CHAPTER

23

Space- and Object-Based Attention

Michael C. Mozer, Shaun P. Vecera

ABSTRACT

Behavioral studies of visual attention have suggested two complementary modes of selection. In a space-based mode, locations in the visual field are selected; in an object-based mode, organized chunks of visual information—roughly, objects—are selected, even if the objects overlap in space or are spatially discontinuous. Although the two modes are distinct, they can operate in concert to influence the allocation of attention. This chapter presents key experimental results on space- and object-based attention and their interaction, and sketches a theoretical framework in which the two attentional modes can be unified. This chapter also discusses alternative notions of object-based attention, from perceptual grouping of low-level features in a retinotopic reference frame to construction of structural descriptions, and argues that the data are consistent with the former—a simple, low-level mechanism.

I. TWO MODES OF ATTENTIONAL SELECTION

Behavioral studies of visual attention have suggested two distinct and complementary modes of selection, one involving space and the other objects. In a space-based mode, stimuli are selected by location in the visual field (e.g., Ericksen and Hoffman, 1973; Posner, 1980). Evidence for this mode has come from a variety of sources, including spatial pre-cuing tasks, in which an abrupt luminance change (a *cue*) summons attention to a region in space. Observers are faster to detect or identify a subsequent target that appears at the cued location than one that appears at an uncued location (Posner, 1980). The space-based mode of attention has given rise to the attention-as-a-spotlight metaphor, in which attention acts as a beam to illuminate a contiguous region of the visual field. More recently, a zoom-lens metaphor has been suggested (Eriksen and Yeh, 1985), consistent with the finding that the region of space selected by attention can vary in size.

In contrast to the space-based mode, evidence has also been found for an object-based mode in which attention is directed to organized chunks of visual information corresponding to an object or a coherent form in the environment, even if objects overlap in space or are spatially discontinuous. All visual features of an attended object are processed concurrently, and features of an attended object are processed faster and more accurately than features of other objects. In one well-known task (Duncan, 1984), observers view two overlapping objects, a box and a line. Each object varies on two feature dimensions: the box is short or tall and has a gap on its left or right side; the line is dotted or dashed and tilts to the left or right. Observers are instructed to report pairs of features. Observers are more accurate at reporting two features of the same object (e.g., the height and side of gap of the box) than two features that belong to different objects (e.g., the height of the box and the tilt of the line). The cost in accuracy cannot be attributed to spatial factors, because the two objects overlap in space; rather, the cost must be attributed to the switching of attention from one object to the other. Indeed, Vecera and Farah (1994) have shown that no additional cost is incurred if the two objects are separated in space, suggesting that spatial factors are not at play in the object-based deployment of attention.

Some studies have shown that both spatial and object factors can simultaneously influence the allocation of attention. Egly et al. (1994) presented displays containing two rectangles (Fig. 23.1a). One end of one of the rectangles is cued with a brief flicker (Fig. 23.1b); a target then appears, and observers make a key-press response to the appearance of the target. The target

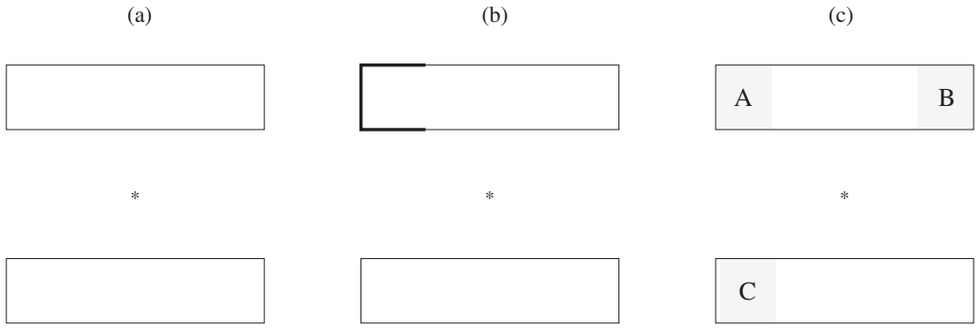

FIGURE 23.1 The Egly et al. (1994) experiment: (a) start of trial, (b) cue, (c) possible target locations.

appears either at the cued location, at the uncued end of the cued object, or in the uncued object (locations A, B, and C in Fig. 23.1c). Observers show a cue validity effect: detecting a target at the cued location is fastest. This result can be interpreted in terms of ordinary space-based attention. However, although the distance from the cued location to the target is the same for the noncued target locations, observers are faster to detect targets at the uncued end of the cued object (B) than those in the uncued object (C), indicating that the cue summoned attention to the entire cued object. This object-based effect is nonetheless modulated by spatial proximity: If the two objects are moved close together, the object effect is reduced in magnitude (Vecera, 1994). Further interactions between space- and object-based attention have been found via demonstrations that object-based effects occur only within the focus of spatial attention (Lavie and Driver, 1996), and the outputs of preattentive grouping processes influence the allocations of spatial attention (e.g., Baylis and Driver, 1992; Kramer and Jacobson, 1991; Logan, 1996).

