Do Visual Attention and Perception Require Multiple Reference Frames? Evidence from a Computational Model of Unilateral Neglect

Michael C. Mozer
Department of Computer Science and Institute of Cognitive Science
University of Colorado
Boulder, CO 80309–0430

Abstract

A key question motivating research in perception and attention is how the brain represents visual information. One aspect of this representation is the coordinate or *reference* frame with respect to which visual features are encoded. 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 (Mozer, 1991; Mozer & Sitton, 1998), we argue that this inference is not logically necessary: object-based attentional effects in neglect can be obtained without object-based frames of reference.

Introduction

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. 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.

Determining which reference frame or frames are used by the brain to encode visual features is a key step to understanding the mechanisms of visual cognition and attention. For this reason, there has been intense interest in neurological patients with unilateral neglect, who provide a rich source of data diagnostic of the reference frames involved in human perception.

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). 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? Clever behavioral experiments have been designed to dissociate various reference frames and determine the contribution of each to neglect. In several 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 attentional effects*, the inference is often made that attentional allocation must be operating in an object-based frame of reference, and consequently, object-based representations are key to visual information processing. The point of this paper is to show that this inference is not logically necessary: *Object-based attentional effects can be obtained without object-based reference frames*. We argue this point via a computational model that utilizes only viewer-based frames, yet can account for data from experimental studies that were interpreted as supporting object-based frames.

MORSEL

MORSEL (Mozer, 1991; Mozer & Sitton, 1998) is a connectionist model of visual perception and attention, which has previously been used to explain a large corpus of experimental data, including reading deficits in neglect dyslexia (Mozer & Behrmann, 1992), and line bisection performance in neglect (Mozer, Halligan, & Marshall, 1997). MORSEL (Figure 1) includes a *recognition network* that can identify multiple shapes in parallel and in arbitrary locations of the visual field, but has capacity limitations. MORSEL also includes an *attentional mechanism* that determines where in the visual field to focus processing resources.

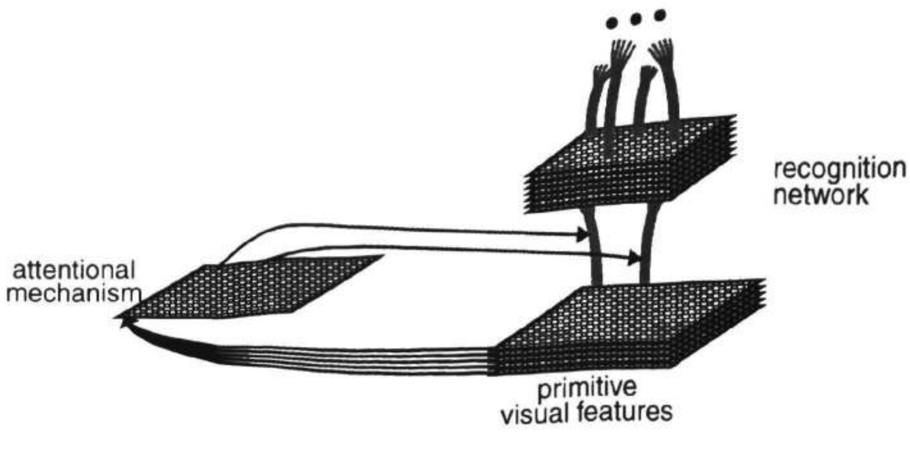

FIGURE 1. Key components of MORSEL (Mozer, 1991) include a recognition network, the first stages of which are depicted against a grey background, and an attentional mechanism.

Visual input presented to MORSEL is encoded by a set of feature detectors arrayed on a topographic map. Activity from the topographic map propagates through both the recognition network and the attentional mechanism. The topographic map is in a viewer-based reference frame, meaning that the input representation changes as the viewer moves through the world.

The attentional mechanism

In the present work, only the attentional mechanism, or *AM* for short, is required to account for data from unilateral neglect. The AM 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 1 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.

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 1 by the connections from the primitive features to the AM). Given the exogenous input, cooperative and competitive dynamics within the AM cause a subset of locations to be activated.

Figure 2 shows an example of the AM in operation. Although the model appears to have formed a spotlight of attention, the dynamics of the model do not mandate the selection of a contiguous or convex 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.

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:

(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 2, where the initial activity of the AM (upper-middle frame) is based on the exogenous input (upper-left frame).

