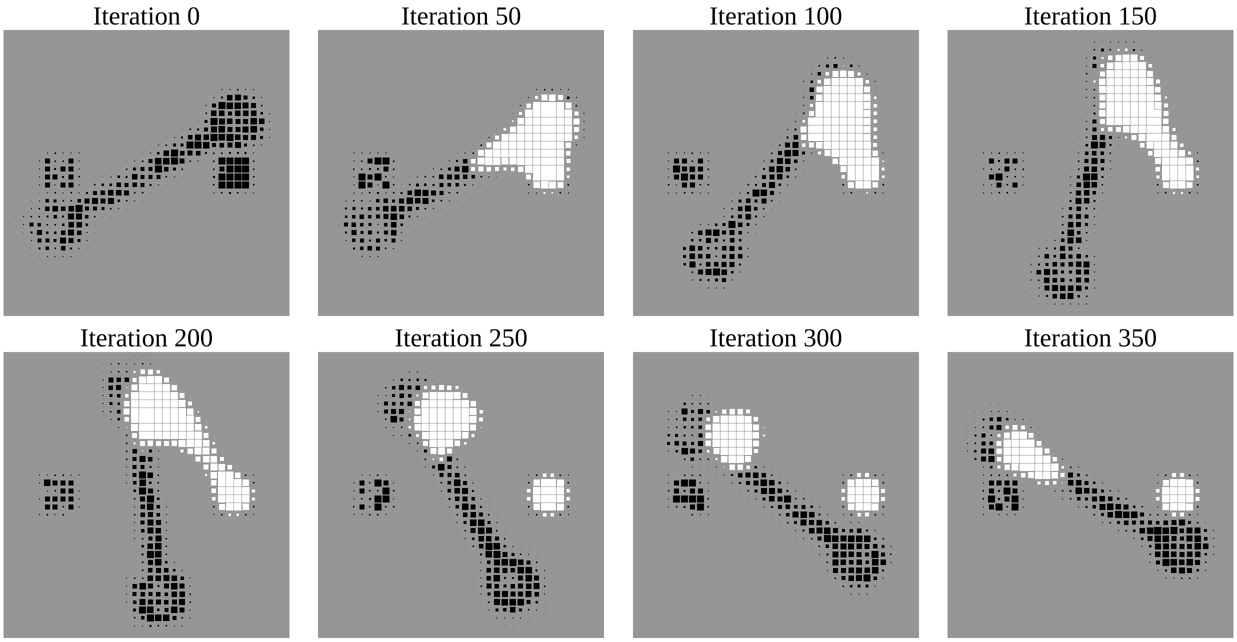


FIGURE 14. One trial of the lesioned model on the Behrmann and Tipper (1999) barbell-square experiment.

TABLE 2. Read-out activity from the lesioned AM for the experiment of Behrmann and Tipper (1999)

condition	left disk	right disk	left square	right square
moving	.21	.04	.00	.99
static	.00	.90	.00	.91

TABLE 3. Read-out activity from the lesioned AM for the experiment of Behrmann and Tipper (1999) under different target contingencies

condition		left disk	right disk	left square	right square
target more likely on disks	moving	.21	.02	.00	.78
	static	.00	.91	.00	.73
target more likely on squares	moving	.16	.04	.00	.99
	static	.00	.72	.00	.92

target is more likely to appear on the squares. Again, the AM suggests that the findings of Behrmann and Tipper can be accounted for without invoking object-based reference frames.

Although the AM replicates the pattern of data of Behrmann and Tipper (1999), the account is intuitively unsatisfying in one regard. As Figure 14 shows, the AM's initial blob of activation encompasses both the right disk and the right square. The activation rules giving rise to this blob depend on the spatial proximity of the two shapes, but ignore the fact that the two shapes are distinct objects. One might view this failure to prevent the spread of activation across object boundaries as a deficiency of the AM. A simple repair might be made by increasing the resolution of the topographic map, resulting in additional locations separating the two objects, and less likelihood of activation spreading from one object to the other. However, the problem seems more fundamental: the AM has no embodiment of Gestalt grouping principles such as good continuation, closure, or similarity, but only the principle of proximity. Such principles can account for object-based effects in normals without requiring object-based representations (e.g., Behrmann, Zemel, & Mozer, 1998; Mozer, 1999; Mozer, Zemel, Behrmann, & Williams, 1992; Vecera & Farah, 1994). The AM might incorporate grouping principles by using the grouping principles to specify which pairs of units cooperate and compete (Grossberg, Mingolla, & Ross, 1994; Mozer & Sitton, 1998). For example, the current implementation of the AM incorporates the principle of proximity via mutual excitation between adjacent units in the topographic map. The AM might be extended to include cooperation between, say, two locations containing the same features, or between two locations whose features form a smooth contour. Essentially, we are proposing to modulate the connectivity within the AM based on how features in the display are grouped according to the Gestalt principles.

We digressed in order to explain how the AM might be made more sensitive to object boundaries. Assuming such a modification, the AM is still able to explain the Behrmann and Tipper (1999) results. The modified AM, when lesioned, should select *either* the right side of the barbell *or* the right square. If the barbell is selected, covert attentional tracking should still lead to "object-based" neglect for the barbell. If the right square is selected, "viewer-based" neglect will be obtained for the squares. The magnitude of object-based and viewer-based effects will depend on how often one object or the other is selected, which in turn should be influenced by target contingencies.

To summarize, the AM—whether in its present form or modified to incorporate Gestalt grouping principles—is able to explain the simultaneous observation of object-based and location-based effects of neglect in the experimental paradigm studied by Behrmann and Tipper (1999).

Pavlovskaya, Glass, Soroker, Blum, and Groswasser (1997)

If object recognition requires that the visual input be encoded in object-based coordinates, one might expect recognition accuracy to be impacted if the object-based frame is imposed incorrectly. Pavlovskaya et al. (1997) studied a task that might seem to affect the establishment of an object-based frame. They supposed that the *luminance centroid* of a visual shape—the center of mass of its light distribution—serves as a natural origin for an object-based frame. They further supposed that a cue prior to the onset of the shape might serve to bias or shift the origin, which would hamper recognition. Stimuli were brief, masked, letter-like stimuli preceded by a location cue. The cue could be at the luminance centroid, or to its left or right (Figure 15). Subjects were instructed to report the stimulus identity. Normal subjects perform best if the cue is at the luminance centroid (Figure 16, upper left graph), consistent with the hypothesized role of the cue in biasing the establishment of an object-based frame. The neglect data (Figure 16, upper right graph) is also consistent with the hypothesis: one might suppose that neglect causes a rightward bias in specifying the origin, thereby impeding recognition, but that a left cue offsets this bias.

However, the data does not demand an explanation involving object-based frames, or an effect of cues and neglect in establishing an object-based frame. Figure 16 (bottom row) shows a simulation of this experiment using the AM. In the simulation, the cue was presented for 50 iterations, followed by the stimulus for 20 iterations. Stimuli tested consisted of 100 trials of each of four different shapes studied by Pavlovskaya et al.—an H, a left-pointing T, a right-pointing T, and a horizontal bar. Over the 20 stimulus iterations, the mean attentional activity of locations corresponding to the stimulus was computed. We assume that recognition accuracy is related to the attentional activity that accumulates while the stimulus is present. The attentional activity was computed by first calculating the mean activity of stimulus locations in each column, and then calculating the mean across columns. This measure gave a better sense of how far the attentional activation had spread in a left-right direction than a direct average across locations. However, both measures yielded the same results.

The simulation shows the same qualitative pattern of performance as the human subjects.[6] For the normal model, activation of the stimulus following an LC cue is greater than following a left or right cue (LC versus left: $F(1,396)=23.2$, $p<.001$; LC versus right: $F(1,396)=23.2$, $p<.001$), but left and right cues obtained the same activation

($F(1,396)<1$); for the lesioned model, activation of the stimulus following an LC cue is less than that when following a left cue ($F(1,396)=1758$, $p<.001$) and is greater than that when following a right cue ($F(1,396)=4098$, $p<.001$). One gains an intuition about these results when examining typical activation patterns for the normal and lesioned model, with both left and luminance center cues (Figure 17). When a cue is first presented, the cue location becomes active. To a first approximation, when the shape appears, activation spills out from the cue location to the remaining locations of the shape. The "spilling" is due to the cooperation rule: a location becomes activated more readily if its neighbors are already active. For the normal model, when the cue is at the center of the object, the tide of activation has an equal number of locations to cover to the cue's left and right, allowing for an efficient parallel propagation of activation. When the cue is to the left or right, however, activation has further to travel before it reaches the shape's extremities. This asymmetry results in less attentional activation for left and right cues than for the luminance centroid cue.

For the lesioned model, a different asymmetry comes into play: to a degree, attentional activation must spread from the cue location to other locations of the object (the cooperation rule), but the weaker exogenous input to the left of the cue causes the activation to travel more readily to the right than to the left. Cueing to the left appears to partly compensate for this asymmetry, resulting in the most rapid rise of activation for left cues.

Cueing effects arise in the simulation not because of interference or facilitation in constructing an object-based representation, but because of attentional dynamics in a viewer-based frame. Driver and Pouget (in press) have also argued that the Pavlovskaya et al. data can be explained without recourse to object-based frames. Their argument is completely consistent with the AM's account, although the AM adds to the story by providing a working computational mechanism that removes all doubt about the plausibility of the explanation, by modeling normal as well as patient performance, and by avoiding assumptions specific to the Pavlovskaya et al. paradigm (e.g., one of their accounts assumes that accuracy is dependent on the balance of neural response over left and right sides of a stimulus).

Arguin and Bub (1993)

Several studies have tried to disentangle the contributions of various frames of reference to neglect by manipulating the location of a target in one reference frame while keeping it fixed in another (e.g., Behrmann & Moscovitch, 1994; Calvanio, Petrone, & Levine, 1987; Farah, Brunn, Wong, Wallace, & Carpenter, 1990). Arguin and Bub (1993) performed such a study in which the two frames were viewer based and object based.[7] Subjects were asked to name a target letter presented in a horizontal array containing four elements. The other three elements were filled circles. The target could appear in one of eight positions on the screen, called the *viewer-relative position*. The target could also appear in one of four positions relative to the circles, called the *object-relative position*. Viewer-relative and object-relative positions were varied independently, producing 32 different display configurations (Figure 18). Response time to name the target was measured. Arguin and Bub viewed response time as a "direct measure of allocation of attention across space": the more attention allocated to a position, the faster the response times. This paradigm allows for the comparison of performance across object-relative position when the viewer-relative position is held constant.

Whereas normal subjects showed no effect of object-relative position, neglect patient B.A. showed increasing response time with leftward target displacement in the array. Because an effect of object-relative position was obtained when unconfounded with viewer-relative position, the data were interpreted as supporting the hypothesis that neglect can occur with respect to an object-based frame of reference.

In our simulation of this experiment, each display element was mapped to a 4×3 pattern of exogenous input to the AM, with a one-column gap between display elements. Because the target could appear in eight different viewer-

---

[6] Although Figure 16 does not appear to show a precise quantitative fit to the data, one should not be concerned. First, the mapping from attentional activation to recognition accuracy need not be linear. If the mapping curve has a positive acceleration, small differences for higher mean activities will be amplified and the fit will improve. Second, little or no attempt was made to adjust the stimulus representation and timing of a trial to achieve an exact fit.