II. CLARIFYING THE NOTION OF OBJECT-BASED ATTENTION

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: An object-based effect is observed in any experimental study in which attentional allocation or performance depends not merely on the location of an object in space, but on the extent, shape, or movement of the object itself. "Object based" can also be a characterization of processes and internal representations. Object-based representations arise from processes that use object-based frames of reference to transform visual features to achieve partial or complete view invariance. Object-based effects do not require object-based representations or frames of reference (Mozer, 2002; Vecera, 1994).

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 attentional mechanisms and representations might be considered object based. Examples of at least four alternatives can be found in the literature. Ordered from weakest to strongest notions of object based, these alternatives are as follows. (See Driver (1999) for a similar enumeration.)

1. *Grouping in a viewer-based frame* (Grossberg and Raizada, 2000; Mozer et al., 1992; Vecera, 1994; Vecera and Farah, 1994). Attention might act to select the set of locations in which visual features of an object are present. The resulting segmentation has been referred to as a *grouped array representation* (Vecera, 1994), because visual features are coded in a viewer-centered (e.g., retinotopic) array of locations, and labeled to indicate their grouping (Fig. 23.2). The segmentation can be achieved via heuristics, such as the Gestalt grouping principles, or might exploit low-order statistical regularities in visual scenes (Mozer et al., 1992). Whole-object knowledge is not required, nor are object-based frames of reference.
2. *Grouping and determination of principal axis* (Driver, 1999). In addition to performing segmentation in a viewer-based frame, attentional processes 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, the shape in Fig. 23.3a evokes a principal axis from which the midline of the shape can be

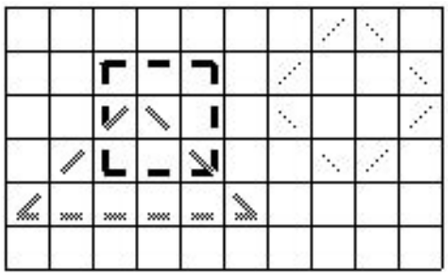

FIGURE 23.2 Illustration of the grouped array representation. Grouping of the visual features is indicated by their shading.

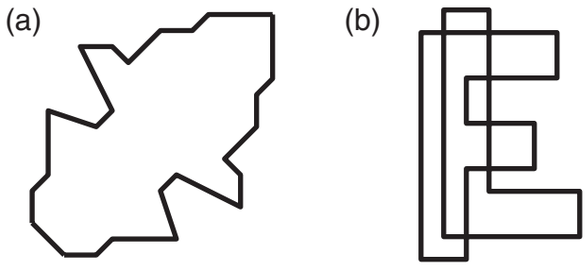

FIGURE 23.3 (a) A shape and its principal axis. (b) Sample stimulus from Vecera and Farah (1997).

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" can be derived from the viewer-based frame.

3. *Grouping and determination of an object-based frame of reference* (Marr and Nishihara, 1978). Attentional processes 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. It is difficult to imagine that this process could operate in a purely bottom-up fashion; most likely contact with object representations stored in long-term visual memory would be necessary.
4. *Grouping and determination of a structural description* (Biederman, 1987). For complex, articulated objects, attention might operate on a structural description of an object—a description that decomposes an object or scene into its parts, and that characterizes the relationships among the parts in terms of multiple allocentric frames of reference. Attention would then act to select a subtree of the structural description.

These alternatives vary on two dimensions: the reference frame in which object-based attention operates—egocentric for (1) and (2), allocentric for (3) and (4)—and on the degree of interaction with object knowledge required for selection—from weak, low-order statistics for (1) to high-order object knowledge for (4). The key issue is the degree to which object-based attentional effects require the explicit computation of object properties, such as a principal axis, frame of reference, or structural description, and the degree to which the data mandate such computations. All four alternatives have been invoked in the literature to explain object-based attentional effects.