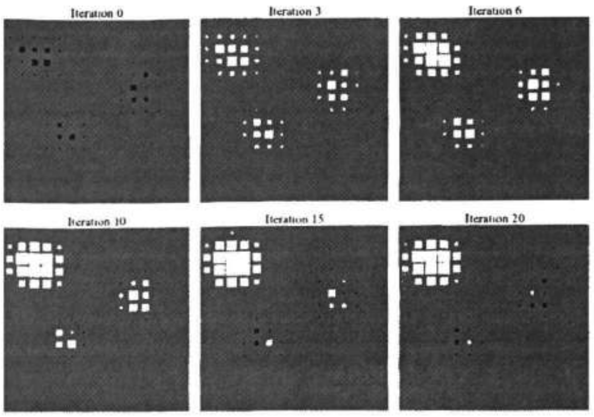

FIGURE 2. Example of the operation of the AM. Each panel contains a 15×15 topographic map depicting the state of the AM at a particular processing iteration The area of a black square is proportional to the exogenous input at that location. The area of a white square is proportional to the AM activity. 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.

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

(3) *Locations whose activity grows the slowest should be suppressed.* This rule results in *competition* between units, and is manifested in Figure 2 by the decrease in activity for the two lower blobs once the upper-left blob begins to dominate in activity. This rule allows a large region to become activated, if the activity of all units in the region rises at more-or-less the same rate.

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 (Mozer, 1991, 1999).

Lesioning the AM to produce neglect

In our earlier modeling of neglect, 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 will throughout this article). The graded damage we propose is inspired by Kinsbourne's (1987) orientational bias account of neglect. The damage affects the probability that primitive visual features are detected by the AM. The specifics of the damage are described in Mozer (1999). We emphasize that the damage is to a viewer-centered representation of space.

Simulations 1 and 2

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 3a) such that the two disks exchanged places (Figure 3b). 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 3b). 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.

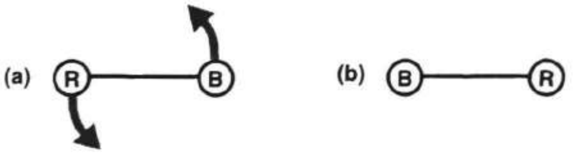

FIGURE 3. Barbell stimulus used in the Behrmann and Tipper experiment. The disk labeled "R" is colored red, the disk labeled "B" blue. In the moving condition, the initial display (panel a) was rotated 180°, resulting in the left and right disks exchanging places (panel b). In the static condition, no rotation occurred (panel b).

Tipper and Behrmann (1996) showed that the phenomenon appeared to depend on the disks being encoded as one object: in contrast to the condition depicted in Figure 3 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.

General simulation methodology

The AM as described is identical to the model used in our earlier simulation studies of neglect. Because our earlier simulations involved static displays and the present simulations involve dynamic displays, one minor technical change was made to the nature of input noise, as described in Mozer (1999).

To simulate an experimental task, the experimental stimuli are 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 generating a pattern of activity based on the silhouette, emphasizing the object borders.

The experimental task has as its dependent variable the response time to detect 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 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. In all results reported, we average activity over a window of 20 iterations following target onset, over all locations of the

