Journal of Experimental Psychology: Human Perception and Performance
August 1998 Vol. 24, No. 4, 1011-1036

Object-Based Attention and Occlusion
Evidence From Normal Participants and a Computational Model

Marlene Behrmann
Department of Psychology Carnegie Mellon University
Richard S. Zemel
Department of Psychology University of Arizona
Michael C. Mozer
Department of Computer Science University of Colorado at Boulder

ABSTRACT

One way of perceptually organizing a complex visual scene is to attend selectively to information in a particular physical location. Another way of reducing the complexity in the input is to attend selectively to an individual object in the scene and to process its elements preferentially. This latter, object-based attention process was examined, and the predicted superiority for reporting features from 1 relative to 2 objects was replicated in a series of experiments. This object-based process was robust even under conditions of occlusion, although there were some boundary conditions on its operation. Finally, an account of the data is provided via simulations of the findings in a computational model. The claim is that object-based attention arises from a mechanism that groups together those features based on internal representations developed over perceptual experience and then preferentially gates these features for later, selective processing.

This research was supported by Grant MH47566-07 from the National Institute of Mental Health, the National Science Foundation's Presidential Young Investigator's Award IRI-9058450, Grant 90-21 from the James S. McDonnell Foundation, and Digital Equipment Corporation External Grant 1250. Portions of this article were presented at the 35th Annual Meeting of the Psychonomic Society, St. Louis, Missouri, November 1994. We thank David Plaut, Allison Sekuler, and Steve Yantis for their helpful comments and Susan Watt, Reza Habib, Alice Wong, John McGoldrick, and Jim Nelson for help with data collection.
Correspondence may be addressed to Marlene Behrmann, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213-3890.
Electronic mail may be sent to behrmann+@cmu.edu
Received: March 8, 1995
Revised: March 20, 1997
Accepted: May 6, 1997

Humans are exceptionally good at recognizing objects in natural visual scenes despite the fact that such scenes usually contain multiple, overlapping objects. One way in which individuals organize this complex input to minimize the processing load is to divide the field on the basis of spatial location and then to attend selectively to particular physical regions. This selective attentional spotlight "illuminates" areas of interest and facilitates preferential processing of information from those chosen areas (e.g., Broadbent, 1982 ; B. A. Eriksen & Eriksen, 1974 ; C. W. Eriksen & Yeh, 1985 ; Posner, 1980). There is now much evidence supporting this location-based selection, all of which shows that information from selected regions is processed faster and more accurately than equivalent information from unattended regions (Posner, 1980 ; Posner, Snyder, & Davidson, 1980). The idea that location-based selection plays an exclusive role in organizing visual information, however, has been increasingly challenged in recent years. Studies have shown, for example, that humans can select one of two superimposed figures even when there is no spatial basis for selection (Rock & Gutman, 1981) and can allocate attention to perceptual groups independent of the spatial proximity and contiguity of the component elements (e.g., Behrmann, Vecera, & McGoldrick, 1998 ; Driver & Baylis, 1989 ; Duncan, 1984 ; Kramer & Jacobson, 1991 ; Kramer & Watson, 1995 ; Lavie & Driver, 1996 ; Prinzmetal, 1981 ; Vecera & Farah, 1994). To account for these findings, an alternative selection process, in which attention is directed to objects, rather than to locations or unsegmented regions of space, has been proposed. This object-based mechanism, in which complex visual input is parsed into discrete units for further processing, has received considerable empirical, neuropsychological, and computational support in recent years.

Object-Based Visual Attention

An early but compelling empirical illustration of the view that attention can be directed to objects, rather than to spatial locations per se, comes from work by Duncan (1984) . In these studies, participants were shown displays consisting of an outline box on which a diagonal line was superimposed, thereby occupying roughly the same spatial region as the box. Participants were then required to make judgments about two features that were present in the display, both of which appeared on the same object (e.g., line orientation and texture from the diagonal line, or box size and gap side from the box) or one of which appeared on each of the two different objects (e.g., line orientation and box size). The critical result was that participants showed a cost in accuracy in reporting the features from the two different objects compared with features from a single object. Indeed, Duncan (1984 , (1993) showed that, under simple conditions, participants could identify two properties of a single object just as accurately as they could identify one. Duncan interpreted all these results as favoring a view that the visual field can be segmented or parsed into separate objects and that attention can then be directed selectively to a single object, thereby facilitating the processing of its features (see also Lappin, 1967 ; Neisser, 1967). The two-object cost is attributed to the time taken to switch object-based attention between the two objects.

Although the results from these experiments are consistent with the object-based view, there are other possible explanations that might also account for these findings. For example, Posner (see Footnote 3 in Duncan, 1984) has suggested that if an attentional spotlight were to operate in three-dimensional (3-D) rather than two-dimensional (2-D) space, the overlapping box and line might be separable in depth and then attention may select one of the objects spatially in depth. A second concern is that Duncan's results may reflect a difficulty in attending to different spatial frequencies rather than to different objects; whereas the two attributes of the line (texture and line orientation) are primarily available at high spatial frequencies and the two attributes of the box (height and gap) are available primarily at low spatial frequencies, the attributes of the box and of the line might be segregated not by object-based attention per se but by setting a spatial frequency filter at different levels of coarseness (see Baylis & Driver, 1992 , 1993 ; Lavie & Driver, 1996 ; and Watt, 1988 , for further discussion).

Recent researchers have circumvented these potential shortcomings and have shown that the features of a single object can indeed be preferentially selected and processed. For example, in Lavie and Driver's (1996) study, participants judged whether two odd elements (e.g., two dots or a dot vs. a gap) in a display of two crossed dashed lines were the same or different. The results revealed an advantage for decisions of elements from a single line relative to two lines even when the spatial distance between the judged elements was wide and exceeded 8°. Neither a depth account nor a difference in spatial frequency could account for these results: The crossed lines were clearly 2-D in appearance and the elements to be judged were equivalent in spatial frequency. Similarly, when spatial frequency was controlled by Baylis and Driver (1993 ; also see Baylis (1994) , an advantage for a single object was also obtained. In these latter experiments, participants made position judgments about parts of a display when the display could be parsed as one or two objects depending on the participants' perceptual set. Using the physically identical display with attention directed to different components by color cues, Baylis and Driver found that the judgment of relative position of two parts was accomplished better when the two parts came from a one- rather than from a two-object display. Taken together, all these findings suggest that attentional selection can operate on an object-based description and that the results are not simply attributable to artifacts of the display.

The findings from these experiments are consistent with the idea of a selection process in which visual input can be organized by segmenting the image into discrete objects or groups. This mechanism need not, however, be mutually exclusive with a location-based mechanism. For example, Egly, Driver, and Rafal (1994) have demonstrated the coexistence of both location (space)- and object-based processes in the same participant. In studies with normal and brain-damaged participants (Egly, Driver, & Rafal, 1994 ; Egly, Rafal, Driver, & Starrveld, 1994), they examined the cost incurred during target detection when attention was initially cued to a particular location and when the target then appeared in a different location. In these latter uncued or invalid trials, the target could appear either in the same object as the initial precue (two locations within a single object) or in a different object (two locations between two objects), and the cost in accuracy of detection was measured. Relative to the validly cued trials, there was a cost when attention was switched between two spatial locations within the same object. Interestingly, there was an additional cost when attention was switched between two locations, each of which was occupied by a different object, and this was so even though the two between-objects locations were spatially closer than the two within-objects locations. Furthermore, Egly, Driver, and Rafal (1994) and Egly, Rafal, et al.(1994) provided evidence that the neural substrate mediating these two components of visual selection differed, with the left and right hemispheres likely subserving switches across objects and across space, respectively (see Kramer, Weber, & Watson, 1997 ; Vecera, Strayer, & Chamberlain, 1996 ; see also Lavie & Driver, 1996 , for ideas on how object- and location-based mechanisms may be reconciled).