III. EVIDENCE POINTS TO ONE ACCOUNT OF OBJECT-BASED ATTENTIONAL SELECTION

In terms of the computations required, the grouped array account of object-based selection is the simplest and can operate at the earliest point in the visual processing stream, and hence should be preferred on the grounds of parsimony. Mozer (2002) and Vecera (1994) have argued that a variety of experimental results seeming to call for more complex accounts of object-based attention can in fact be explained in terms of the grouped array account, and object-based representations and contact with object knowledge are not required. Although the existence of multiple attentional processes raises the possibility of multiple types of object-based selection, the grouped array account can explain most, if not all, of the object-based attention literature.

A complete account of object-based attention in terms of grouped arrays requires an understanding of the cues that determine grouping and, hence, that specify objects. The Gestalt cues of similarity, closure, and connectedness are all grouping cues that can influence the allocation of attention. For example, Kramer and Jacobson (1991) reported that a target that was physically connected to adjacent flanking items was attended as a single unit or group; a target that was not connected to the flankers could be selectively attended with little influence from the surrounding flankers (also see Baylis and Driver, 1992). Thus, grouping cues, including grouping, by similarity, connectedness, and good continuation, can determine which stimuli or visual features are attended simultaneously.

Grouping cues need not be primitive and innate; they might also be learned and influenced by familiarity with a visual environment. For example, Vecera and Farah (1997) presented displays consisting of two overlapping outline block letters, in either the upright or the inverted position (Fig. 23.3b). Observers were

asked to determine whether an "X" in the display was contained in one of the forms or in neither of the forms. Response times were faster for upright displays than inverted displays, suggesting that familiarity with upright letter forms was at play in the allocation of object-based attention. This result seems to argue against an account that relied solely on Gestalt grouping cues for segmentation, but might nonetheless be explained by the existence of adaptive grouping mechanisms that exploit low-order statistical regularities in the environment (Zemel et al., 2002).

IV. RELATIONSHIP BETWEEN SPACE-BASED AND OBJECT-BASED ATTENTION

Although many grouping cues have been identified, the relationship between these cues, which direct attention to objects, and the factors directing attention to locations in space has proved elusive. Initially, evidence for both space-based and object-based attentional selection led to a debate about whether selection was object based or space based. The current consensus is that both of these attentional modes coexist in the visual system and may influence one another. Despite this emerging consensus, many studies continue to address the relationship between these two modes of selection as if one mode of selection is more important than the other. For example, Lavie and Driver (1996) suggested a space-then-object account by demonstrating that object-based effects occur only within the focus of spatial attention. Other theorists have argued for an object-then-space relationship in which that the outputs of preattentive grouping processes influence the allocation of spatial attention (e.g., Baylis and Driver, 1992; Kramer and Jacobson, 1991; Logan, 1996).

Neither of these accounts is completely satisfactory. For example, data exist that appear inconsistent with the object-then-space account. Mack et al. (1992) demonstrated that if spatial attention is occupied at fixation by a visually demanding discrimination, grouping fails. If grouping occurred before attention, then occupying spatial attention with a demanding task should not impair object-based grouping. Similarly, a space-then-object account has difficulty explaining results from object-based attention studies. In many studies (e.g., Kramer and Jacobson, 1991), observers are instructed to perform a discrimination on a centrally presented target item. With such instructions, spatial attention should be focused centrally, and grouping outside this central region should not influence observers' responses to this central target.

The alternative to accounts supposing primacy of either space-based or object-based attention is an interactive account in which space- and object-based attentional processing operate simultaneously, each one helping to guide the other. However, interactive accounts face a serious computational problem. Object attention requires a search to partition visual features into objects; spatial attention requires a search for salient locations in the visual field. Each of these searches entails distinct and possibly conflicting computational objectives and, hence, incompatible solutions. Consequently, the reality of interactive accounts is that they are tricky to implement: The solutions reached are often suboptimal, where each search converges but the two outcomes are inconsistent with one another and each is suboptimal within its own domain (Hinton and Lang, 1985; Mozer et al., 1992).

Thus, a significant challenge lies ahead to unify mechanisms of space- and object-based attention. The grouped array view of object-based attention provides one key insight toward a coherent theory, via its proposal of a common substrate for the two varieties of attention: a topographic, viewer-based representation of space, often referred to in the attentional modeling literature as a saliency map. Another key insight concerns the role of strategic control. Because one form of attention does not always dominate over the other, it is likely that task demands and stimulus structure influence the relative contribution of each form of attention. Thus, a complete theory of attention requires claims concerning the processes by which the flexible attentional system is configured to operate for a given task in a given environment.