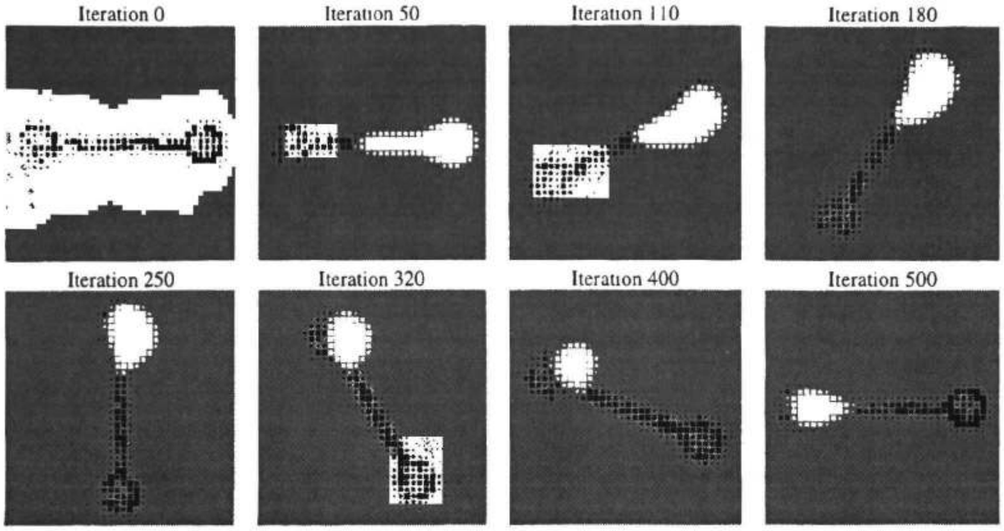

FIGURE 4. One trial of the lesioned model on the Behrmann and Tipper (1994) rotating-barbell stimulus.

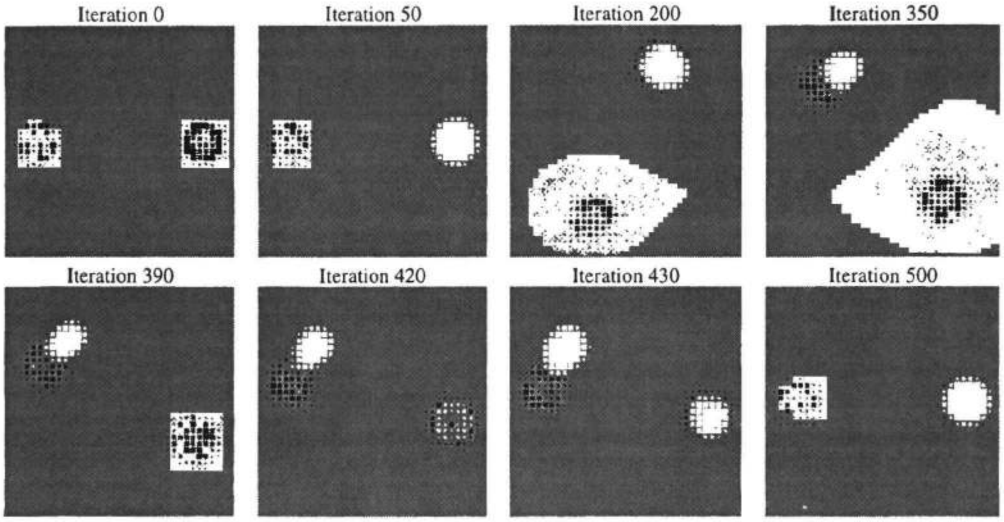

FIGURE 5. One trial of the lesioned model on the Tipper and Behrmann (1996) rotating disconnected disks.

disk on which a target appears, and over 200 trials. We call this the *read-out activity.* Greater read-out activity for a disk indicates a shorter response time to the target appearing in that disk.

Results

For the unlesioned model, the read-out activity on left and right disks and in the static and moving conditions is the same. The lesioned AM shows quite different behavior (Figure 4). A relative degradation to the exogenous input on the left side of the barbell can be observed due to the damage, 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. As rotation continues, attentional activity lags slightly behind the exogenous input, 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 4 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 5). 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 kicks in 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. (4) For patients, there is a main effect of target side: left is slower than right.

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, 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: the state at some iteration is a function of both the exogenous input and the state at the previous iteration. Attention is not ordinarily 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 number of activation rules that support it. 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 the disconnected-disk trials. Based on an exploration of alternative stimuli, it appears that the flipping occurs for the disconnected but not connected trials due to the presence of the *neck* of the barbell on connected trials—the region where the disk makes contact with the bar. The neck provides an region of exogenous input adjacent to the disk, and by the cooperation rule, therefore provides an environment that supports attentional activity. Figure 4 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.

TABLE 1

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

Simulation 3

Recently, Behrmann and Tipper (1999) have explored an intriguing variation of the rotating-barbell experiment in which the display contains, in addition to the rotating barbell, two elements—squares which remain stationary during the trial (Figure 6). 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.