Mechanisms Underlying Object-Based Attention

The existence of an object-based attentional process is now well accepted and no longer particularly controversial. What is still not obvious from this empirical work, however, is what exact mechanisms underlie this selection process. That features of objects benefit from selective attention implicates the existence of a process by which features belonging to the same object are bound or grouped together before they are selectively enhanced. Such a grouping process can account not only for the single-object advantage but also for the two-object cost: Once it is determined which features belong to which objects in the image, then judgments about features assigned to the same object are carried out more quickly than judgments about features belonging to different objects. We therefore posit that an essential component of object-based attention is feature grouping and that if researchers understand how grouping and perceptual organization operates, this will greatly inform their understanding of object-based attention. A second focus of this article, then (after the empirical work), was to examine the hypothesis that feature grouping mediates object-based attention.

One long-standing proposal of how perceptual grouping works is that the visual world is parsed preattentively into discrete chunks defined according to Gestalt principles of perceptual organization. Through this parsing process, elements that share continuity of, for example, contour, color, or movement are bound together and then attention is directed to these grouped components (Desimone & Duncan, 1995 ; Duncan, 1984 ; Neisser, 1967 ; Prinzmetal, 1981 ; Wertheimer, 1923/1955). The Gestalt heuristics provide the primitive grouping rules for linking together elements of a visual scene that likely belong together, reducing the complexity of the proximal stimulus and binding together components that can serve as input for object recognition. This process of decomposing or partitioning an image into coherent components that can be independently characterized is often referred to as *segmentation, grouping,* or parsing.

However, Gestalt rules of perceptual organization need not be the only possible principles that guide the parsing of elements of a display. Computer vision researchers have long been involved in developing grouping algorithms that use shape cues to organize image features into independent parts of a scene. One fairly common approach to choosing image features that belong to the same object is based on strategies that rely on particular local relations between primitive elements in the display. These elements are then grouped to form salient, coherent groupings in the image. For example, one metaheuristic involves the determination of nonaccidental regularities or feature combinations that are unlikely to occur by chance when several objects are juxtaposed (Kanade, 1981 ; Lowe & Binford, 1982 ; Witkin & Tenenbaum, 1983). The idea is that because the grouping of elements is not accidental and truly reflects interdependent elements, the algorithm provides a reliable means for segmenting the image.

In previous work, a possible computational mechanism in which feature grouping may be achieved in early vision was explored (Mozer, Zemel, Behrmann, & Williams, 1992). In particular, instead of simply assuming that grouping is driven by Gestalt rules or heuristic nonaccidental properties, Mozer et al. were interested in understanding what types of statistical regularities might be discovered by an adaptive model trained to segment images. The computational model–multiple-object adaptive grouping of image components (MAGIC)–was initially trained on a set of presegmented images containing two superimposed objects in which each elementary feature was labeled as to which object it belonged. Grouping of the features was performed by a relaxation network that attempted to bind related features. Over time and adaptively, MAGIC learned to detect configurations of the image features that had a consistent labeling in relation to one another across the training examples. When presented with novel displays after training, MAGIC successfully segregated the features into independent objects. These findings suggested that MAGIC had learned a number of cues that support accurate grouping of object fragments and had captured some important principles of segmentation or grouping. If object-based attention relies on robust feature grouping, then we would expect that MAGIC would show the single-object advantage and two-object cost in the same way as do the human participants. If so, this would provide insight into a potential mechanism underlying object-based attention and would support the central role of feature grouping. In this article, we replicate the data from the human empirical experiments in MAGIC and show that the computations embodied by MAGIC may serve as the key component in a model of object-based attentional effects.

Occlusion and Object-Based Attention

If grouping processes are indeed used to select objects as we suggest, an outstanding question concerns the potential limitations of such processes. Most researchers have used visual displays in which the features of an object are unobscured and therefore relatively easy to select. A more stringent condition arises when features from cluttered or more complex images need to be grouped and selected. A particularly difficult situation that would seem to challenge any kind of object-based grouping process is that of occlusion: Not only are some of the elements of a single occluded object obscured but those that are visible are often spatially distant and discontinuous. The crucial issue, then, is, in a display in which a single object is occluded such that it has two disparate parts (an amodal shape), do object-based processes work sufficiently well such that the disconnected parts can be bound together and preferentially enhanced? Furthermore, are the features of the occluded object integrated with the same speed and accuracy as a single, uninterrupted shape (modal shape), or is the participants' performance more akin to the two-object condition?

The problem of occlusion has a long history in the study of visual processing, and the facility with which a fragmented proximal stimulus is completed has been the focus of much research (see, e.g., Kanisza & Gerbino, 1982 ; Kellman & Shipley, 1991 , 1992 ; Koriat, 1994 ; Marr, 1977 ; Shimojo, Silverman, & Nakayama, 1989 ; Yantis, 1995). Not only can people easily identify objects that are occluded or that fall in the region of a visual scotoma (Ramachandran, 1992), but this completion process is rapid, automatic, and spatially parallel (Enns & Rensink, 1996 ; Nakayama, Shimojo, & Ramachandran, 1990). Although most studies of object completion endorse the ease with which fragmented objects are perceived and interpreted, only recently has there been concern with the nature of the representation subserving completion. Using a high-speed priming paradigm, for example, Sekuler and Palmer (1992 ; Sekuler, Palmer, and Flynn (1994) found that the representation of a partly occluded object changes over time and that the occluded object is represented fully only as a completed object at about 100—200 ms. An outstanding question about the final representation is whether the occluded object truly has the integrated status of a single, coherent object. If so, and if the object-based attention process is sufficiently robust to apply to occluded objects, then judging two features from each of the two noncontiguous parts of an occluded object should be done as well as judging two features from a single, uninterrupted object. If, however, the features of an occluded object are not strongly bound together, then reporting two features from an occluded object will not show the single-object advantage and will more closely parallel the two-object condition.

In this article, in a series of experiments with human participants, we demonstrate that object attention is indeed robust under conditions of occlusion and that identifying two features from noncontiguous parts of an occluded object is achieved as well as when the features come from a single object. This result, however, holds only under certain conditions; when the perceptual evidence no longer clearly supports the existence of a unified object, the features of the noncontiguous parts are not grouped into a single entity and are not afforded preferential processing (also see Kellman & Shipley, 1992 ; Yantis & Moore, 1995). We also show that MAGIC groups features from an occluded object as well as those from a single object and that, like the human participants, MAGIC is also sensitive to perceptual constraints: When the evidence is not consistent with the presence of a single, occluded object, MAGIC parses the discontinuous parts into two separate objects.

In summary, the goal of this article is to examine the processes that underlie object-based attention. We begin by empirically demonstrating the preferential enhancement afforded features of a single object and then probe the generality of these object-based attentional effects: In a series of behavioral experiments, we replicate the single-object advantage (and two-object cost) both for fully completed and for occluded objects when participants make decisions about local features of objects and when they make segmentation decisions at the more global object level. We also verify the representation mediating the occluded object in a task in which participants explicitly categorize the displays according to their phenomenological experience. Because all of these decisions are made in response to spatially overlapping stimuli, these findings support the claim that features of a single and an occluded object are processed preferentially by virtue of being grouped together into a coherent whole rather than by virtue of sharing a spatial location. We go on to show that this object-based attention process operates flexibly and generally over different types of displays but that limitations appear under certain perceptual conditions. We then demonstrate that the performance of the human participants can be simulated by an existing computational model, MAGIC, and that even under conditions of occlusion, a particularly stringent test of the grouping process, the performance of the model is fast and accurate. MAGIC serves not only as an existence proof of one potential mechanism mediating the grouping and selective enhancement of the grouped features but it also makes strong predictions concerning how humans should perform across a range of related tasks. Taken together, the findings from the empirical human studies and the computational simulations constrain our understanding of the object-based attention process.

Experimental Data From Humans

Experiment 1a

This first experiment had two main purposes. The initial goal was to replicate previous findings that participants perform more accurately when making judgments about two features of a single object than when making judgments about the same two features when they come from two different objects. Because the displays contain two objects that are spatially superimposed, selection on the basis of physical location cannot produce this result. Similarly, because there are no spatial frequency differences between the two objects in the display, this potential artifact also cannot explain the findings we obtain. The second goal was to examine object-based attention under a more stringent testing condition: Participants were required to make the same decisions about two features as they did on the single and two-object conditions discussed earlier, but now each of the two features was located on one of the noncontiguous portions of a single, occluded object. The question was whether feature judgments of the occluded object are made as quickly and as accurately as those made on the single, spatially coherent object.