[7] Arguin and Bub distinguish object-based frames from *stimulus-based* frames. An object-based frame "depicts the spatial relations between the *parts of a single object*," whereas a stimulus-based frame "represents the relative locations of *spatially distinct stimuli*" (Arguin & Bub, 1993, p. 350; italics in original). We see no clear cut distinction between these two situations. Many objects can be drawn in a way that their parts are not physically connected, e.g., the word DOG. If one accepts a hierarchical organization of objects and their parts, there is no in principle distinction between "objects" and "stimuli" because a "stimulus" at one level of the hierarchy (made up of multiple objects) is an object at the next level up the hierarchy. Until some compelling evidence is presented for a dissociation of object-based and stimulus-based frames of reference, we argue the two terms should be treated as equivalent. Even if the two frames are dissociated, the simulation to be reported is still of value: Arguin and Bub argue for the psychological reality of stimulus-based frames based on their data, but the present simulation replicates the pattern of data without either a stimulus-based or object-based level of representation.

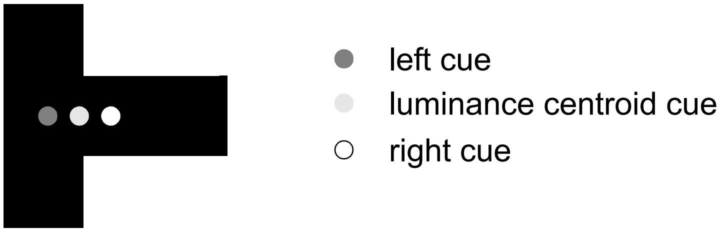


FIGURE 15. Sample shape used in Pavlovskaya et al. (1997), and the locations of the three different cue types.

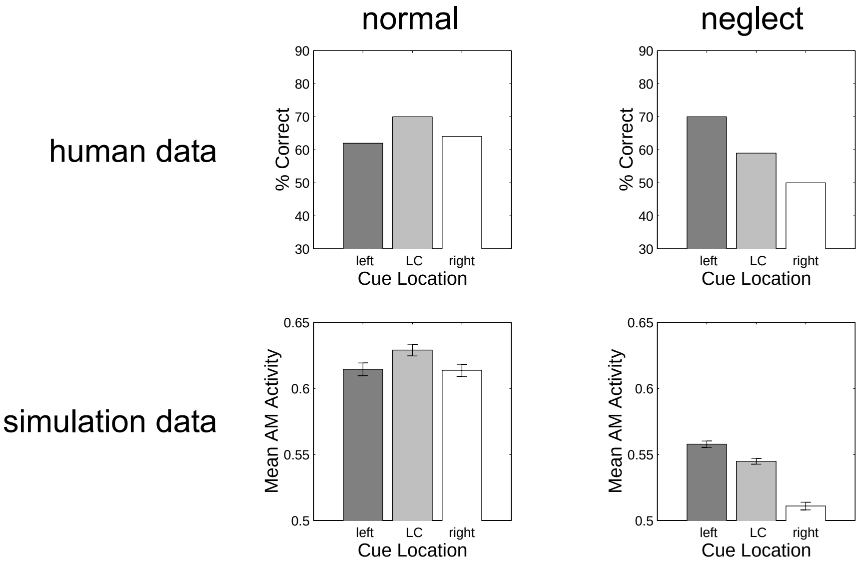


FIGURE 16. Human and simulation data (upper and lower rows) from the Pavlovskaya et al. (1997) study, both for normal adults and neglect patients (left and right columns). The human and simulation data show an excellent qualitative match. The error bars for the simulation indicate one standard error of the mean.

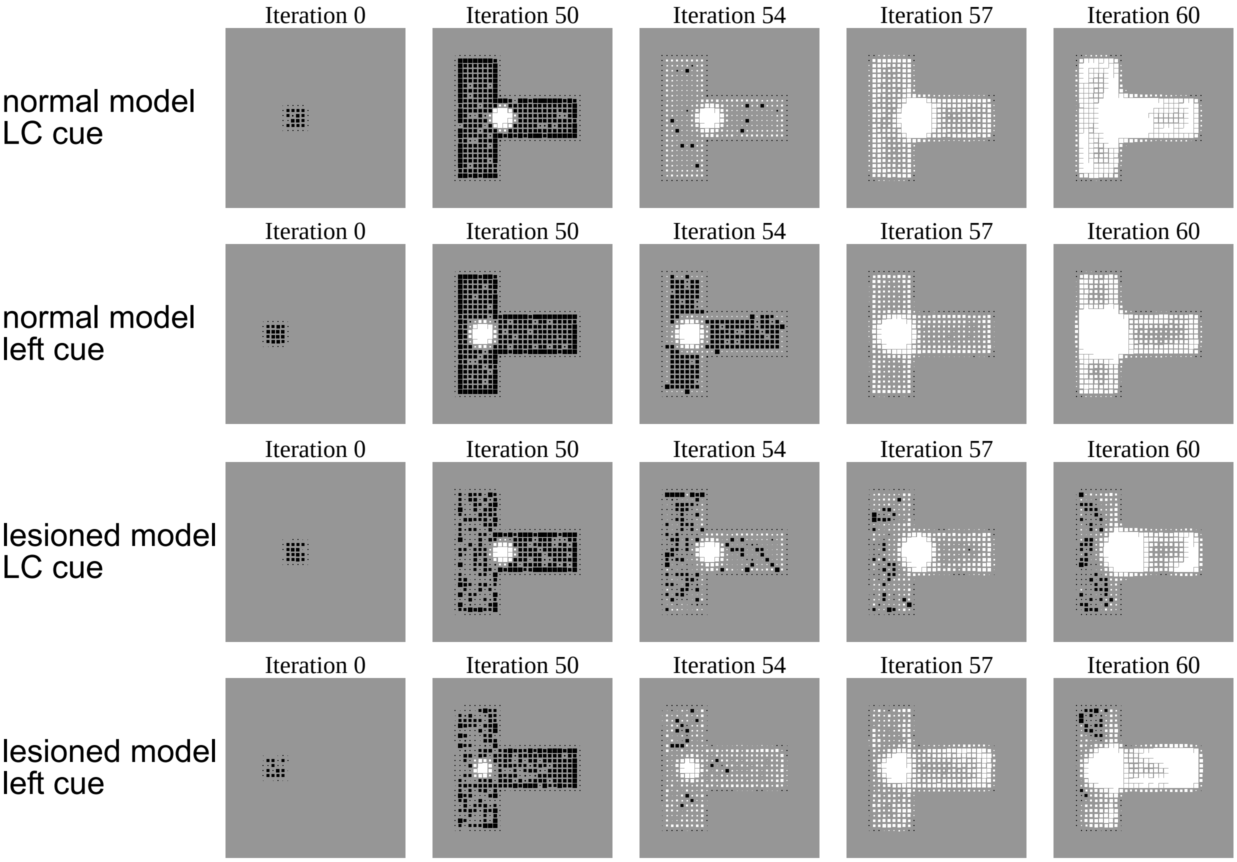


FIGURE 17. Typical activation patterns on trials of the normal and lesioned model, with luminance centroid (LC) and left cues.

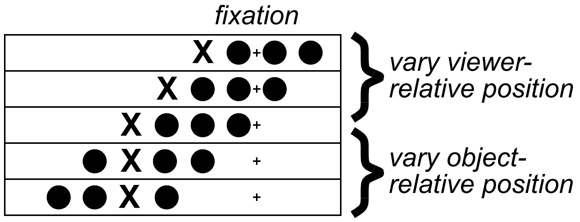


FIGURE 18. Each rectangle, containing a a letter and three filled circles, depicts a possible stimulus display in the Arguin and Bub (1993) study. The fixation point is indicated by the "+". In the top three rectangles, the position of the letter is varied with respect to the viewer-based frame whereas the position with respect to the object-based frame is fixed. In the bottom three rectangles, the position of the letter is varied with respect to the object-based frame, whereas the position with respect to the viewer-based frame is fixed.

centered locations, and it was necessary to allow for three additional display elements to the left and right of the target, the topographic map was designed to accommodate fourteen distinct locations. Figure 19 shows an example of the lesioned model's performance. As the Figure illustrates, the exogenous input tends to be weaker for the elements further to the left. However, because the stimulus display is on the left side of the viewer-centered frame, the exogenous input is degraded even for the rightmost element. Although all four elements capture attention on this trial, attention builds most rapidly for the rightmost elements, suggesting that the read-out activity (the mean attentional activity over the twenty iterations following stimulus onset) should be larger for the rightmost elements, and hence response time should be faster. This observation is confirmed by running fourteen presentations of the complete experimental design (32 trial types).

The simulation data are summarized in Figure 20. For the normal model, no effect is found for either object- or viewer-relative position (object: $F(3,416) < 1$; viewer: $F(7,416) = 1.1, p > .3$). For the lesioned model, however, main effects are obtained for both object- and viewer-relative position (object: $F(3,416) = 80.8$, $p < .001$; viewer: $F(7,416) = 109.9$, $p < .001$), and there is no interaction ($F(21,416) = 1.14$, $p > .3$). These results replicate the main findings of Arguin and Bub. The only significant discrepancy between the human and simulation data is that Arguin and Bub observed effects of retinal eccentricity that influenced performance as a function of viewer-relative position. The model obviously does not address retinal acuity effects, because its visual field is homogeneous. However, such effects could readily be incorporated by, for example, assuming read-out time that increases with distance from fixation.

How can it be that the model has only a viewer-based representation yet its performance is affected by the object-relative position of a target? The story depends on two factors. First, the attentional gradient causes exogenous input from the relative left of a stimulus to be weaker than exogenous input from the relative right (the bias rule), regardless of the absolute position of the stimulus. Second, attentional activation initially depends on the strength of the exogenous input, but then competition comes into play and the relatively stronger locations suppress the relatively weaker locations. Suppression of the leftmost elements by the rightmost elements was also evident in the Behrmann and Tipper simulation (Figure 10), where only the right half of the barbell was attended, even prior to the onset of rotation. The suppression does not occur in the normal model because the exogenous input to the left and right elements is balanced; consequently, their support is roughly equal, and competition does not come into play.

Central to this account is the fact that competition arises only from active locations, making the key consideration in the determination of the activation of an attentional unit its viewer-based horizontal location with respect to other locations where features are present, not its location with respect to the retina as a whole. The idea of relative neglect arising from an attentional gradient has been proposed previously (e.g., Driver, 1999; Kinsbourne, 1993; Pouget & Sejnowski, 1997). However, the AM explains how this idea, in the context of a dynamical system, can give rise to the detailed properties of patient performance.

Mozer and Behrmann (1992) modeled effects of object-relative position in a reading task with the lesioned AM similar to those found in the present simulation. However, the effect of viewer-relative position on performance is somewhat dependent on the task and the specific read out assumptions. For example, when Mozer, Halligan, and Marshall (1997) simulated line bisection using the lesioned AM, they found very minor viewer-relative effects when response formulation was assumed to depend on the asymptotic activity of the AM.