Without delving into the details of the simulation, Figure 7 presents a single trial of the lesioned AM which gives an intuition of how the model can explain the data. 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 as the disk and square separate, the attentional blob connecting them is broken into two blobs. One might expect one blob to be suppressed due to competition between the blobs, but the competition is weak, for the following reason. The competition rule depends on the *rate* of activity growth in one blob versus another, not the total activity in one blob versus another. By the time the two blobs are formed, the activities of individual locations in the two blobs are comparable and near asymptote. Consequently, the competition rule does not produce significant inhibition of either blob. Quantitative measures of read-out activity are consistent with the example presented in Figure 7 and with the

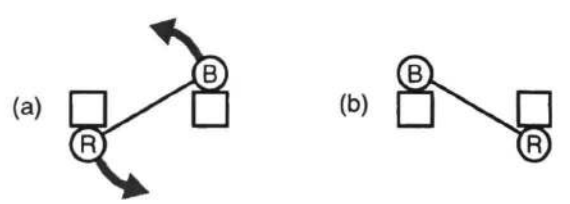

FIGURE 6. The multiple-object display studied by Behrmann and Tipper (1999). In the moving condition, the initial display (panel a) consists of two stationary squares and a barbell, which—as in the earlier studies—rotates such that its two disks exchange horizontal positions (panel b). Subjects were asked to detect a target that could appear on either disk or either square. In the static condition, no rotation occurred (panel b).

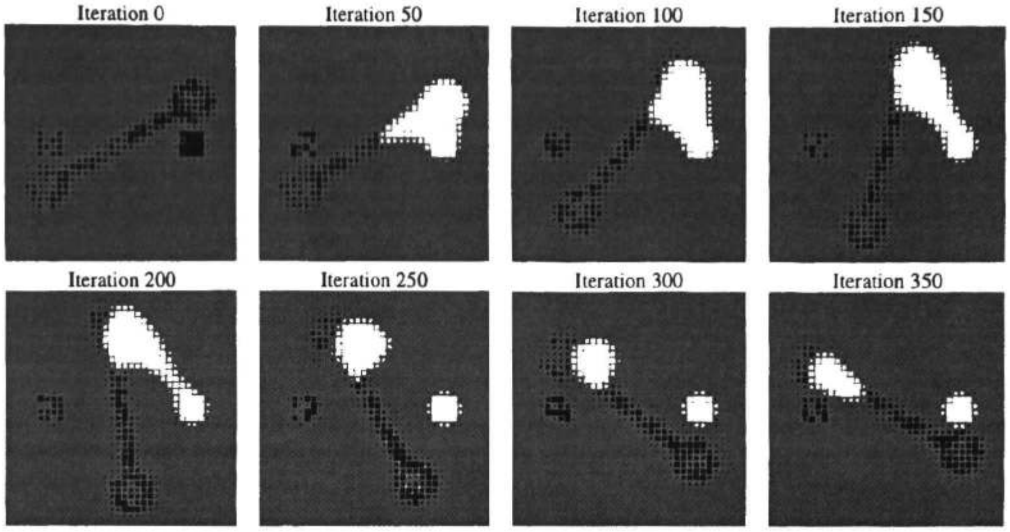

FIGURE 7. One trial of the lesioned model on the Behrmann and Tipper (1999) barbell-square experiment.

results of Behrmann and Tipper (see Mozer, 1999, for details).

Discussion

The neuropsychological studies of Behrmann and Tipper have been taken as support for the hypothesis that neglect, and hence attention, can operate in object-based coordinates. However, simulations of the AM provide an alternative explanation, because the AM operates only in a viewer-based frame of reference, yet can account for the data. The AM's account involves covert attentional tracking. But the AM's account is more complex, because the AM can in addition explain the lack of neglect reversal for disconnected displays or for the stationary squares.

The AM has also been successful in explaining a variety of other neglect studies that have been interpreted to support the psychological reality of object-based frames of reference (Mozer, 1999). The accumulation of such studies has caused many researchers to believe in the existence of object-based reference frames and object-based representations in the brain. But the AM strikes a blow against this interpretation. In the absence of strong empirical support for object-based frames, it would seem more parsimonious to suppose that object-based frames of reference play little or no role in the course of ordinary perception.

Acknowledgments

This work greatly benefited from the comments and insights of Paul Smolensky, Brenda Rapp, John McGoldrick, and Yuko Munakata. This research was supported by Grant 97-18 from the McDonnell-Pew Program in Cognitive Neuroscience, and by NSF award IBN-9873492.

References

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.

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

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.

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

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

Mozer, M. C. (1999). Frames of reference in unilateral neglect and spatial attention: A computational perspective. *Submitted for publication.*

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.

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