Consider the displays made up of two overlapping rectangles as shown in Figure 1 . As is evident, at two ends of the four possible edges of the X figure, a set of features or "bumps" appears that is made of either two or three divisions of the end of the bar. The sets of bumps or features appear either at each of the two ends of a single object (e.g., Figures 1 a and 1d), at the ends of two separate objects (e.g., Figures 1 b and 1e), or at the ends of a single but occluded object (e.g., Figures 1 c and 1f). The features had either the same (e.g., Figures 1 a—1c) or different (e.g., Figures 1 d—1f) number of bumps, and the participants were required to indicate by a keypress whether the number of bumps was the same or different. We hypothesized that, consistent with the findings on object-based attention, participants would be able to judge the number of bumps on the ends of the single object without loss of accuracy or speed compared with the two-object display. Furthermore, if the object-based selection process was sufficiently robust, the superiority in making decisions about a single object would hold even under conditions in which only two, noncontiguous bars of a single object were observable because of occlusion by a second object (Figures 1 c and 1f). If this were so, performance on the occluded condition would be no different from that on the single-object condition.

Method Participants.

Seven men and 9 women (aged 19—24 years) were recruited from the introductory psychology subject pool at the University of Toronto. All received course credit for their participation. All had normal or corrected visual acuity by self-report, and all were right-handed.

Apparatus and materials.

The experiment was conducted on a Macintosh IIci computer. Stimuli were presented on a 13-in. (33.02 cm) color monitor using Psychlab experimental software version 1.0 (Bub & Gum, 1991). The displays contained two rectangles crossing each other in the center to form an X. On each trial, the features (bumps) appeared at two of the four ends of the two rectangles. The end was divided into two equal parts for the two-bump and into three equal parts for the three-bump displays. Examples of the displays appear in Figure 1 , with the rows and columns illustrating the different conditions and judgments (*same—different*), respectively.

The displays fell into three different conditions: (a) single (or unoccluded) *object,* in which the two sets of bumps appeared at each end of a single rectangular bar (e.g., Figures 1 a and 1d); (b) *two objects,* in which each of the two sets of bumps appeared at the end of bars belonging to two different rectangles (e.g., Figures 1 b and 1e); and (c) occluded object, in which the two sets of bumps appeared at each end of a single rectangular bar that was occluded (e.g., Figures 1 c and 1f).

There were an equal number of *same* and *different* judgments in each of the three conditions. On *same* trials, there were either both two bumps (known as a 2-2 trial; Figures 1 a—1c) or three bumps (known as a 3-3 trial) at the two ends and there were an equal number of 2-2 and 3-3 *same* trials. On *different* trials, there were always two bumps on one end and three bumps on the other (Figures 1 d—1f) and the locations of the two and three bumps were evenly counterbalanced.

The displays were presented as black-and-white line drawings on a white background. Viewing distance was approximately 50 cm. Each rectangular bar was 8.7 cm in length (10.2°) and 2.5 cm (2.9°) in width. The straight line drawn from the midpoint of one end of a rectangle to the midpoint of the adjacent rectangle either horizontally or vertically was 6.2 cm (7.8°). Note that the spatial distance between the bumps in the *single* and occluded conditions always exceeds that of the two-object condition. This manipulation ensures that any advantage afforded by spatial proximity worked against the single and occluded object and favored the two-object condition. On half the trials the single object was oriented from left to right as in Figure 1 , and on the remaining half the orientation of the bar was right to left. The orientation of the bar was crossed orthogonally with the other variables.

The participant's task was to decide whether the number of bumps on the two ends of any of the rectangular bars was the same or different. Responses were indicated with the Z or M keys with the left and right index fingers on the standard keyboard. The assignment of keys to *same* or *different* responses was counterbalanced across participants. Reaction times (RTs) to make the decision were recorded in milliseconds and accuracy noted.

Design.

The design was entirely within subjects, with the independent variables being condition (single, two, and occluded) and judgment (same or different). There was an equal number of trials drawn from each of the three conditions and an equal number of *same* and different trials in each of the two different orientations (slanting from left to right or vice versa). This core set of displays was replicated for a total of 288 trials.

Procedure.

Participants were shown a display that appeared on the computer screen and were told to make same—different judgments on the number of bumps as accurately and quickly as possible. The sequence of events on any one trial was as follows: A black fixation dot appeared in the center of the screen for 500 ms and then disappeared. After a delay of 1 s, the stimulus appeared, centered over the fixation point, and remained on the screen until a response key was pressed. An interval of 1 s followed the response and the sequence was repeated. The experiment was run in three blocks of 96 randomized trials, including appropriate crossing of all the variables, with a few minutes' break between blocks. Before starting the experiment, the participants were shown examples of the trials and completed a block of 24 practice items, including instances of all possible trials.

Treatment of results.

The data from the practice trials were discarded from the analysis. The data were collapsed across the three experimental blocks, and the error trials were excluded from the RT analysis. The median RT and mean error for each crossing of judgment, orientation, and condition were calculated for each participant and were then subjected to analyses of variance (ANOVAs). Post hoc comparisons, using a Tukey test with a probability level of .05, were conducted to evaluate pairwise differences, and this procedure was also used in all subsequent experiments.

Results and Discussion

Because there was no difference in RT patterns as a function of whether the modal object in the display was oriented to the left or right ($F < 1$), we pooled the data across the two orientations for the remainder of the analysis. The means of the participants' median RTs for the remaining six cells of this experiment are shown in Figure 2 , together with the associated mean error rates for each condition shown in parentheses.

A two-way ANOVA with judgment (*same* or *different*) and condition (single, occluded, and two) as within-subject variables was conducted on error rates and RTs. As is evident from Figure 2 , the error rates were low, constituting 1.9% of the total trials. The error rates were not affected significantly by the type of judgment, F (1, 15) = 0.54, p > .1, or by the condition of the display, F (1, 15) = 0.77, p > .1. For RTs, *same* judgments were significantly faster than *different* judgments by an average of 43 ms, F (1, 15) = 18.8, p < .001. More important, a highly significant difference was noted across conditions, F (2, 30) = 15.9, $p < .0001$. Planned pairwise comparisons using Tukey tests with a probability level of .05 revealed that, for both *same* and *different* judgments, responses to the single and occluded displays did not differ from each other but that both were significantly faster than responses to two-object displays. The equivalence between single and occluded and their difference from the two-object condition held to an equal extent across both same and *different* judgments, F (2, 30) = 0.4, p > .5.

The major findings of this experiment are clear. In a task in which participants were instructed to make judgments about local features of objects, participants' decisions were significantly influenced by whether these features appeared on one versus two objects. The single-object advantage is consistent with the results of many experiments (Baylis & Driver, 1993 ; Duncan, 1984 ; Vecera & Farah, 1994), showing a superiority in judging parts of one over two objects even when, as in our displays, the features within a single object were farther from each other than the features on the two different objects (see also Lavie & Driver, 1996). The novel contribution from this study, however, is that the time taken to judge the similarity of features of a single occluded object was not significantly different from that of a single, uninterrupted object. This finding suggests that participants treat the two discontinuous bars of an occluded object as though they were drawn from a single object rather than from two disconnected objects. The single- over two-object superiority and single and occluded equivalence were not accounted for by a speed—accuracy trade-off because the error rate was low, but, to the extent that it differed at all, there were fewer errors in the single and the occluded displays than in the two condition. These data, then, provide support for the object-based superiority in human participants in displays in which there are two overlapping outline geometric shapes. Moreover, the findings suggest that participants can attend to features of single objects preferentially even when the objects are partially occluded.