Driver and Halligan (1991)

Like Arguin and Bub, Driver and Halligan (1991) studied a task that involved independently manipulating the location of a target in viewer-based and object-based frames. The task involved detecting whether a pair of nonsense shapes differed in a subtle detail. A sample shape is shown in Figure 21a. The shapes were vertically elongated and relatively bottom heavy, two cues which would suggest an unequivocal principal (up-down) axis. When presented upright and aligned with the patient's sagittal midline, the viewer-centered and object-centered frames are confounded. However, when the shapes are presented at an orientation of 45°, the two frames can be dissociated. In Figure 21b, the detail labeled "L" appears on the object left but the viewer right, and the detail labeled "R" appears on the object right but the viewer left. Driver and Halligan observed an object-based effect: details on the left side of a shape were more often neglected than details on the right side, regardless of whether the detail appeared to the left or right of the sagittal midline, suggesting that neglect can operates in coordinates defined relative to the principal axis of a shape.

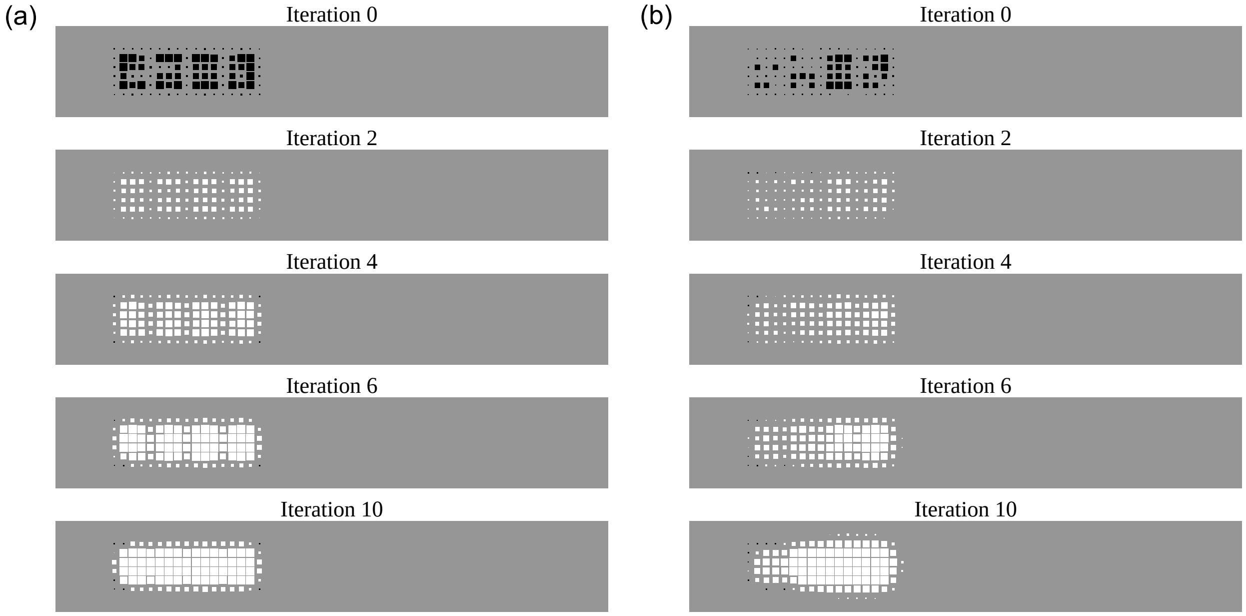


FIGURE 19. Performance of the (a) normal and (b) lesioned model when presented with a sample display from the Arguin and Bub (1993) study.

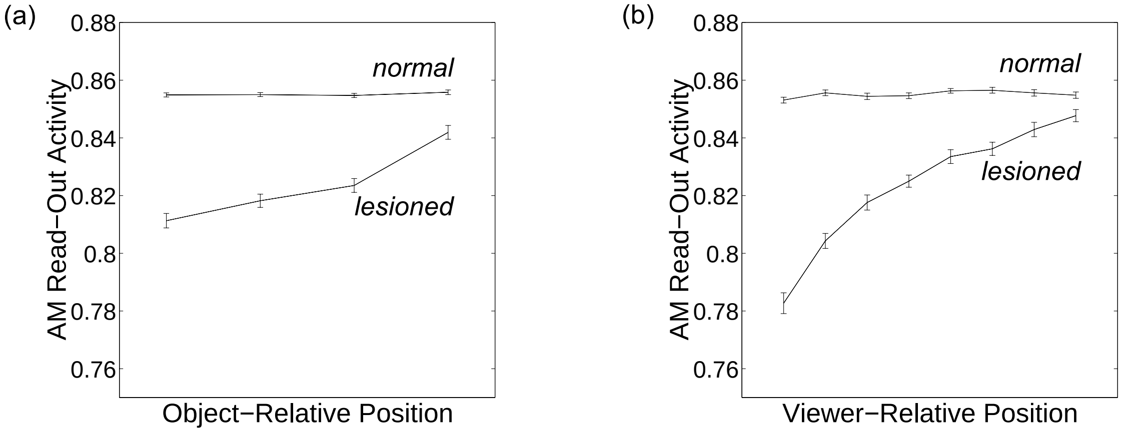


FIGURE 20. Simulation performance of the normal and lesioned AM on the Arguin and Bub (1993) task.

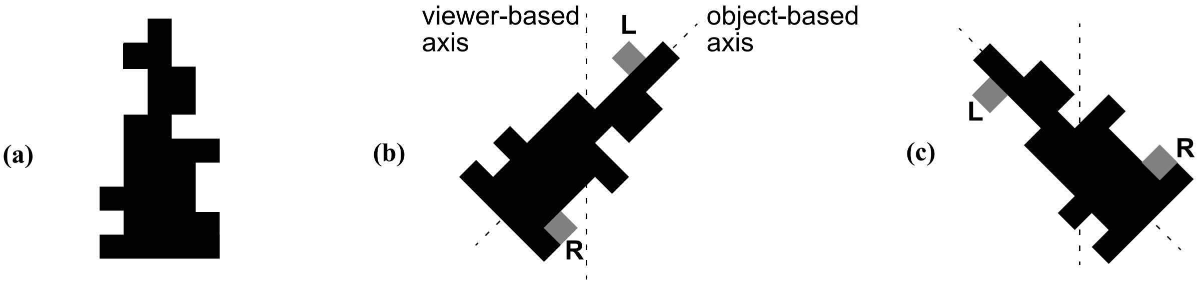


FIGURE 21. Example of an elongated nonsense shape used by Driver and Halligan (1991). (a) When the shape is presented upright, the viewer-based and object-based frames are in alignment. (b) At an orientation of 45°, the frames are dissociated. The detail labeled "L" appears on the object left but the viewer right. The detail labeled "R" appears on the object right but the viewer left. (c) At the other diagonal orientation, the two frames are again in alignment, as in (a).

The critical trials of the Driver and Halligan experiment were those in which the pair of nonsense shapes differed in some detail. Eight shapes were used to form these displays. Each shape was used in one upright and one diagonally oriented display, and in mirror images of these displays, resulting in a total of 32 critical trials.

Our simulation utilized the same eight shapes and 32 critical trials. For each trial, we presented a single shape, with the detail present. The shapes drawn by Driver and Halligan were composed on a 10×5 grid, and each grid cell was translated to a 2×2 region of activity on the topographic map. The activity pattern for the shape in Figure 21a is shown in Figure 22a; activity is enhanced around the contour of the shape, as in all other simulations. To generate the shapes at a diagonal orientation, we designed an activity pattern for the upright shape and then rotated it on the grid 45° clockwise or counterclockwise using the same algorithm as the rotation of the barbell in the Behrmann and Tipper simulation. (The rotation was a mathematical operation to map pixels in the upright shape to pixels in the diagonal shape. It was not part of the simulation; that is, the model was presented with a static stimulus at a diagonal orientation; it did not view the stimulus as it was rotated into position.) Sample exogenous input to the lesioned model for the shape in Figure 21b is presented in Figure 22b. The AM was allowed to settle for 100 iterations, and then the read-out activity was determined at the locations corresponding to the detail. We assume that read-out activity is monotonically related to the likelihood of noticing the detail. Figure 22c depicts the final state of the lesioned AM. Although the difference is not completely apparent in the Figure, the read-out activity of details L and R on this trial was .48 and .79, respectively.

The critical trials fall into four conditions, determined by crossing the side of the shape on which the detail appears with the side of the viewer on which the detail appears. In the Driver and Halligan experiment, the two conditions in which the detail appears on the same side of the two reference frames were studied using upright shapes (e.g., Figure 21a). We chose instead to use shapes oriented at 45° from upright, in either a clockwise or counterclockwise direction such that the relevant detail appeared on the same side of both the viewer-based and object-based frames (e.g., Figure 21c). We made this decision after observing that the unlesioned model produced significantly higher read-out activity for upright shapes than diagonal shapes, due to the coarse resolution of the topographic map and quantization effects in the diagonal displays. The stimulus in Figure 21c allows one to test the same conditions as the stimulus in Figure 21a, but is a better matched control to the stimulus in Figure 21b.

Our simulation involved ten replications of each critical trial. An ANOVA performed using the stimulus as the random factor revealed a main effect of object-relative position (object-left details had read-out activity .57; object-right details had read-out activity .67; $F(1,7) = 6.07$, $p < .05$), a main effect of viewer-relative position (viewer-left details had read-out activity .43; viewer-right details had read-out activity .80; $F(1,7) = 51$, $p < .001$), but the interaction did not reach significance ($F(1,7) = 1.6$, $p = .24$). Most importantly, regardless of whether the detail appeared on the left or the right with respect to the viewer-based frame, the read-out activity was higher for details on the right side of the shape than on the left side of the shape. Thus, the model replicated the key finding from the patient data.

The simulation and the patient data did not match in that viewer position had the larger effect in the simulation, whereas object position had the larger effect in the patient data. However, the human data was from a single patient, and the simulation data assumed a specific lesion. One should therefore not expect an exact match between the simulation and patient, just as one would not expect an exact match between the data from two different patients. Nonetheless, the model does replicate the key result which suggested to Driver and Halligan the role of object-based reference frames.[8]

Driver, Baylis, Goodrich, and Rafal (1994) explain how effects of within-object position could arise from viewer-based frames of reference. We summarize their explanation, because it serves as a helpful lead-in to the explanation suggested by the AM. If one traces a horizontal path in the viewer-based frame intersecting detail L in Figure 21b, detail L always appears to the left of other shape information along this path, and similarly, detail R always appears to the right of other shape information along a horizontal path intersecting it. Thus, the viewer-based and object-based frames have not been completely unconfounded, and the advantage for detail R over detail L could be explained by neglect for information that appears on the *relative* left in the viewer-based frame, i.e., information whose viewer-based horizontal position is on the left relative to other information with the same viewer-based vertical position. The AM suggested essentially this relative neglect explanation for the Arguin and Bub data. Although

---

[8] The larger effect of viewer position in the model is responsible for another discrepancy between the model and patient performance: the patient made fewer errors when the detail was object-right/viewer-left than when the detail was object-left/viewer-right (14 versus 23 errors on 32 trials, respectively). However, the importance of this comparison is questionable, for the reasons explained in the text, and because Driver and Halligan (1991) did not report the statistical significance of the comparison. Rather, they reported only Chi-squared tests for the main effects of viewer position and object position, and their significance results matched those of the model.