V. TOWARD A UNIFIED THEORY OF SPACE- AND OBJECT-BASED ATTENTION

Rather than viewing space-based and object-based attention as two qualitatively different forms of attention, unification is possible by conceptualizing attentional processing as fundamentally aimed at grouping related locations in the visual field. One can think of space-based and object-based processes as providing weak constraints concerning which locations belong together: Space-based attention suggests that adjacent, contiguous locations be grouped; object-based attention suggests that locations containing visual features likely to belong to the same object be grouped. The attentional state is then determined by a constraint-satisfaction search that attempts to identify groupings that are consistent with as many of the suggestions as possible. Thus, the operation of attention is viewed as

a single search, not two searches with distinct goals; this unification is possible via a shared representation of space.

This view suggests a weaker role for grouping processes than is ordinarily considered. Grouping processes can be heuristic and spatially local and can operate on multiple dimensions (e.g., color, shape) independently, and the global attentional state results from resolving the assorted grouping constraints with the space-based constraints.

Given the contribution of constraints from many different processes converging in attentional selection, the weighting of constraints becomes a key issue. Space-based attentional states result when space-based constraints dominate; object-based attentional states result when object-based constraints dominate. Because the data suggest the nature of the task and the stimulus display can influence which form of attention dominates, it seems natural to suggest that the weighting of constraints is under strategic control. One particularly elegant, computationally motivated form of control might involve selection of task- and environment-specific weightings that yield optimal performance, e.g., minimal response time or maximal accuracy. For example, reinforcement learning (Sutton and Barto, 1998) might be used to fine-tune the operation of the attentional system to achieve optimal performance. Beyond the virtue of integrating space-based and object-based attention, this perspective has the additional potential virtue of explaining integration of the various and diverse Gestalt grouping cues in determination of the attentional state.

References

Baylis, G. C., and Driver, J. (1992). Visual parsing and response competition: the effect of grouping factors. *Percept. Psychophys.* **51**, 145–162.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94**, 115–147.

Driver, J. (1999). Egocentric and object-based visual neglect. *In* "The Hippocampal and Parietal Foundations of Spatial Cognition" (N. Burgess, K. J. Jeffery, and J. O. Keefe, Eds.), pp. 67–89. Oxford Univ. Press, New York.

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

Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* **113**, 501–517.

Egly, R., Driver, J., and Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *J. Exp. Psychol. Gen.* **123**, 161–177.

Eriksen, C. W., and Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Percept. Psychophys.* **14**, 155–160.

Eriksen, C. W., and Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *J. Exp. Psychol. Hum. Percept. Perform.* **11**, 583–597.

Grossberg, S., and Raizada, R. D. S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Res.* **40**, 1413–1432.

Hinton, G. E., and Lang, K. J. (1985). Shape recognition and illusory conjunctions. *In* "Ninth Annual Joint Conference on Artificial Intelligence," pp. 252–259. Morgan–Kaufmann, Los Altos, CA.

Kramer, A. F., and Jacobson, A. (1991). Perceptual organization and focused attention: the role of objects and proximity in visual processing. *Percept. Psychophys.* **50**, 267–284.

Lavie, N., and Driver, J. (1996). On the spatial extent of attention in object-based visual selection. *Percept. Psychophys.* **58**, 1238–1251.

Logan, G. D. (1996). The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychol. Rev.* **103**, 603–649.

Mack, A., Tang, B., Tuma, R., Kahn, S., and Rock, I. (1992). Perceptual organization and attention. *Cogn. Psychol.* **24**, 475–501.

Marr, D., and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proc. R. Soc. London Ser. B* **200**, 269–294.

Mozer, M. C. (2002). Frames of reference in unilateral neglect and visual perception: a computational perspective. *Psychol. Rev.* **109**, 156–185.

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

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* **32**, 3–25.

Sutton, R. S., and Barto, A. G. (1998). "Reinforcement Learning." MIT Press, Cambridge, MA.

Vecera, S. P. (1994). Grouped locations and object-based attention: comment on Egly, Driver, and Rafal (1994). *J. Exp. Psychol. Gen.* **123**, 316–320.

Vecera, S. P., and Farah, M. J. (1994). Does visual attention select objects or locations? *J. Exp. Psychol. Gen.* **123**, 146–160.

Vecera, S. P., and Farah, M. J. (1997). Is visual image segmentation a bottom-up or an interactive process? *Percept. Psychophys.* **59**, 1280–1296.

Zemel, R. S., Behrmann, M., and Mozer, M. C. (2002). Experience-dependent perceptual grouping and object-based attention. *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 202–217.