Experiment 1b

Although the data favor the view that object-based enhancement applies to single and occluded objects, but not to the two-object condition, the evidence that participants parse the display into two separable objects is somewhat indirect. In Experiment 1a, participants were making decisions on features that were superimposed on the contour of the objects. According to Lavie and Driver (1996) , elements superimposed on a contour need not form an integral part of the object and might be coded as distinct entities independent of the object. If this were the case and participants were simply attending to local features, it would be difficult to explain why we obtained the object advantage found in Experiment 1a. Nevertheless, to verify that the advantage for the single- and occluded-object conditions over the two-object condition would still be observed when participants attended to the whole display and made judgments (rather than comparisons about features) based on the entire display, we repeated the same experiment but altered the task instructions. In this experiment, participants made decisions about the objects themselves rather than about the local features. Using the identical displays and procedure as in Experiment 1a, in this next experiment we simply changed the instructions and told participants to ignore the exact number of the bumps and to indicate whether the bumps fell on the same object or on different objects. RTs and accuracy to make these *same—different* object judgments were recorded. If object-based attention provides superior processing of the elements of a single (occluded or not) object, then we should see the same results as those in Experiment 1a when the "read-out" of the task was now at the object rather than at the featural level.

Method Participants.

Eleven men and 5 women (aged 18—42 years, M = 25.2) were recruited via the bulletin boards at Carnegie Mellon University. All consented to participate and received payment for their participation. All had normal or corrected visual acuity by self-report, and all but 2 were right-handed.

Apparatus and materials.

The apparatus, stimuli, and material were identical to those used in Experiment 1a.

Design and procedure.

The timing and procedure were identical to Experiment 1a except that the participant's task was to decide whether the bumps fell on the same object or on two different objects. Responses were indicated with the *Z* or *M* keys with the left and right index fingers on the keyboard, and the assignment of keys to *same* or *different* object decisions was counterbalanced across participants. Reaction times to make the decision were recorded in milliseconds and accuracy was noted. Participants were told that the number of bumps was not relevant for their decision and should be ignored. The design, number of trials, practice procedure, and analysis were identical to those Experiment 1a. The data were collapsed across the three blocks and the error trials excluded. The median RTs and mean error rates for making *same* responses to the single and occluded condition and *different* judgments to the two-object condition were calculated. Even though the number of bumps was irrelevant, we included this variable in the analysis to determine whether there might be any interference from a mismatch in the number of bumps and the type of decision (e.g., when participants made a *same* -object decision on a *different* -bumps 2-3 display or a different -object decision on a *same* -bumps 2-2 display). Post hoc Tukey tests were conducted to examine pairwise differences.

Results and Discussion

Participants made fewer than 2% errors. An ANOVA on these error data showed that the number of bumps did not affect accuracy, F (1, 15) = 0.18, p > .5, nor did it affect the participants' ability to decide whether the features belonged to one object, F (2, 30) = 1.9, $p > .1$. The error rate, however, did differ as a function of condition, F (1, 15) = 6.3, $p < .01$, with significantly fewer errors being made when the bumps fell on a single object (.8%) than on an occluded object (1.6%) or with two different objects (2.9%). The latter two conditions did not differ significantly.

As was the case with errors, there was no difference in participants' RTs as a function of the number of bumps in the display (*different*, 2-3; same, 2-2 and 3-3 trials), F (1, 15) = 0.47, p > .1, nor did this interact with the judgment of whether the bumps fell on the same object, F (2, 30) = 0.12, p > .5. There was, however, a significant difference in RTs as a function of condition, *F* (2, 30) = 16.1, p < .0001. Pairwise comparisons revealed no difference between responses to single (*M* = 640.9 ms) and occluded (M = 656.9 ms) trials, *F* (1, 31) = 3.4, *p* > .05, but a significant difference between each of these and the two-object condition (M = 709.54 ms): single and two-object, F (1, 31) = 50.7, p < .0001; occluded and two-object trials, *F* (1, 31) = 16.9, *p* < .005.

The results of this experiment are compatible with the previous findings and endorse the view that features from single and occluded objects are preferentially and equally enhanced relative to features from two separate objects. The findings also indicate that the single-occluded superiority holds irrespective of whether the decision is made at the local level of the features or at a more global object level. This suggests that the system as a whole settles in favor of a particular interpretation of the display, that this interpretation is propagated through the system, and that it is upheld wherever the read-out occurs.

Experiment 1c

Although the results of Experiment 1b are strongly compatible with those from Experiment 1a and clearly favor the superiority of features from a single object, whether occluded or not, there is an alternative interpretation for these latter results. In Experiment 1b, the single and occluded displays were both assigned the response of *same,* whereas the two-object display was assigned the response of *different.* As is well-known, *same* responses are generally made faster than *different* responses (Nickerson, 1965), and so the observed single-occluded object advantage might simply be attributable to a response advantage rather than to an object-based facilitation per se. Notwithstanding the consistency of these findings with those of Experiment 1, we undertook yet a third experiment with these same displays to obtain unequivocal empirical evidence for the equivalence of occluded and nonoccluded single displays. Moreover, this experiment was also designed to probe the participants' representation of the occluded object in a more explicit fashion rather than having to infer it indirectly from the equivalence of the RTs for the single and occluded conditions.

To test the participants' representation of the occluded displays, we adapted the matching task used by Gerbino and Salmaso (1987) to examine the completion of amodal, occluded displays into an explicit categorization task. [1] Gerbino and Salmaso compared the speed with which participants responded *same* to a display containing pairs of items. One half of each pair consisted of a fully completed shape. The other half consisted of either the identical completed shape, the identical form partially occluded, or the identical form that was explicitly truncated. Because participants made *same* judgments to the pair containing the occluded form but not the truncated form as rapidly as to the completed form, Gerbino and Salmaso concluded that participants were indeed completing the amodal form and representing it as a whole shape.

In this experiment, instead of having participants make *same—different* responses to pairs of stimuli, we instructed them to categorize the occluded display into the same category as either the single or the different displays. Thus, for one group of participants, single and occluded objects were both categorized as belonging to A and the two-object displays were categorized as belonging to B, whereas for a second group, the single objects alone were categorized as A and the occluded objects were classified together with the two-object displays as B. If participants represented the occluded objects more like single objects than like two objects, then the speed of correct categorization of occluded objects should be faster in the former type of categorization (i.e., when it was assigned into the same class as the fully completed display) than in the latter type of categorization. By comparing the speed of categorizing the occluded objects under these two different assignments, we could therefore obtain evidence for whether participants interpreted the occluded form as completed or not. In such a paradigm, we circumvented the problem of assigning *same* and *different* responses to the conditions and more directly evaluated what representations participants were using for occluded objects. This type of experiment follows the trend toward testing phenomenological hypotheses by objective experimental techniques (see also Kellman & Spelke, 1983 ; Pomerantz & Kubovy, 1981 ; Shipley & Kellman, 1992).

Method Participants.

Fifteen men and 29 women (aged 18—42 years, *M* = 25.2) were recruited via the bulletin boards at Carnegie Mellon University. All consented to participate and received payment for their participation. All had normal or corrected visual acuity by self-report, and all but 4 were right-handed. They were consecutively assigned to Group 1 or Group 2.

Apparatus and materials.

The apparatus, stimuli, and material were identical to those used in Experiments 1a and 1b.

Design and procedure.

Participants were instructed that they were to perform a categorization task and were shown a sample set of displays in which the category assignment was demonstrated. Participants in Group 1 were instructed to classify single and occluded objects into one category and the two-object stimuli into a second category, whereas participants in Group 2 were told to classify single objects into one category and occluded and two-object stimuli into a second category. The categories were assigned labels A and B, and the label for the category was counterbalanced in both groups. The categorization decision was made by pressing the M and Z response keys, and these, too, were counterbalanced across participants. Accuracy and RT of categorizations were recorded. As in the previous experiments, participants completed three blocks of trials, each containing 96 trials for a total of 288 trials. Participants received 24 practice trials before the experimental trials.