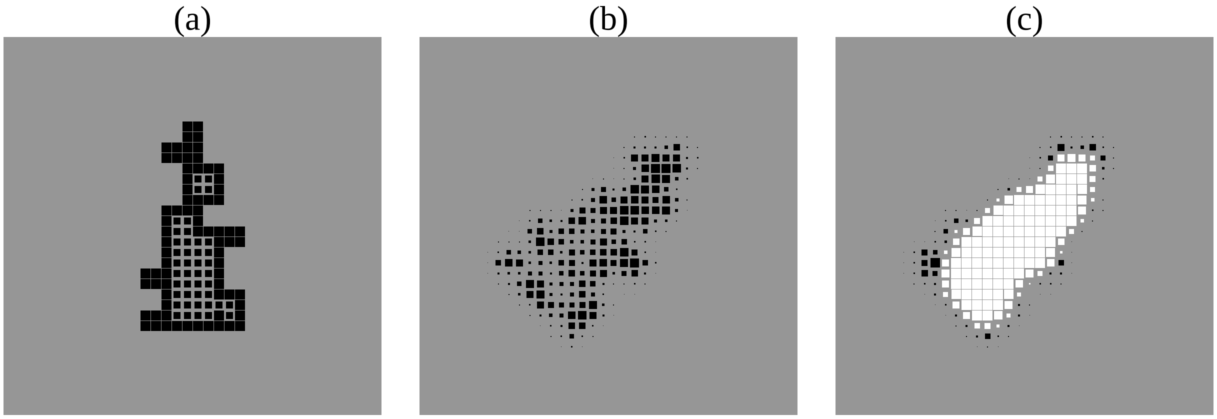


FIGURE 22. (a) Exogenous input to the unlesioned model for an upright Driver and Halligan nonsense shape (Figure 21a). (b) Exogenous input to the lesioned model for a rotated Driver and Halligan nonsense shape (Figure 21b). (c) Asymptotic AM state for the rotated Driver and Halligan nonsense shape.

the AM does not employ competition between left and right along a horizontal path—competition in the AM is *global*, that is, between every pair of locations—the simple account was adequate because the Arguin and Bub stimuli extended primarily in the horizontal direction. However, the explanation requires some elaboration for the Driver and Halligan stimuli, which extend nonuniformly in the vertical direction, and therefore cannot be treated as a one-dimensional shape. The explanation for the Driver and Halligan stimuli is based on the observation that along every horizontal path, the strength of exogenous input to the AM on the right is stronger than on the left. Although competition is global, cooperation is local; the features along the right edge of the shape are neighbors and hence are mutually supportive, as are the features on the left. A gang effect emerges whereby the stronger features on the right edge support one another and hence—due to competition—become even stronger.

Although relative neglect may qualitatively characterize the phenomena observed in the Arguin and Bub and the Halligan and Driver data, the AM offers a deeper explanation with greater predictive power. Relative neglect arises in the AM due to an interaction between the competition and cooperation rules and the specific pattern of exogenous input provided by the bias rule. It is even conceivable that for some configurations of visual features in a display, the balance of competition and cooperation may be such that relative neglect is minimized or suppressed. Thus, the AM suggests that the pattern of neglect will depend not simply on the count of features to the left or right of some location, but also on the arrangement of features. Such effects have been observed even for simple bisection tasks (Halligan & Marshall, 1994), and serve to explain the next two experimental results.

Driver, Baylis, Goodrich, and Rafal (1994)

Because neglect of the relative left in the viewer-based frame could explain the results of Driver and Halligan as well as Arguin and Bub, Driver et al. (1994) explored a task that additionally controlled for the relative horizontal position of the critical information, ruling out relative position as a complete explanation for the object-based neglect they observe. They presented an array of equilateral triangles, and the patient's task was to detect a gap in the center triangle. Two different configurations were studied; in one, the edge in which the gap appeared—the *critical edge*—was to the left of midline with respect to the reference frame established by the array (Figure 23a; call this the *left* condition), and in the other, the critical edge was to the right of midline (Figure 23b; call this the *right* condition). In the two conditions, the absolute and relative horizontal viewer-based position of the critical edge was identical. The difference between the two conditions can be described only with respect to the principal axes of the array.

The simulation of this experiment involved presenting each of the arrays 200 times to the lesioned AM and measuring the mean activity of locations corresponding to the critical edge over the first twenty iterations of a trial. Because it was difficult to construct equilateral triangles given the limited resolution of the AM, the simulated configurations were tilted slightly, from 60° to 45°, but this modification did not alter the essential features of the arrays. Figure 24a shows the exogenous input to the AM for the left condition; one can readily see the seven triangles.

The read-out activity of the critical edge in the left and right conditions for the lesioned model was .52 and .66, respectively, a statistically reliable difference ($F(1,199)=466$, $p<.001$). Thus, the AM shows more neglect for the left condition than the right—just as the patients do—despite the fact that the critical edge is in the same absolute and relative viewer-based position in the two conditions, and it has the same number of features to its left and right in the two conditions. The second two panels of Figure 24 present typical AM states on the final iteration for the left and right conditions, respectively. The critical edge—pointed to by the grey arrow—has less activation for the left condition than for the right.

Observing the AM in operation, one gains an insight as to its behavior. Initially, activity begins to grow around the entire set of triangles. However, the activity of the leftmost triangles rises at a slower rate, their exogenous input being the weakest, and their activity is quickly suppressed by the competition from the rightmost triangles (the competition principle). Roughly, one can conceive of the three rightmost triangles as the "strong locations," and the three leftmost triangles as the "weak locations." The critical edge is closer to the strong locations in the right condition than the left condition; one can readily see this fact in Figure 23. Because of the cooperation principle, which causes the spread of activity to neighboring locations, activation is likely to spill from the strong locations to the critical edge in the right condition, but less likely due to the increased distance in the left condition.

Driver et al. suggest that neglect must operate with respect to the principal axes of the array, although they acknowledge that axis-based neglect might result from an interaction between object-based and viewer-based factors. This account requires that the principal axes be identified, which is tantamount to establishing the object-based frame of reference. The AM can replicate axis-based neglect, yet has no explicit representation of the axes or object-based frame.

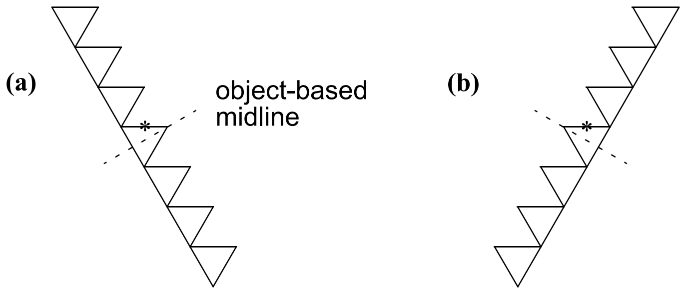


FIGURE 23. Stimuli studied by Driver et al. (1994). The asterisk indicates the edge on which a gap was to be detected. When the critical edge is characterized with respect to the principal axis of the object, it is to the left of the center in (a) but to the right of the center in (b).

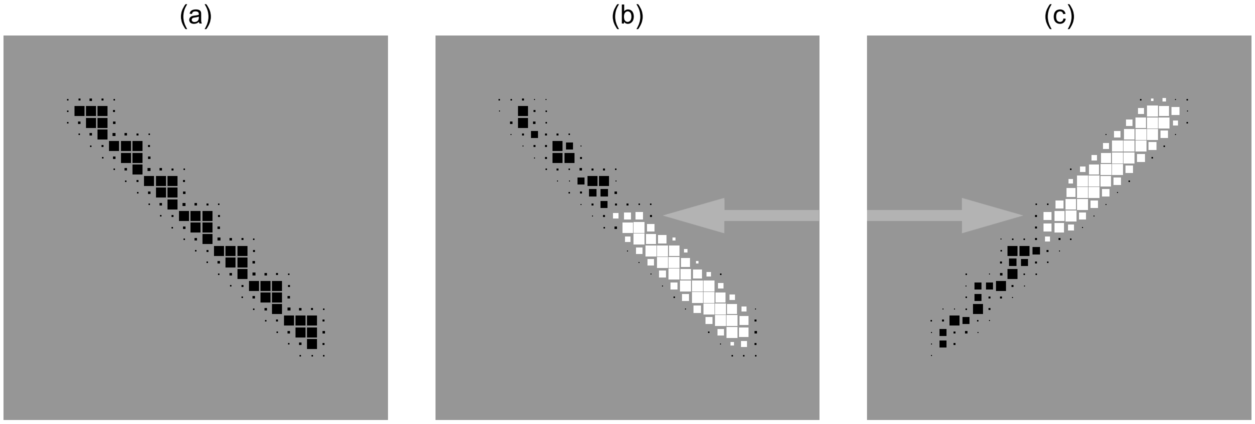


FIGURE 24. Simulation of the Driver et al. (1994) task of detecting a gap in the center triangle. (a) Exogenous input for the normal model in the left condition. (b) State of the lesioned AM at iteration 20 in a typical trial of the left condition. (c) State of the lesioned AM at iteration 20 in a typical trial of the right condition. In (b) and (c), the grey arrows point to the triangle edge in which the gap appears. The critical locations are less active in the left condition than in the right condition.

Driver and Baylis (unpublished)

Driver (1999) briefly describes a follow-up to the equilateral-triangle experiment that is troublesome because the explanation we gave for the AM's behavior in the equilateral-triangle experiment fails to predict the outcome of the follow-up study. Although the follow-up study by Driver and Baylis is unpublished at present, it seems sufficiently important to the line of argument presented in this paper that we attempted a simulation.

As in the equilateral-triangle experiment, the task involved detecting a gap in a center triangle. In this experiment, the triangle was embedded in a context of isosceles triangles (Figure 25a). Symmetry and elongation of the display configuration suggest a principal axis, indicated in the Figure by a dashed line. In the *left* condition, the configuration of triangles is such that the critical edge is left of the principal axis; in the *right* condition, the critical edge is right of the principal axis. If neglect occurs with respect to a reference frame in which the principal axis divides the object into left and right sides, one would expect patient performance in the left condition to be worse than in the right condition. This is indeed what Driver and Baylis found (no quantitative results were reported).

Examining the Figure, the AM would seem to predict exactly the opposite result. In the left condition, the equilateral triangle closer to the critical edge is on the right; in the right condition, the equilateral triangle closer to the critical edge is on the left. By the account we gave for the equilateral-triangle experiment, the adjacent triangle should provide support that spills over to the critical edge, and because the equilateral triangle on the left should provide less support than the equilateral triangle on the right, performance should be better in the left condition than in the right.

Our simulation consisted of a 45×45 retinotopic array designed such that the shortest distance from a point on the isosceles triangle to a point on the equilateral triangle was the same for both isosceles triangles and in both conditions. As in the simulation of the Driver and Halligan (1991) data, we designed upright triangles and then rotated them on the grid to achieve the diagonal orientations. (The rotation was a mathematical operation to map pixels in the upright orientation to pixels in the diagonal orientation; it was not part of the simulation.) The larger array was necessary to achieve the necessary symmetries given quantization effects due to finite resolution of the array.

The simulation of this experiment involved presenting each of the arrays 200 times to the lesioned AM and measuring the mean activity of locations corresponding to the critical edge over the first twenty iterations of a trial. The mean activity was reliably different in the left and right conditions—0.719 and 0.742, respectively ($F(1,199)=182$, $p<.001$). This result is in accord with the human patient data and is astonishing because we expected the opposite effect based on the argument presented earlier. To verify that the result was not due to an artifact in the stimuli, we simulated the unlesioned AM and verified that performance was the same in left and right conditions. To better understand the difference between the left and right conditions, we superimposed the two stimuli such that the equilateral triangles in the two conditions were overlaid on one another (Figure 25b). One curious difference is that, because the isosceles triangles are placed the same distance from the equilateral triangle in each condition, the two stimuli have different horizontal extents: the two isosceles triangles are offset to the viewer left in the right condition relative to the left condition. This offset can be seen in the Figure by comparing the dashed lines indicating the left and right extent of the isosceles triangles. The consequence of this offset is that the exogenous input to the two isosceles triangles will be weaker in the right condition than in the left, and hence could result in less competition with the critical edge. This explanation is probably not the whole story, because even if the offset were eliminated, a crucial factor distinguishes the stimuli in the two conditions: the horizontal distribution of locations occupied by visual features Examining the superimposed image in Figure 25b, the two isosceles triangles on the viewer left of the center triangle are mirror reversed along the horizontal axis. Consequently, in the left condition the tip of the isosceles triangle is further to the viewer right and the base is further to the viewer left. Because the isosceles triangles have nonhomogeneities—for example, the density and mass of features near the tip is greater than near the base—one should expect the distribution of features in the triangles, when modulated by the gradient of attention, to influence competition in the model.[9]

The two factors distinguishing left and right conditions that we identified both contribute to the activation of the locations of the two isosceles triangles to be lower in the right condition than in the left: 0.508 and 0.746 for the two

---

[9] To alleviate any concern that our explanations of the isosceles-triangle and equilateral-triangle studies are inconsistent with one another, note that the distribution of visual locations in the equilateral-triangle study was the same in left and right conditions. Hence, the factor we identified as critical in the isosceles-triangle study is irrelevant in the equilateral-triangle study. And although the factor we identified as critical in the equilateral-triangle experiment—the local neighborhood around the critical edge—could have played a role in the isosceles-triangle experiment, we can speculate that it was not as significant a factor as the distribution of visual locations containing features.

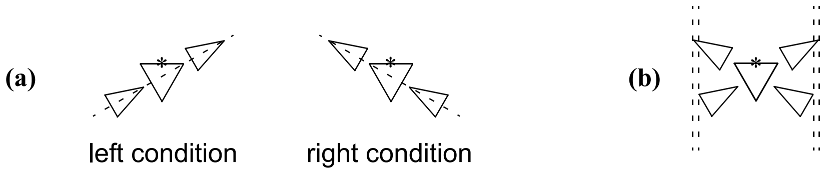
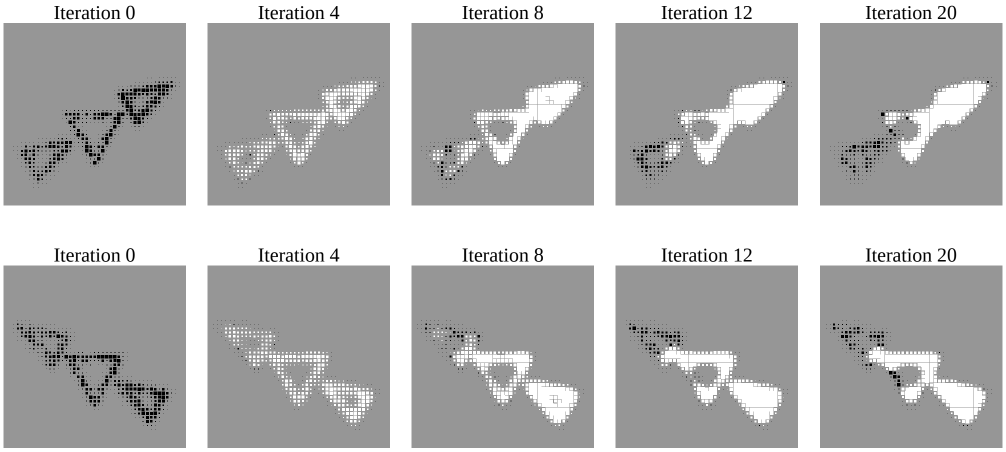


FIGURE 25. (a) Stimuli studied by Driver and Baylis (unpublished). The task is to detect a possible gap in the center triangle, which appears in the location indicated by the asterisk. Depending on the configuration of triangles, the gap will appear either left or right of the principal axis, indicated by the dashed line. (b) An image of the two stimuli superimposed such that the equilateral triangles are overlaid on one another. The dashed lines indicate the left and right edges of the two pairs of isosceles triangles.



Iteration 0 Iteration 4 Iteration 8 Iteration 12 Iteration 20

Iteration 0 Iteration 4 Iteration 8 Iteration 12 Iteration 20

FIGURE 26. Simulation of the Driver and Baylis (unpublished) task of detecting a gap in the center triangle. Each frame shows the state of the lesioned AM after a given number of iterations in typical trials of the left and right conditions (upper and lower frames, respectively).

isosceles triangles in the right condition, versus 0.597 and 0.774 in the left. Less activation of the isosceles triangles means less competition for the equilateral triangle, and more activation of the critical edge. If weaker competition is the cause of the higher activation in the right condition, then the proximity of the isosceles triangles to the equilateral triangle should not matter, because the offset will be present for any fixed separation of the triangles, and inhibition in the model is global and does not depend on local configurations of features. Indeed, in displays with greater space between triangles, the right condition still results in more activation than the left, consistent with our hypothesized explanation.

Our goal in the preceding paragraph was to offer several factors that distinguish the left and right conditions and that are salient to a model like the AM but that have nothing to do with the claimed distinction—the location of the critical edge with respect to the principal axis. The AM once again points to the challenge of designing carefully controlled stimulus displays that admit only one explanation for a patient's performance: the psychological reality of object-based frames.

Discussion

The neuropsychological studies we addressed are concerned with the issues of what internal representations are constructed in the ordinary course of visual information processing, and whether attention can be directed in coordinates defined by the object itself. In one study, Behrmann and Tipper (1994) observed that neglect remains with the left side of an object when the object is inverted. However, this reversal of neglect is not observed for displays in which the left and right sides of the object are disconnected (Tipper & Behrmann, 1996) or for additional, fixed objects in the display (Behrmann & Tipper, 1999). Pavlovskaya, Glass, Soroker, Blum, and Groswasser (1997) improved accuracy of object identification by cuing patients to a location left of the center of the object. Arguin and Bub (1993) found reaction times longer for a stimulus located on the left side of an object than the right side, controlling for retinal position of the stimulus. In two different paradigms, Driver and Halligan (1991). Driver, Baylis, Goodrich, and Rafal (1994), and Driver and Baylis (unpublished) observed neglect operating with respect to the principal axis of an object. These results were interpreted by the authors of the studies to support the psychological reality of a frame of reference other than the viewer-based frame:

> "The findings suggest that attention operates on object-centered as well as location-based representations, and thus accesses multiple reference frames." (Tipper & Behrmann, 1996)
>
> "...[P]atients with visual neglect...represent information in both location- and object-centered frames simultaneously in the same task...These results cannot be explained by any model of attention that argues solely for a location-based medium...Rather, the current findings can only be explained by assuming that the positions of the targets are defined with respect to particular frames of reference, and that spatial position (and neglect thereof) is determined with respect to these frames." (Behrmann & Tipper, 1999)
>
> "Our data might be interpreted in terms of neglect operating in the object-centered coordinate frame." (Pavlovskaya et al., 1997)
>
> "...A stimulus-centered spatial reference frame...may be affected in the visual hemineglect syndrome...Thus, we suggest that the concept of stimulus-centered reference frame corresponds to a level of spatial representation that is generally used in human vision..." (Arguin & Bub, 1993)
>
> "...Our results imply that [patient] PP's stroke has damaged neural systems which normally code visual information in...object-centred co-ordinates..." (Driver & Halligan, 1991)
>
> "A *purely* egocentric neglect is unable to explain the present axis-based result...[O]ur experiment shows...that the *dividing line* between neglected and spared sides...can be determined by the principal axis assigned to a shape during perception." (Driver et al., 1994)

Although the authors are careful to state their conclusions tentatively, the accumulation of such studies has led to widespread acceptance of the psychological reality of object-based frames of reference in visual object recognition. Because the existence of object-based frames is a fundamental source of evidence supporting allocentric theories of object recognition, allocentric theories are generally viewed as canonical in the cognitive neuroscience community, (e.g., Caramazza & Hillis, 1990a), and provide a motivation, often implicit, for experimental work (e.g., the studies analyzed here).

Contrary to the predominant view in cognitive neuropsychology, the simulations reported in this paper suggest that object-based frames of reference are not necessary to explain the neglect data. Our model can account for the data yet operates only in a viewer based frame.[10] Consequently, one must be cautious in interpreting even seemingly clear-cut experimental results without the aid of a computational model.

Evaluating the model

Computational modeling is a valuable exercise for many reasons. A computational model provides a concrete embodiment of a theory. It forces one to be explicit about one's claims. It allows one to examine interactions among assumptions. However, a computational model makes its greatest contribution when it offers a new or fundamentally different conceptualization of data. The model described here, the AM, has succeeded in this regard. The neglect data seemed to demand an explanation involving object-based frames. No qualitative model could convincingly argue otherwise; only a simulation model could resurrect a class of explanations that would otherwise be ruled out. Even if the internal dynamics of the AM were incomprehensible to human observers, it would still provide an existence proof—a detailed model having only viewer-based frames that can nonetheless explain the data.

The inner workings of the AM are indeed difficult to comprehend. We are sometimes successful in explaining its behavior in qualitative language that can be communicated to others, but not always. The model's behavior is an emergent property of the interaction of cooperative and competitive forces. One should not expect that such complex dynamics can be reduced to a simple explanation that sidesteps the dynamics. To the degree we have succeeded in characterizing the model's performance, the characterizations come via post hoc analysis of the simulation results.

Indeed, the AM has sufficiently complex dynamics that its creator has difficulty in predicting the outcome of a simulation. Many results we have modeled using the AM were unexpected and surprising. In the simulations reported here, the disconnected disk condition of Tipper and Behrmann (1996) and Driver et al. (1994) are two cases in point. The model is far more interesting and subtle than we first realized. However, its success in explaining a wide variety of data is undeniable. Each time that the model, with only trivial extensions, can explain a diverse corpus of data it was not designed around, one must increase one's confidence in the model. A relatively simple model like the AM could not continue to provide accounts of data were it not in some basic sense correct.