Results and Discussion

The crucial question concerned the difference between speed of categorization of the occluded object when it was assigned along with the single object display compared with when it was assigned along with the two-object displays. To examine this, we performed an ANOVA with four within-subjects variables, including type of categorization group (occluded along with single or occluded along with two-object) and condition (single, occluded, and two). The other two variables were the number of bumps (same or different) to ensure that the number of bumps, although irrelevant to the categorization decision, did not interfere with the categorization in any way. We also evaluated the participants' performance across the three blocks of the experiment to determine whether, as participants became more familiar with the displays, there would be a change in their categorization. The most interesting and critical result was a difference in the speed of categorizing the occluded object for the two different participant groups. This difference diminished somewhat over the three blocks of the experiment, F (4, 176) = 2.2, p = .06, being most pronounced initially but still holding, albeit to a lesser degree, in the final block. Of most relevance was that participants in Group 1 (occluded assigned with single) categorized the single and occluded displays equally quickly (671.4 and 679.8 ms, respectively), whereas those in Group 2 (occluded assigned with two objects) took 44 ms longer to categorize the occluded than the single display (804.1 and 760.7 ms, respectively). This difference was largest in the first block, but the difference between single and occluded for the two groups still remained significant even in the third block. Interestingly, across all blocks, but especially in Block 1, the categorization times of the occluded object for participants in Group 2 were slower even than the two-object display (804.1 vs. 765.4 ms), suggesting that there might have been some additional interference or incongruence for the occluded display: Presumably, the participants perceived the occluded display as a single object, but, because it was designated as belonging to the same category as the two-object displays, this mismatch gave rise to the disproportionately long decision times for the occluded displays for these participants.

In addition to the critical Condition ? Group interaction, along with the influence of block, there were several other significant effects. Group 1 participants were 101 ms faster overall than Group 2, *F* (1, 44) = 4.53, p < .05, and all participants showed faster RTs in later than earlier blocks, *F* (2, 88) = 19.2, *p* = .0001, although participants in Group 2 showed even greater speedup from Block 1 to Block 2 and from Block 2 to Block 3 than did the participants in Group 1, *F* (2, 88) = 5.6, p < .01.

The findings from this experiment provide clear and unequivocal evidence for the view that occluded objects are completed and interpreted as single objects; participants were able to categorize the occluded display as rapidly as the completed single display when they were assigned to the same label. When the assigned category conflicted with their perceptual interpretation, however, RTs were significantly lengthened. These data suggest that occluded displays are well suited for evaluating the extent to which features are preferentially enhanced by an object-based mechanism that selects features from a single object and that this paradigm is a robust method for studying the representation of occluded items. That the participants perceived the occluded displays as phenomenally complete is consistent with the previous findings of Gerbino and Salmaso (1987 ; see also Sekuler and Palmer (1992) that an amodally completed figure, as in the case of our occluded displays, is functionally equivalent to a completed figure.

Experiment 2a

The explicit categorization task in Experiment 1c confirmed the functional equivalence of the occluded display and the single object display, and both the local feature (Experiment 1a) and object-level (Experiment 1b) versions demonstrated the predicted superiority for one over two objects even when the single object was occluded. We interpret these findings as evidence for a difference in selective attention to a single object relative to two objects and have argued that this selective mechanism applies equally well to occluded objects. There is, however, an alternative, perhaps simpler, interpretation of these data. A consideration of the displays in Figure 1 shows that whenever the two sets of bumps appeared on a single object (occluded or not as in Figures 1 a and 1d and Figures 1 b and 1e vs. Figures 1 c and 1f), the two sets of bumps fell along a straight line. By contrast, the two sets of bumps making up the two-object condition were always at right angles to each other. A possible explanation for the superior performance in the single- and occluded-object condition over the two-object condition therefore might be unrelated to the one- versus two-object distinction but might simply arise from the fact that it is easier to scan along a straight line than to process information along a 90° angle. Alternatively, because the features of one object always fall on the same rectangle and therefore appear at the end of two parallel lines, whether occluded or not, the superior performance for a single over two objects might have arisen from the salience afforded by the parallelism or the powerful perceptual cue of collinearity. Even if this were the case, it is still of interest that the two noncontiguous bars of the occluded object are afforded the status of a single object, but it is important to know more precisely whether it is these perceptual cues that are driving the effect rather than the object-based attentional facilitation. One question addressed in this next experiment, then, was whether the object-based effects are simply a function of the perceptual cues of continuity and collinearity or whether object-based attention for occluded and nonoccluded objects is also observable for similar objects that do not have these properties.

Another potential alternative explanation for the findings in Experiment 1 revolves around the contingencies that existed in the design of the experiment. Because there were an equal number of single, occluded, and two displays, there were more instances of the two sets of bumps appearing at the opposite ends of the corner of the display (as in the single and occluded conditions) than at adjacent corners (as in the two condition). Thus, if, for example, one set of bumps appeared in the upper left corner (as in Figure 1 a), the probability of the other set of bumps appearing at the opposite end of the same rectangle (occluded or not) was twice that of it appearing in an adjacent corner. Thus, the difference in making judgments in the single and occluded condition over the two condition might have arisen not because of the superiority of a single over two objects but simply because there was a higher probability of the second set of bumps appearing in the opposite corner. The second question addressed here, then, concerned whether these previous findings are simply the result of the particular contingencies embedded in the experimental design.

To verify that the critical effect observed in Experiment 1 was truly a consequence of a single- versus two-object distinction rather than an artifact of scanning speed, perceptual organization factors such as parallelism, or of unevenly weighted contingencies, we repeated the experiment with a different display in which the two sets of bumps of both the single- (occluder and occluded) and the two-object displays appeared at right angles to each other (see Figure 3). This display simply required the rearrangement of some of the lines from the X display to form two overlapping Vs. If the results found in Experiment 1 are indeed attributable to a difference in object-based processing, rather than to any of the possible artifacts, and this object-based procedure is robust across different types of displays, then we would expect to find the same pattern of single-object superiority (occluded or not) over two objects in this experiment as we did in Experiment 1.

Method Participants.

Ten men and 6 women (aged 18—22 years) were recruited from the introductory psychology subject pool at the University of Toronto. No one had participated in any previous experiment. All received course credit for their participation. All had normal or corrected visual acuity by self-report, and all were right-handed.

Apparatus and materials.

The apparatus was identical to that used in Experiment 1, but, whereas in that experiment, the display stimulus was made up of two rectangular bars that crossed in the midline making an X-shaped stimulus, in this experiment, the display consisted of two Vs, one rotated 180° with their apices overlapping (see Figure 3). The dimensions of the V displays were identical to the X stimuli.

As is evident from Figure 3 , the same three conditions were used as in Experiment 1: (a) single (or unoccluded) object, in which the bumps appeared at each end of a single V (Figures 3 a and 3d); (b) two objects, in which the bumps appeared at one end of the two different Vs (Figures 3 b and 3e); and (c) occluded object, in which the bumps appeared at each end of a single V that was occluded (Figures 3 c and 3f).

The orientation of the Vs was balanced such that on an equal number of trials, the Vs were superimposed on each other at the upper and lower edge of the display or at the right and left of the display and the bumps could appear equally at the top, bottom, left, or right. In all the conditions, the two sets of bumps were on adjacent corners and never fell along the diagonal. Because of this arrangement, there were no longer unequal contingencies on the locations of the bumps that could be used strategically by the participant while scanning the image. The rest of the experiment followed the same design and procedure used in Experiment 1a, and participants made local decisions on the number of bumps at the two corners of the display. RTs and accuracy were measured, and participants received a block of 24 practice trials at the beginning.

Results and Discussion

The mean of the median RTs across the participants as a function of judgment and condition is shown in Figure 4 , and the associated mean error rates are displayed in parentheses. A two-way ANOVA with judgment (same and *different*) and condition (single, occluded, and two) as within-subjects variables was conducted on error rates and median RTs. Errors constituted 1.8% of the total trials and were not affected by judgment, *F* (1, 15) = 0.33, p > .5, or by condition, *F* (1, 15) = 0.88, p > .1. In RTs, participants responded 24 ms faster on *same* than *different* trials, F (1, 15) = 9.7, *p* < .01, and there was a highly significant difference as a function of condition, *F* (2, 30) = 70.5, p < .0001, but no interaction between judgment and condition, *F* (2, 30) = 0.83, p < .1. Post hoc Tukey tests with a probability level of .05 revealed that responses to single and occluded displays were not significantly different from each other but that responses in each of these conditions were significantly faster than responses to the two-object trials.

The results of this experiment replicate those of Experiment 1a and demonstrate the generality of the object-based selection process. Irrespective of whether the two objects crossed each other to form an X or whether they were aligned as two overlapping Vs, participants were faster at making decisions about a single (occluded or not) object relative to two different objects. That the findings remained unchanged across display types suggests that the superiority for processing a single object cannot be attributed solely to the collinearity of the lines or to the specific design used in the first experiment. [2] Instead, the findings suggest that the ability to attend selectively to features of an object, occluded or not, is a robust and general ability that applies across different displays.