As the careful reader has no doubt noticed, the AM sometimes produces curious behaviors that might have a correlate in patient behavior. We list three examples from the Behrmann and Tipper simulations. (1) The AM produces a reversal of neglect in the barbell displays even if the center portion of the bar is missing (Figure 12, lower right panel). (2) When the barbell has been rotated 180°, there appears to be some asymmetry in the distribution of attention to the disk on the viewer left, suggesting that patient reaction times might be faster to a target on the right side of the disk than on the left. (Figure 10, lower right panel). (3) In the disconnected disk condition, there appears to be a short period of time during which both disks are attended (Figure 11), suggesting that if the target appeared at the critical instant, reaction times might be comparable for the two disks.

We are currently testing patients on these and other experimental tasks in which contrasting predictions are made by the AM and an account relying on object-based reference frames (McGoldrick, Mozer, Munakata, & Reed, in preparation). One natural reaction to the present simulation studies is that the value of the model cannot be ascertained until the model can be shown to predict to results of *novel* experiments. However, this reaction is misguided, because in fact the present simulation studies already represent predictions of the model. One should not be concerned with whether the experimental data explained by the model existed in the literature prior to the simulation. Rather, the essential considerations in evaluating the predictive power of a model on an experiment are: whether the model was designed with the experimental data in mind, and whether the model had sufficient degrees of freedom that it could provide an interpretation for any outcome of the experiment. Neither concern is warranted in this case. Thus, in response to the question of whether the AM has any predictive power, the simple answer is that *all* of the results reported here are predictions of an existing, highly constrained computational model. Even if it turns out that further predictions of the model are not supported by patient studies, the model has been valuable in pointing out

---

[10] Further, the model can readily explain other neglect data that has been used in support of, or that has presupposed the existence of, object-based frames, including Grabowecky, Robertson, and Treisman (1993), as well as data from normals which shows neglect-like object-based attentional biases, such as Reuter-Lorenz, Drain, and Hardy-Morais (1996). The explanation of the Arguin and Bub dad can be applied to these cases as well.

alternative explanations for each study simulated, and these alternative explanations could be correct even if the model is ultimately proven inadequate.

An alternative model

Recently, Pouget and Sejnowski (1997) and Pouget, Deneve, and Sejnowski (in press) have proposed a neurobiological model of spatial representations in parietal cortex, and have used the model to explain data from the neglect literature, including phenomena that have been modeled by the AM such as line bisection and relative neglect. We view this model, which we will refer to as the *basis-function model*, as being consistent with and complementary to the AM, focusing on the neurobiological level of description rather than the psychological level. The heart of the basis-function model is the claim that neurons in the parietal cortex are tuned to respond to a stimulus appearing at a specific retinal location and this response is gain modulated by eye position. The model is referred to as a basis-function model because the representation supports both eye-centered and head-centered representations—the former useful for saccadic eye movements and the latter useful for reaching.

The model has two hemispheres, and although neurons in each hemisphere encode all of space, the representation in a hemisphere emphasizes the contralesional side of space. Specifically, a neuronal gradient is assumed in which the right hemisphere contains relatively many neurons coding for leftward retinal locations and leftward eye positions and relatively few neurons coding for rightward retinal locations and rightward eye positions. Consequently, a lesion to the right hemisphere results in a gradient of damage to the representation of space, most severe on the left and least severe on the right. This gradient assumption is the same as that postulated by the AM.

Beyond its claims about parietal representations, the basis-function model makes some fairly standard assumptions about the nature of attention, including a winner-take-all competition among locations, inhibition of return, and detection accuracy and speed being dependent on the saliency of a location. These processes are described at an abstract level, not in terms of neural mechanisms.

Although we view the AM and the basis-function model as being entirely consistent, the basis-function model has some virtues. First, the model makes contact with the underlying neurobiology. Second, the model deals with sensorimotor behavior as well as perception. Third, the model addresses the specific nature of the viewer-based representation in terms of both eye-centered and head-centered coordinate frames. Fourth, the model shows that with little more than the gradient assumption, one can explain some of the data concerning relative neglect; the complex dynamics of the AM are not essential.

The basis-function model also has its limitations. First, the central claims of the model concern the representation of space, but additional claims are required about the processes that operate on the representation in order to extract responses from the model. Consequently, the model lacks a uniform mechanistic foundation to explain data from diverse experimental paradigms, and hence, each new experimental paradigm seems to require additional assumptions, and the ratio of assumptions to data explained seems high relative to that of the AM. Second, although the basis-function model can explain some if not all cases of relative neglect (e.g., Arguin & Bub, 1993) without explicit object-based frames, it does not provide an explanation for the challenging studies of Behrmann and Tipper (1994, 1999; Tipper & Behrmann, 1996), or an explanation of the Driver et al. (1994) gap-detection study without invoking object-based frames.[11] The complex dynamics of the AM are what give it the power to explain these phenomena.

Rather than viewing one model or the other as superior, we feel it quite plausible that the strengths of the basis-function model and the AM can be integrated into a single model that accounts for an even broader range of behavioral and neurophysiological data with even fewer assumptions than either of the current models.

The meaning of “object based”

The phrase “object based” is ambiguous, and a lack of clarity as to its intended meaning has resulted in some confusion in the literature. “Object based” can be a *descriptive term for experimental results*. All of the experimental studies simulated in this paper showed object-based effects, in that the behavior of patients was dependent not merely on the location of an object with respect to the viewer but on the extent, shape, or movement of the object itself.

---

[11] Deneve and Pouget (1998) do address the Driver et al. result, but to explain the result in terms of the basis-function model, they must suppose that the orientation of the row of triangles is factored out, i.e., the row of triangles is rotated such that its principal axis is aligned with the viewer left-right axis. This assumption is consistent with the conclusion of Driver et al., and essentially involves the establishment of an object-based frame. In contrast, the AM does not require the determination or the explicit representation of the principal axis of the row of triangles.

"Object based" can also be a *characterization of processes and internal representations*. Object-based representations arise from processes that utilize object-based frames of reference to transform visual features to achieve partial or complete view invariance. Our simulations show that object-based *effects* can be obtained without object-based *representations* or *frames of reference*.

The distinction between object-based effects and object-based representations does not entirely remove the ambiguity in the phrase "object based." One can conceive of a continuum of senses in which a model's processes and representations might be considered object based. Examples of at least four alternatives can be found in the literature, which we present in order from weakest to strongest notions of object based. (See Driver, 1999, for a similar enumeration of alternatives.)

(1) *Segmentation in a viewer-based frame* (Grossberg & Raizada, in press; Mozer et al., 1992; Vecera & Farah, 1994). Some models attempt to perform segmentation, i.e., grouping together the visual features that belong to one object. Segmentation can be performed in a viewer-based reference frame utilizing grouping heuristics that exploit the statistics of objects in visual scenes. For example, the cooperation rule of the AM results in grouping of neighboring locations in a viewer-based frame, but because neighboring locations in a viewer-based frame tend to be neighboring locations in an object-based frame, the AM tends to select objects. If other grouping heuristics were incorporated into the AM, as we suggested earlier might be necessary in a full-blown implementation of the model, then the AM's behavior would appear even more object based (and more like the model of Mozer et al., 1992).

(2) *Segmentation and determination of principal axis* (Driver, 1999). In addition to performing segmentation in a viewer-based frame, a model might also determine the principal axis of an object—the axis of symmetry or elongation. Using the axis to establish a partial frame of reference—such as an up-down direction—visual features could be reinterpreted with respect to the partial frame. For example, Driver et al. (1994) suggest that the shape in Figure 23a evokes a principal axis from which the midline of the shape can be determined, and the left-right position of visual features is then determined with respect to the midline, although the specification of which direction is "left" and which is "right" arises from the viewer-based frame (Driver, 1999).

(3) *Segmentation and determination of an object-based frame of reference* (Marr & Nishihara, 1978). A model might determine not only the up-down direction of an object, but also its left-right and front-back direction, allowing for the establishment of a full-blown object-based frame of reference.

(4) *Segmentation and determination of a structural description* (Biederman 1987; Hummel & Biederman, 1992; Hummel & Stankiewicz, 1998). To handle complex, articulated objects, a model might construct a structural description that decomposed an object to its parts and described the relationships among the parts; this is a generalization of an object-based frame of reference.

Alternative (3) is the common notion of object based in the cognitive neuropsychology literature, although Driver (1999) has argued that alternative (2) is sufficient to explain key neglect data. Modern allocentric theories of object recognition assume the necessity of alternative (4). The AM is an example of alternative (1). Indeed, it is a simple formulation of alternative (1), in that segmentation depends only on the proximity of features, although we are ready to concede that a more complex formulation, still within the framework of alternative (1), will be necessary to explain a broader corpus of data. Alternative (1) provides the weakest notion of object based, in that it does not require the explicit computation of a principal axis, frame of reference, or structural description. Because it is consistent with the data—as demonstrated by simulations of the AM—it seems to provide the most parsimonious account.

One might ask whether the AM model an object-based account. The preceding discussion is intended to forestall this question, and to argue that the question itself is indicative of a lack of clarity concerning the various notions of "object based." Clearly, the AM produces object-based effects. The AM also has some knowledge, albeit extremely weak and low order, about objects. However, the question should not be whether or not a model is object based, but the degree to which it requires the explicit computation of object properties, such as a principal axis, frame of reference, or structural description, and the degree to which the data mandates such computations.

The status of object-based frames of reference in neglect

Several researchers have noted caution in invoking object-based frames of reference to explain data from neglect patients. Buxbaum (1995), Pouget and Sejnowski (1997), Pouget, Deneve, and Sejnowski (in press), and Driver and Pouget (in press) argue—consistent with the present work—that apparent object-based deficits in neglect might arise from attentional gradients in viewer-based space. Driver et al. (1994) acknowledge that the object-based effects of Driver and Halligan (1991) might be attributable to a relative deficit of attention in viewer-based space. Farah (1990; see also Vecera & Farah, 1994) argues for location-based encoding of object properties, and attributes object-based effects to the fact that parietal attentional processes are part of an interactive system that includes other parts of the brain that recognize objects.

Some studies have found no or limited support for object-based frames in neglect. Farah, Brunn, Wong, Wallace, and Carpenter (1990) failed to obtain neglect in an object-based frame for familiar objects. Behrmann and Moscovitch (1994) replicated this finding for both objects and symmetric-letter stimuli, but did observe neglect for the intrinsic left side of asymmetric-letter stimuli. However, the finding for asymmetric letters may be attributable to their having more complex structure on the right (Drain & Reuter-Lorenz, 1997). Buxbaum, Coslett, Montgomery, and Farah (1995) eliminate object-based neglect by manipulating task instructions, and suggest that mental rotation may underlie at least some cases of object-based neglect.