Experiment 2b

The results from Experiment 2a suggest that object-based facilitation applies generally across a range of stimulus displays and confirms the findings obtained in Experiment 1. As in Experiment 1, before concluding definitively that participants interpreted the occluded display as a single, completed form, a more explicit probe of the way in which the occluded form was represented is necessary. The same phenomenological categorization task (used in Experiment 1c) was therefore repeated using the V displays. As was the case previously, the occluded display was categorized either along with the single or the two-object displays with the prediction that it should be faster in the former than in the latter case if participants were representing it as a completed figure.

Method Participants.

Thirteen men and 7 women (aged 18—25 years, *M* = 20.7) were recruited either via the bulletin boards or from the undergraduate subject pool at Carnegie Mellon University. All consented to participate and received either payment or course credit for their participation. All had normal or corrected visual acuity by self-report, and all but 2 were right-handed. Participants were consecutively assigned to Group 1 or Group 2.

Apparatus and materials.

The apparatus, stimuli, and material were identical to those used in Experiment 2a.

Design and procedure.

The method was identical to that used in Experiment 1c. The two groups of participants classified the occluded objects along with the single objects or with the two objects, and the time to make the categorization decisions was recorded. The labels A and B that were used in making the categorization decision were counterbalanced within each of the two groups. Participants completed three blocks of 96 trials and received 24 practice trials before the experimental trials.

Results and Discussion

The major comparison of interest was the difference in categorizing the occluded displays when they were assigned together with the single- versus the two-object displays. As in Experiment 1c, we performed an ANOVA with four within-subjects variables, including group assignment (occluded along with single or occluded along with two object), condition (single, occluded, and two), number of bumps (*same* or *different*), and blocks (one, two, and three) of the experiment. The important finding was that participants in Group 1 (occluded assigned with single) categorized the single and occluded displays quickly (631.8 and 591.1 ms, respectively) with even a slight advantage for the occluded object, whereas those in Group 2 (occluded assigned with two) took far longer to categorize the occluded than the single display (615.3 and 649.5 ms, respectively), *F* (2, 36) = 5.13, p = .01. The difference in speed of categorization of the occluded objects between the two groups was 58.4 ms. Although participants increased their speed of categorizations over the three blocks of the experiment, *F* (3, 54) = 8.89, p < .0001, this did not affect the occluded display differentially (*F* < 1).

The major finding from this experiment was a difference in the speed of categorization of the occluded displays depending on whether they were assigned along with the single- or the two-object displays, with a significant advantage in the former over the latter case. This result suggests that the representation derived for occluded V displays was more consistent with the modal, completed object than with the two-object displays. This finding replicates the result from Experiment 1c and shows that in both experiments, participants phenomenologically experienced the occluded displays as being complete. There were some minor differences in the results from these two experiments: Whereas in the case of the Xs (Experiment 1c), the categorization effect was more dramatic in the first than in the third block, that was not the case with the Vs (Experiment 2b). Also, in the case of the Xs, the decision time for the occluded displays was significantly longer than the two-object displays when they were categorized together, suggesting that the mismatch between the perception and the categorization led to even slower responses. This mismatch effect did not manifest with the Vs. Exactly why these minor differences arose is not clear. What is most pertinent for our purposes, however, is that in both the X and the V case, the occluded object was treated more like a single object than like two separate objects when the participants' representation was probed directly and explicitly.

Experiment 3

The results of Experiments 1 and 2 are consistent in showing the superiority of processing a single object compared with two objects and reflect the generality of the object-based selection process. However, because in each of these two experiments we exclusively used one type of display (X or V), it was possible that the observed effects occurred through stimulus-specific expectations and that the object-selection process was really not as general as might be thought and instead was tailored to the specific display. If the object attention process is indeed general and flexible, as we claimed earlier, then we might expect to see the advantage for the single object, relative to the two-object condition, when both the X and V displays are mixed in the same block of trials in a within-subjects design. In this experiment we used the same basic design as in Experiments 1a and 2a, but every participant saw both X and V displays, randomized and presented in a mixed block, and performance on the two types of displays were compared within subjects. To our knowledge, this was the first object-based attention experiment in which two different forms of displays were used from trial to trial, and, as such, it provided a test of the flexibility of object-based attentional selection.

Method Participants.

Nine men and 16 women (aged 18—23 years) were recruited from the undergraduate subject pool at Carnegie Mellon University. No one had participated in any of the previous experiments. All received course credit for their participation. All had normal or corrected visual acuity by self-report, and all except 2 were right-handed.

Apparatus and materials.

The apparatus and materials were identical to those used in the previous experiments, except that the computer monitor was 14 in. (35.56 cm) rather than 13 in. (33.02 cm) in size. Both the X and V displays appeared with equal probability, and the conditions and judgment variables were the same as those used in the previous experiments.

Procedure.

Participants completed five blocks of 96 trials, making a total of 480 trials. The number of trials was increased relative to the other experiments to account for the additional within-subjects variable of display type. Each block contained an equal number of X and V displays, with display type crossed orthogonally with the variables of condition and judgment. The trials were randomized in a block, and participants saw 24 practice trials, 12 X and 12 V displays, before starting the experiment. Participants again made local decisions about the number of bumps on the two ends as in Experiments 1a and 2a. RTs and accuracy were measured, and the median RTs and mean error rates per condition were calculated per participant. The type of display, judgment, block (one through five), and condition all served as within-subjects variables in the three-way ANOVA, with median RTs and error rates as the dependent measures.

Results and Discussion

Error rates were low, constituting fewer than 3% of the total trials, with more errors in the occluded than in the single or two condition, *F* (2, 30) = 3.8, p < .05, and more errors for *same* than for different judgments, *F* (1, 15) = 10.6, p < .001. There were, however, equivalent errors for X and V display types (*F* < 1). The three-way ANOVA conducted on the RT data revealed that the crucial variable of display type (X or V) did not interact significantly with condition (single, occluded, or two), F (2, 30) = 0.66, p > .5, although participants made decisions 8.5 ms faster overall on the X than the V displays, *F* (1, 15) = 6.5, p < .05. The difference between the X and V displays was also marginally affected by judgment (*same* or different), *F* (2, 30) = 4.4, p = .05, and this interaction can be seen in Figure 5 . Post hoc Tukey testing with a probability level of .05 showed that the interaction arose largely because the time to make *different,* but not *same* judgments, was slower for the Vs than for the Xs (cf. Figure 1 f and Figure 3 f). *Same* and different judgments also manifested differently for the three conditions, *F* (2, 30) = 16.1, p < .0001. There were, however, joint effects of condition, judgment, and display type, *F* (2, 30) = 5.5, p < .01, revealing that the discrepancy between *same* and *different* judgments was exaggerated for the V displays relative to the X displays, particularly on the *different* occluded condition.

The most important result from this study was that there was no significant interaction between display type and condition, suggesting that the X and V displays were processed equivalently and showed the same basic single-object advantage. The results of this experiment confirm the finding that participants were able to attend selectively to features of single and occluded objects better than to features of two different objects even when two different displays, the X and V, were presented randomly intermixed in the same block of trials to the same participant. These data suggest that the process by which this featural enhancement takes place is not specific to a particular configuration and that it applies to occluded objects irrespective of display type.

Of interest in this mixed experiment was that the expected pattern of findings (single = occluded and both better than two objects) was stronger for the *same* than for *different* judgments. Exactly why this was so when there was no significant interaction between judgment and condition in Experiments 1 or 2 (when the displays were run individually) was not clear. Although there were no immediately obvious explanations for the weakened object-based effects for the *different* trials, there may be some potential perceptual or response-related explanations. One possibility suggested to us is that the three-bump end appeared to be slightly wider than the two-bump end and that this illusory perceptual effect might have weakened the object advantage for the different occluded displays. [3] Another possibility is that some aspects of the response requirements bring about this oddity. For example, it is known that under somewhat reduced certainty about a decision, participants typically engage in a verification process, particularly when the response is *different*, and that this verification process typically increases RTs (Nickerson, 1965). This, however, does not explain specifically why the RTs were so exaggerated for the *different* occluded V trials, but not X trials. Yet, another plausible, albeit not watertight, possibility is that there was some Stroop-like interference in that participants were responding different when the sets of bumps were on the same single or occluded object. This incongruence might lead to the lengthened RTs in these two conditions. None of these explanations fully accounts for the specific pattern of interaction, but all may play some role. What is most compelling from this experiment, for our purposes, is that the overall pattern of data shows the object-based selection benefit even in a mixed-presentation format.