Two neglect-related phenomena have been reported that are not immediately explained by the AM and its viewer-based representation of space. Caramazza and Hillis (1990b; Hillis & Caramazza, 1995) have studied patients who show neglect for the right side of a word, in both perception and production, across perceptual modalities, and—most pertinent to the issue of frames of reference in visual perception—irrespective of the topographic arrangement of letters (neglect was observed for words whose letters are arranged in a column or are mirror reversed). Humphreys and Riddoch (1994, 1995; Riddoch, Humphreys, Luckhurst, Burroughs, & Bateman, 1995) have observed patients who manifest left neglect in single words but right neglect in multiple-stimulus displays. These phenomena, though undoubtably real, do not necessarily conflict with the perspective we have presented. One might accommodate these phenomena by positing specialized visual representations or processing mechanisms for words versus other visual stimuli. Driver et al. (1994) note the role of handedness in specifying identity of letters and words as one factor that might make words unique. For the Caramazza and Hillis (1990) patient, Driver (1999) comments that neglect is found for letter strings presented aurally, suggesting a deficit arising at a high level of representation, not in the visual processing system. One can also argue that the phenomena implicating object-based representations arise from a complex interaction of perceptual and motor processing that is both strategic and task dependent, in contrast to the seemingly more pure-perception tasks simulated in this paper (Buxbaum et al., 1995). For example, reading a mirror-reversed word may involve piecing together the letters one at a time in a verbal or visual short term store (Farah & Buxbaum, 1997); and right-sided motor neglect could explain difficulty in processing multi-item displays via the guidance of eye movements. Such motor and strategic factors contaminate the interpretation of patient performance on complex, temporally-extended tasks in terms of perceptual reference frames.[12]

Other evidence for object-based frames of reference?

In this section, we discuss three sources of evidence for object-based frames in the cognitive neuroscience literature: neurophysiological studies, object-based attentional effects in normals, and configuration effects.

In an investigation of the neural basis of object-based representations of space, Olson and Gettner (1995, 1996) trained monkeys to make saccades to the left or right side of an object, and recorded activity of neurons in the supplementary eye field. They observed object-based directional sensitivity of neural activity: the response of certain neurons depended not on the direction of the saccade or the location to which the saccade was directed, but on the side of the object to which the saccade was directed. These data would seem to support the neurobiological reality of object-based representations. However, Deneve and Pouget (1998) present a model that accounts for the data without relying on explicit object-based representations. Instead, neurons in the model have receptive fields defined in eye-centered coordinates, i.e., they represent the direction and amplitude of saccades with respect to the fixation point, and are modulated by task instructions—whether the saccade is to be directed to the left or right side of the object.

Object-based attentional effects in normal subjects (e.g., Duncan, 1984; Egly, Driver, & Rafal, 1994; Vecera, 1993) are no more a source of evidence for object-based frames than are object-based attentional effects in neglect

---

[12] Many authors (e.g., Behrmann & Tipper, 1999; Humphreys & Riddoch, 1995; Vecera & Farah, 1994) have argued that task contingencies and instructions can influence whether the allocation of attention is primarily object based or location based. The availability of multiple attentional allocation strategies is pretty much orthogonal to, though consistent with, the perspective presented in this paper because object-based attentional effects can be explained without object-based reference frames. However, the flexibility of attentional allocation, and strategic effects more generally, do make it more difficult to conclusively interpret data as implicating object-based or multiple reference frames in perception.

patients. Object-based attentional effects can be readily explained via a viewer-based encoding of object properties (Mozer et al., 1992; Mozer & Sitton, 1998; Vecera & Farah, 1994).

Configuration effects (Attneave, 1968; Palmer, 1980; Rock, 1984) are an intriguing source of data concerning the role of object-based frames in perception. Consider the sort of displays studied by Attneave (1968), shown in Figure 27. The elements in the configuration on the left are interpreted as squares, whereas the identical elements appearing in the configuration on the right are interpreted as diamonds. A natural account of these data, which we call the *principal-axis* account, proposes that the principal axis of the configuration is first identified, and the elements are then interpreted with respect to the principal axis. Although the use of a principal axis is weaker than establishing a full blown object-based frame of reference (see previous discussion of different notions of "object based"), it is nonetheless a computation that was not required by the AM to explain the neglect data. Other data also support the role of principal axes; for example, Ling and Sanocki (1995) found that major axes can be primed to facilitate identification. Nonetheless, one could conceive of an account that does not require of explicit determination of a principal axis. The account is in the spirit of models that activate multiple interpretations based on all sources of evidence, and then allow constraint satisfaction to select a single interpretation (e.g., Kintsch, 1988; Mathis & Mozer, 1996; McClelland, Rumelhart, & Hinton, 1986). By this account, which we call the *construction-integration* account following the terminology of Kintsch (1988), the display elements in Figure 27 would activate detectors that corresponded to both "diamond at 0° orientation" and "square at 45° orientation" interpretations. In addition, for the display on the left, low spatial-frequency detectors would indicate high energy at 45° in the display. During constraint satisfaction, the 45°-square and 45°-energy detectors would mutually cooperate, the 0°-diamond and 45°-energy detectors would compete with one another, and the 45°-square and 0°-diamond detectors would be also compete with one another, resulting in selection of the 45°-square interpretation of the display elements. The construction-integration and principal-axis accounts both involve two stages, but the stages are very different in nature. The elegance of a construction-integration account comes from the fact that each stage can be relatively simple, heuristic, and error prone, yet the result of the computation can be reliable. However, it remains to be seen through computational modeling whether such an account will turn out to be more parsimonious and consistent with the data than a principal-axis account.

In summary, although one cannot completely discount the support for object-based frames of reference from areas of cognitive neuroscience other than neuropsychology, the support seems weak in that alternative explanations of the data are readily available.

Conclusions

The simulations presented in this paper severely weaken the case for object-based frames of reference in visual perception. The existence of viewer-based frames is undisputable: Early visual information is encoded with respect to retinal location and gaze direction. However, the conjecture supported by our computational model is that object-based frames of reference are not required for ordinary visual perception. The issue of object-based frames is central because it divides the two classes of theories of object recognition (Figure 2). Allocentric theories suppose that—at bare minimum—object-based frames are utilized in constructing visual representations from which recognition is performed, whereas egocentric theories utilize visual representations in a viewer-based frame to perform recognition. In the absence of strong empirical support for object-based frames, either from neuropsychology, behavioral studies in normals, or neurophysiological studies, a class of theories utilizing only viewer-based frames—egocentric theories—would seem the more parsimonious.

One should not interpret our line of argument as supporting the stronger conjecture that object-based frames are nonexistent in the brain. Surely if demanded by the task, people can mentally construct visual object-based representations. However, the AM suggests that such a complex cognitive ability is not called upon in the course of ordinary visual perception in naturalistic environments, and is built on top of a more basic perceptual apparatus that operates using viewer-based frames of reference.

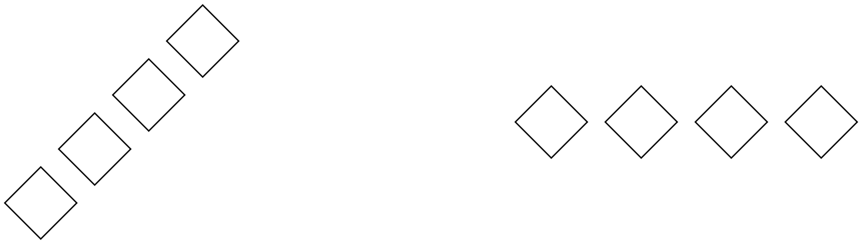


FIGURE 27. Configurations of elements studied by Attneave (1968). The elements of the two configurations are identical, but the elements on the left are interpreted as squares and the elements on the right are interpreted as diamonds, due to the arrangement of elements.

Acknowledgments

This research greatly benefited from the comments and insights of Paul Smolensky, Jon Driver, Glyn Humphreys, John Duncan, John Hummel, Brenda Rapp, Jeff Elman, James Reggia, John McGoldrick, Yuko Munakata, and Alex Pouget. This research was supported by Grant 97-18 from the McDonnell-Pew Program in Cognitive Neuroscience, and by NSF award IBN-9873492.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum Associates.

Arguin, M., & Bub, D. N. (1993). Evidence for an independent stimulus-centered spatial reference frame from a case of visual hemineglect. *Cortex*, *29*, 349–357.

Attneave, F. (1968). Triangles as ambiguous figures. *American Journal of Psychology*, *81*, 447–453.

Behrmann, M., & Moscovitch, M. (1994). Object-centered neglect in patients with unilateral neglect: Effects of left-right coordinates of objects. *Journal of Cognitive Neuroscience*, *6*, 1–16.

Behrmann, M., & Tipper, S. P. (1994). Object-based attentional mechanisms: Evidence from patients with unilateral neglect. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious processing and cognitive functioning* (pp. 351–375). Cambridge, MA: MIT Press.

Behrmann, M. & Tipper, S. P. (1999). Attention accesses multiple reference frames: Evidence from visual neglect. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 83–101.

Behrmann, M., Zemel, R. S., and Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal subjects and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*,*24*, 1011–1036.

Beymer, D. J. (1993). Face recognition under varying pose. A.I. Memo No. 1461. Cambridge, MA: Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.

Biederman, I., & Bar, M. (1999). One shot viewpoint invariance in matching novel objects. *Vision Research*, *39*, 2885-2899.

Biederman, I., & Gerhardstein, P.C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1162-1182.

Biederman, I., & Gerhardstein, P.C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1506-1514.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London, Series B (Biological Sciences)*, 352, 1203-1220.

Bülthoff, H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Science*, *89*, 60-64.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *5*, 247-260.

Buxbaum, L. J. (1995). Visual attention to objects and space: Object-based neglect or location focus? Paper presented at the *International Neuropsychological Society Conference*, Seattle, WA, February 1995.

Buxbaum, L. J., Coslett, H. B., Montgomery, M. W., & Farah, M. J. (1996). Mental rotation may underlie apparent object-based neglect. *Neuropsychologia*, *34*, 113–126.

Calviano, R., Petrone, P. N., & Levine, D. N. (1987). Left visual spatial neglect is both environment-centered and body-centered. *Neurology*, *37*, 1179–1183.

Caramazza, A., & Hillis, A. E. (1990a). Levels of representation, coordinate frames, and unilateral neglect. *Cognitive Neuropsychology*, *7*, 319–445.

Caramazza, A. & Hillis, A. E. (1990b). Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature*, *346*, 267–269.

Deneve, S., & Pouget, A. (1998). Neural basis of object-centered representations. In M. I. Jordan, M. J. Kearns, & S. A. Solla (Eds.), Advances in neural information processing systems 10 (pp. 24–30). Cambridge, MA: MIT Press.

Drain, M., & Reuter-Lorenz, P. A. (1997). Object-centered neglect: Do informational asymmetries play a role? *Neuropsychologia*, *35*, 445-456.

Driver, J. (1999). Egocentric and object-based visual neglect. In N. Burgess, K. J. Jeffery, & J. O'Keefe (Eds.), The hippocampal and parietal foundations of spatial cognition (pp. xxx-xxx). New York: Oxford University Press.

Driver, J., Baylis, G. C., Goodrich, S. J., & Rafal, R. D. (1994). Axis-based neglect of visual shapes. *Neuropsychologia*, *32*, 1353–1365.

Driver, J., & Halligan, P. W. (1991). Can visual neglect operate in object-centered coordinates? An affirmative single-case study. *Cognitive Neuropsychology*, *8*, 475–496.

Driver, J., & Pouget, A. (in press). Object-centered visual neglect, or relative egocentric neglect? *Journal of Cognitive Neuroscience*.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517.

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Science*, *1*, 296-304.

Edelman, S., & Bülthoff (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, *32*, 2385-2400.

Edelman, S., & Duvdevani-Bar, S. (1997). A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London, Series B (Biological Sciences)*, *352*, 1191-1202.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161–177.

Ellis, R., Allport, D. A., Humphreys, G. W., & Collis, J. (1989). Varieties of object constancy. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *41*, 775-796.

Farah, M. J. (1990). *Visual agnosia*. Cambridge, MA: MIT Press/Bradford Books.

Farah, M. J., Brunn, J. L., Wong, A. B., Wallace, M. A., & Carpenter, P. A. (1990). Frames of reference for allocating attention to space: Evidence from the neglect syndrome. *Neuropsychologia*, *28*, 335–347.

Farah, M. J., & Buxbaum, L. J. (1997). Object-based attention in visual neglect: Conceptual and empirical distinctions. In P. Thier & H.-O. Karnath (Eds.), *Parietal lobe contributions to orientation in 3D space* (pp. 385–400). Berlin: Springer-Verlag.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition*, *15*, 455–469.

Grabowecky, M., Robertson, L. C., & Treisman, A. (1993). Preattentive processes guide visual search: Evidence from patients with unilateral visual neglect. *Journal of Cognitive Neuroscience*, *5*, 288–302.

Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, *101*, 470–489.

Grossberg, S., & Raizada, R. D. S. (in press). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*.

Halligan, P. W., & Marshall, J. C. (1993). The history and clinical presentation of visual neglect. In I. H. Robertson, & J. C. Marshall (Eds.), *Unilateral neglect: Clinical and experimental studies.* London: Erlbaum Associates.

Halligan, P. W., & Marshall, J. C. (1994). *Focal and global attention modulate the expression of visuo-spatial neglect: A case* study. *Neuropsychologia*, 32, 13–21.

Heilman, K. M., Watson, R. T., & Valenstein, E. (1993). Neglect and related disorders. *Clinical neuropsychology* (2nd edition) (pp. 279–336). New York: Oxford University Press.

Hillis, A. E., & Caramazza, A. (1995). Spatially specific deficits in processing graphemic representations in reading and writing. *Brain and Language*, *48*, 263–308.

Hillis, A. E., Mordkoff, J. T., & Caramazza, A. (1999). Mechanisms of spatial attention revealed by hemispatial neglect. *Cortex*, *35*, 433–442.

Hinton, G. E. (1981). A parallel computation that assigns canonical object-based frames of reference. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (pp. 683–685). Los Altos, CA: Morgan Kaufmann.

Hinton, G. E., & Lang, K. (1985). Shape recognition and illusory conjunctions. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence* (pp. 252–259). Los Altos, CA: Morgan Kaufmann.

Hubel, D. H., & Wiesel, T. N. (1979). Brain mechanisms of vision. *Scientific American*, *241*, 150–162.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480-517.

Hummel, J. E., & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 93-121). Cambridge, MA: MIT Press.

Hummel, J. E., & Stankiewicz, B. J. (1998). Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition*, *5*, 49-79.

Humphreys, G. W., & Heinke, D. (1998). Representation and selection in the brain: Neuropsychological and computational constraints. *Visual Cognition*, *5*, 9-47.

Humphreys, G. W., & Riddoch, M. J. (1995). Separate coding of space within and between perceptual objects: Evidence from unilateral visual neglect. *Cognitive Neuropsychology*, *12*, 283–311.

Humphreys, G. W., & Riddoch, M. J. (1994). Attention to within-object and between-object spatial representations: Multiple sites for visual selection. *Cognitive Neuropsychology*, *11*, 207–241.

Kinsbourne, M. (1987). Mechanisms of unilateral neglect. In M. Jeannerod, *Neurophysiological and neuropsychological aspects of spatial neglect* (pp. 69–86). Amsterdam: North Holland.

Kinsbourne, M. (1993). Orientational bias model of unilateral neglect: Evidence from attentional gradients within hemispace. In I. H. Robertson & J. C. Marshall (Eds.) *Unilateral neglect: Clinical and experimental studies.* London: Lawrence Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 371–379.

Le Cun, Y., Boser, B., Denker, J. S., Hendersen, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*, 541–551.

Ling, X., & Sanocki, T. (1995). Major axes as a moderately abstract model for object recognition. *Psychological Science*, 6, 370–375.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, *5*, 270-288.

Marr, D. (1982). Vision. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London, Series B (Biological Sciences)*, *200*, 269-294.

Mathis, D., & Mozer, M. C. (1996). Conscious and unconscious perception: A computational theory. In G. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 324–328). Hillsdale, NJ: Erlbaum.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1* (pp. 3–44). Cambridge, MA: Bradford Books/MIT Press.

McGoldrick, J., Mozer, M. C., Munakata, Y., & Reed C. (in preparation). Object-based representations and neglect: Tests of a computational model.

McMullen, P. A., & Farah, M. J. (1991). Viewer-centered and object-centered representations in the recognition of naturalistic line drawings. *Psychological Science*, *2*, 275–277.

Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach*. Cambridge, MA: MIT Press/ Bradford Books.

Mozer, M. C. (1999). A principle for the unsupervised hierarchical decomposition of visual scenes. To appear in M. S. Kearns, S. A. Solla, & D. Cohn (Eds.), *Advances in Neural Information Processing Systems 11*. Cambridge, MA: MIT Press.

Mozer, M. C., & Behrmann, M. (1992). Reading with attentional impairments: A brain-damaged model of neglect and attentional dyslexias. In R. G. Reilly & N. E. Sharkey (Eds.), *Connectionist approaches to natural language processing* (pp. 409–460). Hillsdale, NJ: Erlbaum Associates.

Mozer, M. C., Halligan, P. W., & Marshall, J. C. (1997). The end of the line for a brain-damaged model of unilateral neglect. *Journal of Cognitive Neuroscience*, *9*, 171–190.

Mozer, M. C., & Sitton, M. (1998). Computational modeling of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341–393). London: UCL Press.

Mozer, M. C., Zemel, R. S., Behrmann, M., & Williams, C. K. I. (1992). Learning to segment images using dynamic feature binding. *Neural Computation*, *4*, 650–665.

Olshausen, B., Anderson, C. H., & van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, *13*, 4700-4719.

Olson, C. R., & Gettner, S. N. (1995). Object-centered directional selectivity in the macaque supplementary eye field. *Science*, *269*, 985–988.

Olson, C. R., & Gettner, S. N. (1996). Brain representations of object-centered space. *Current Opinion in Neurobiology*, *6*, 165–170.

O'Reilly, R. C., Munakata, Y., & McClelland, J. L. (in press). *Explorations in computational cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

Palmer, S. E. (1980). What makes triangles point: local and global effects in configurations of ambiguous triangles. *Cognitive Psychology*, *12,* 285–305.

Pavlovskaya, M., Glass, I., Soroker, N., Blum, B., & Groswasser, Z. (1997). Coordinate frame for pattern recognition in unilateral neglect. *Cognitive Neuroscience*, *9*, 824–834.

Perrett, D. I. & Oram, M. W. (1998). Visual recognition based on temporal cortex cells: Viewer-centred processing of pattern configuration. *Zeitschrift fuer Naturforschung (C)*, *53*, 518-541.

Perrett, D. I., Oram, M. W., Hietanen, J. K., & Benson, P. J. (1999). Issues of representation in object vision. In M. J. Farah & G. Ratcliff (Eds.), *The neuropsychology of high-level vision* (pp. 33-61). Hillsdale, NJ: Erlbaum.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, *18*, 1–63.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3D objects. *Nature*, *343*, 263-266.

Poggio, T., & Shelton, C. R. (1999). Machine learning, machine vision, and the brain. *AI Magazine*, *20*, 37-55.

Pouget, A., Deneve, S. & Sejnowski, T.J. (in press). Frames of reference in hemineglect: a computational approach. In J. Reggia, E. Ruppin, & D. Glanzman (Eds.), *Neural modeling of brain disorders*. Elsevier.

Pouget, A., & Sejnowski, T. J. (1997). A new view of hemineglect based on the response properties of parietal neurons. *Philosophical Transactions of the Royal Society: Series B*, 352, ???-???.

Reisenhuber, M. & Poggio, T. (1999) Hierarchal models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019-1025.

Reuter-Lorenz, P. A., Drain, M., Hardy-Morais, C. (1996). Object-centered attentional biases in the intact brain. *Journal of Cognitive Neuroscience*, *8*, 540–550.

Riddoch, M. J., Humphreys, G. W., Luckhurst, L., Burroughs, E., & Bateman, A. (1995). “Paradoxical neglect”: Spatial representations, hemisphere-specific activation, and spatial cueing. *Cognitive Neuropsychology*, *12*, 569–604.

Rock, I. (1973). *Orientation and form.* New York: Academic Press.

Sandon, P. A., & Uhr, L. M. (1988). An adaptive model for viewpoint-invariant object recognition. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 209–215). Hillsdale, NJ: Erlbaum.

Seibert, M., & Waxman, A. M. (1990). Learning aspect graph representations from view sequences. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2* (pp. 258-265). San Mateo, CA: Morgan Kaufmann.

Shillcock, R., & Cairns, P. (1995). A connectionist model of hemispheric interaction in unilateral visuospatial neglect. In J. D. Moore & J. F. Lehman (Eds.), Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society (pp. 142–147). Hillsdale, NJ: Erlbaum Associates.

Sperling, G. (1970). Binocular vision: A physical and a neural theory. *American Journal of Psychology*, *83*, 461–534.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, *262*, 658–688.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109-139.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, *1*, 253–256.

Tarr, M. J. (1995). Rotating object to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, *2*, 55–82.

Tarr, M. J. (1999). News on views: Pandemonium revisited. *Nature Neuroscience*, *2*, 932-935.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1494-1505.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, *1*, 275-277.

Tipper, S. P., & Behrmann, M. (1996). Object-centered not scene-based visual neglect. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1261–1278.

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97–159.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, *32*, 193-254.

Vecera, S. P. (1993). Object knowledge influences visual image segmentation. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 1040–1045). Hillsdale, NJ: Erlbaum.

Vecera, S., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, *123*, 146–160.

Vetter, T., Hurlbert, A., & Poggio, T. (1995). View-based models of 3D object recognition: Invariance to imaging transformations. *Cerebral Cortex*, *5*, 261–269.

Wallis, G., & Rolls, E. T. (1997) Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51*, 167-194.

Weinshall, D., Edelman, S., & Bülthoff, H. H. (1990). A self-organizing multiple-view representation of 3D objects. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2* (pp. 274-281). San Mateo, CA: Morgan Kaufmann.

Zemel, R. S., Mozer, M. C., & Hinton, G. E. (1988). TRAFFIC: Object recognition using hierarchical reference frame transformations. In D. Touretzky (Ed.), *Advances in neural information processing systems II* (pp. 266–273). San Mateo, CA: Morgan Kaufmann.