Experiment 4

The results thus far favor the view that participants represent the occluded object as a single, completed object even when there are X and V occluded objects randomly intermixed and that the occluded and single object benefit equally from the object-based selection. In the final empirical experiment, we tested the boundaries of this object-based facilitation for the occluded object. To evaluate the robustness of the occluded object as a single object, we gradually displaced the noncontinuous bars of the occluded object and violated collinearity as the edges to be interpolated were misaligned. The question is, At what point do the features of the occluded object no longer show the single-object advantage relative to the two-object condition. Answering this will shed further light on the nature of the representation of the occluded object.

Previous attempts to examine the nature of the internal representation for occluded objects and their boundary limitations have been made by Spelke and her colleagues in their work with infants. In these studies, Spelke and her associates (Kellman & Spelke, 1983 ; Schmidt, 1985 ; Spelke, 1990) habituated infants to a display containing a center-embedded horizontal bar (much like our X displays) and two protruding occluding bars. They then assessed the children's subsequent looking patterns to displays that did not contain the center occlusion. When 2 1 2 -year-old children were initially shown center-occluded nonsense forms in which the visible surfaces of the occluded object were homogeneously colored and the edges were collinear at the point of occlusion (much like the X displays in this article), the children looked longer at the fragmented forms (i.e., judged as novel) that were not joined than at the single, continuous object. By contrast, when the visible surfaces of the center-occluded objects differed in color or were nonplanar so that extrapolation of the borders produced two distinct forms and were not collinear, the infants looked longer at the display containing a single form, suggesting that they no longer judged these displays as unitary, complete objects. Using a slightly different paradigm, Spelke and her colleagues showed that adults performed in the same manner. These findings suggest some limitations on the coherence of occluded objects; violations of principles such as collinearity or common color led participants to regard the occluded figure as two separate objects. Kellman and Shipley (1992) formalized this result and found that only under relatable conditions were the edges of the amodal object interpolated. More specifically, if the two edges can be connected by a smooth, monotonic curve whose end points match the two edge tangents, the edges are relatable and the occluded object will be perceived as complete.

If the amodal object is completed only under relatable conditions, then we would predict that when the two bars of the occluded object are displaced from the smoothly interpolated edges, the occluded object will no longer be represented as a complete entity and will no longer benefit from object-based facilitation. To evaluate this, we included three conditions in addition to the standard X displays, and, in each of the three, the collinear edges were increasingly misaligned. Participants performed *same—different* judgments on the number of bumps as in previous experiments, and we determined at what point the features of the occluded object were no longer facilitated relative to the two-object condition.

The first column in Figure 6 (Figure 6 a) reflects the standard X display for the single, occluded, and two conditions, used in the preceding experiments, in which the two arms of the occluded object are perfectly aligned. As has been shown in this standard presentation, the time to judge the features of the occluded object was not different from that of a single, coherent object across display manipulations. Figures 6 b—6d show the gradual increase in the displacement of the two bars as the misalignment increases from small to intermediate to large. If the object-based enhancement of the occluded display is restricted to conditions of relatability, for example, then with increasing displacement the RTs in the occluded display should closely approximate the two-object condition, especially for the intermediate (Figure 6 c) and large (Figure 6 d) displacements. If the segmentation process has some tolerance for displacement, as has been suggested previously (Kellman & Shipley, 1992 ; Shipley & Kellman, 1992), then the RTs for the small displacement (Figure 6 b) condition might not yet mirror the two-object condition function and might take up a middle position between the single- and two-object conditions.

Method Participants.

Nine men and 27 women (aged 19—28 years) were recruited to participate in this experiment. Twelve of them were drawn from the undergraduate subject pool in the Department of Psychology at Carnegie Mellon University and received course credit for their participation. The remaining participants were recruited from the bulletin boards and were paid for their involvement. All students had normal or corrected visual acuity by self-report, and all but 3 were right-handed. No one had participated in any of the previous experiments, and all agreed to participate in this one.

Apparatus and materials.

The apparatus was identical to that used in Experiment 3. There were four types of displays, as shown in Figure 6 : (a) standard (as used in Experiments 1a—1c and Experiment 3), (b) small offset, (c) intermediate offset, and (d) large offset. Each offset was half the width of the rectangle (2.98°), and so the offset was 0°, 2.98°, 5.96°, and 8.94°, respectively, from (a) through (d). The single, modal object was oriented left to right in half the trials and in the converse direction in the remaining trials. There were an equal number of same and different trials and an equal number of trials from each displacement type and from each condition. Participants completed six blocks of 96 trials in one session, for a total of 576 trials.

Procedure.

As in the previous experiments, a single display appeared on the screen for an unlimited duration, and the participants made *same—different* local feature judgments on the number of bumps as quickly and accurately as possible. RTs and accuracy were recorded.

Results and Discussion

The overall error rate was low, with a mean of .68% errors, and possibly even lower than in some of the previous experiments. A three-way ANOVA with displacement distance (standard, small, intermediate, and large displacements), condition (single, occluded, and two), and judgment (*same* or *different*) was conducted first with errors and then with median RTs as the dependent measure. The error analysis revealed a marginal three-way interaction among these variables, *F* (6, 210) = 1.1, p = .06: More errors were produced on the *same* two large displacement display than on any other condition. The three-way interaction among displacement distance, condition, and judgment and with RTs as the dependent measure was not significant (*F* < 1). There was, however, a joint effect of condition and size, *F* (6, 210) = 2.7, p < .05. Because this effect was not influenced by judgment, we show the Condition ? Size interaction in Figure 7 , collapsed across judgment. The error data are included in parentheses.

As is evident from this figure and from the post hoc tests, the advantage for processing features from the single over the two trials still held, irrespective of displacement distance. The more relevant finding concerned the status of the occluded object. In the standard condition, the occluded trials (758.8 ms) were equivalent to the single-object (761.1 ms) and were significantly different from the two-object (789.1-ms) trials. This replicates the result of Experiment 1a, and, as was the case in Experiment 3, it reveals that the object-based selective effect still held even when other configurations were shown to the participants in the same block of trials. In the small displacement trials, the occluded (778.4-ms) condition fell in between the other two conditions: The occluded display was not significantly different from the two-object (793-ms) display, but it was also not significantly different from the single-object (758.2-ms) display. For both the intermediate and large displacement display, the occluded condition was now equivalent to the two condition and was different from the single-object condition (although this difference fell short of the critical significant difference of 26.5 ms by just 2 ms in the large displacement trials).

In addition to the critical two-way interaction between condition and size described earlier, the joint effects of condition and judgment affected RTs significantly, *F* (2, 70) = 11.08, *p* < .0001. For same judgments, there was no difference between the single and occluded conditions, but both differed from the two condition, and, for *different* judgments, there was no difference between the occluded and two condition, but both differed from the single condition. This partly paralleled the judgment data from Experiment 3 and, as discussed previously, the exact explanation for the weakened effects in the *different* trials was unclear. The analysis also revealed the predicted significant main effect of condition, F (2, 70) = 36, *p* < .0001, with faster RTs for single and occluded than for the two-object condition and the predicted effects of judgment, *F* (1, 35) = 23.6, p < .0001, with faster *same* than *different* responses. Furthermore, RTs were significantly affected by the displacement distance, *F* (3, 105) = 22.9, p < .0001, with an incremental increase in RTs between standard and small, small and intermediate, and intermediate and large of 7, 16, and 15 ms, respectively.

In the first instance, the results from this experiment replicate the previous findings in showing that, in the standard condition, the occluded object was treated equivalently to the single-object display. Not surprisingly, the RTs even for the standard display were slightly longer in this experiment than in others, presumably because of the added difficulty of the other trials containing the displaced bars in this experiment. This experiment went further than the standard finding and demonstrated the limits of the object-based attention process. As the alignment of the protruding bars of the occluded object were displaced, so the RTs to the bars of the occluded object began to approximate the two-, rather than the single-, object condition. In the standard trials, the occluded object was not distinguishable from the single object in RTs. In the small displacement display, the status of the occluded object was ambiguous: It was not different from two objects, but it also was not completely separable from the single-object condition. In the intermediate and large displacement displays, responses to the occluded object were clearly equivalent to the two-object condition.

These results show that, although the processing of features of an occluded object was robust, this was so only under conditions in which the good continuation or collinearity of the parallel lines was maintained (see Day & Halford, 1994 , for a discussion of sensitivity to displacement on similar displays). When these organizing principles were violated, the object-based process operated differently, dividing the occluded object into two separate objects (essentially indicating that there were now three rather than two objects in the display). Of note, however, is the fact that there was some residual tolerance in the segmentation process: In the small displacement trials, the occluded condition was equivalent to the two-object condition and also was not differentiable from the single-object condition. This finding is consistent with that of Kellman and Shipley (1991 , (1992) , who demonstrated that the process by which the occluded (or illusory) edges are interpolated has some tolerance, albeit small, around collinearity. Whereas Kellman and Shipley (1991) found that this tolerance was around 15 min of arc misalignment, it was somewhat larger in our experiment and on the order of roughly 3° of visual angle. One possibility is that the increased tolerance here arose from the fact that participants' performance might have been contaminated by the presence of the standard display; because one quarter of the trials were of the standard form and contained perfect alignment, participants might then have been induced to be more accepting on the small displacement trials. The major point, then, is that facilitation in making decisions about features of an occluded object changes as misalignment increases. That we observed results similar to those of Kellman and his colleagues (also to Spelke and her colleagues) suggests that this experiment tapped into a similar process of interpolation and completion and that the object-based selection and facilitation operated over a similar internal representation of an occluded object.

A Computational Account

Taken together, the empirical results obtained from the various experiments have demonstrated the advantage afforded features of a single-object relative to two objects and the equivalent advantage afforded features of an occluded and a single, completed object. The findings have also revealed both the generality and the specificity of the object-based mechanism: Whereas facilitation of features of occluded objects was observed across different types of displays, even when they were mixed in the same experiment, the preferential processing of elements of the amodal, occluded object was constrained by the collinearity or relatability of the edges of the occluded object (Kellman & Shipley, 1992). Having characterized some aspects of this object-based process by which facilitation and preferential processing is afforded to features of a single (even occluded) object, we now turn to the question of underlying mechanism.

Thus far, we have followed Duncan (1984) in interpreting the single-object advantage as providing evidence for an object-based attention process. This standard object-based attention hypothesis posits that features can be compared more quickly in single objects because attention is simultaneously directed to the features of that object, whereas these comparisons take longer across objects because attention must be directed sequentially to the two objects. This hypothesis, however, begs the question of which image features are considered as belonging to a single object in the first instance. This question becomes even more relevant in cluttered visual environments, when features of one object may be occluded or obscured by others, further complicating the definition of which features to assign to a single object. Our empirical results (on simple examples of such cluttered environments) show that features of an occluded object as a single object may still be facilitated. We suggest that an essential part of this object-selection process is a mechanism that is responsible for grouping the visual features into objects. This grouping process essentially determines which features belong together and hence may be compared rapidly and which features are not bound together and thus take longer to compare.

In our computational account of these empirical results, we have therefore focused on the grouping of image features into objects. We hypothesize that this grouping, or *segmentation,* component is a key element in object-based attention and is the driving force behind the facilitation: It is by virtue of the fact that the features are segmented together that they may be preferentially gated for further processing. This gating may come about because of competition between the segmented features, with the chosen or marked features ultimately enhanced and winning relative to unattended features. This hypothesis is consistent with a general view of selective attention, such as that recently proposed by Desimone and Duncan (1995) , in which there are competitive and cooperative processes between features (or locations), and that these processes give rise to the benefit for the features that belong together and cohere (see also Duncan, 1996 ; Humphreys, Romani, Olson, Riddoch, & Duncan, 1994). We have taken this view one step further by implementing it in a set of explicit mechanisms and exploring some of the operational details.

In this section, we present an existing model, MAGIC, that embodies the general principles outlined earlier and that can account for the range of human data presented earlier. Before we describe the performance of MAGIC and show that it reproduces the human data remarkably closely, we first describe its representations, architecture, and training (additional details can be found in Mozer et al., 1992 ; Zemel, Williams, & Mozer, 1995). We then describe a decision or read-out process that we have added to the original MAGIC, so that experiments analogous to those that we have conducted with the human participants in this article can be conducted with MAGIC. In this way, the indexes of performance (local feature judgments and object-level judgments) obtained from the model and from the human data are directly comparable. The goals of the simulation studies were twofold: We first want to account for the empirical data and provide proof of our claim that the object-based facilitation arises from the gating of features that come to cohere in a perceptual representation. These simulations verified the accuracy and plausibility of the account. Our second focus, then, because it is still possible that many other models also may be able to simulate these data, was to derive testable predictions from the model. We believe that these specific predictions, outlined in the General Discussion section, further expand our understanding of the mechanisms underlying object-based attention.

Representation and Architecture

MAGIC is an adaptive computational network initially trained using images containing multiple objects that are presegmented; each feature in the image is labeled as to which object it belongs. When tested on novel images that are not segmented, the network must determine which features belong together as part of a single object and which belong to different objects. The system accomplishes this segmentation on the basis of the statistical regularities it extracts from the set of training images; it learns to detect spatially local configurations of the image features that are labeled consistently across the training examples, and this becomes the basis on which subsequent segmentation decisions are made.

The input patterns to MAGIC are visual images containing a variety of geometric contours. The contours for these images are constructed from four primitive feature types–oriented line segments at 0°, 45°, 90°, and 135°–laid out on a 25 ? 25 grid. Feature units that represent each of the four primitive feature types occur at each location on the grid. A given feature unit at a location contains a label that describes the object to which it belongs. During learning, the images are presegmented, and MAGIC is initialized with a random set of feature labels. It is then trained to produce labelings for the features that are consistent with this segmentation. During testing, target labels for the features are not provided and MAGIC is required to produce the labels for each of the features in these novel, unsegmented images.

The representation that allows for the labeling of the features has been inspired by the recent findings of temporal correlations among neural signals, either through the relative timing of neuronal spikes or through the synchronization of oscillatory activities in the nervous system (Eckhorn et al., 1988 ; Gray, Koenig, Engel, & Singer, 1989 ; Singer & Gray, 1995 ; however, see Nelson, 1995 , for a critical review). In MAGIC, each processing unit or feature conveys not just an activation value–average firing frequency–but also a second independent value that represents the relative phase of firing. The dynamic binding of a set of features belonging to a single object, then, is accomplished by aligning the phases of the features. The use of this phase representation is a computational device that allowed us to capture a continuous variable (e.g., time) and thereby to incorporate the principle of temporal synchronicity in a static representation. Phase, then, serves as a proxy for the more neurally realistic property of dynamic spike-dependent correlations. Hummel and Biederman (1992) and Lumer and Huberman (1992) used a similar scheme in their simulations, but, in that work, the pattern of connectivity between the oscillators was either prespecified by simple predetermined grouping heuristics or the groupings that could be learned were direct correlations between features (i.e., there were no hidden units that helped higher order combinations of features to cohere; see also Goebel, 1993). This is not the case in MAGIC, in which the principles that instantiate the phase alignment are acquired adaptively over time.

In MAGIC, each feature unit, then, has a complex-valued activity with both an amplitude and a phase component. The phase represents the labeling of the feature, and the amplitude represents the confidence in that labeling. The amplitude ranges from 0 to 1, with 0 indicating a complete lack of confidence and 1 indicating absolute certainty. There is no explicit representation of whether a feature is present or absent in an image; rather, absent features are clamped off (their amplitudes are forced to remain at zero), and so they are unable to influence other units. The network architecture, as shown in Figure 8 , consists of two layers of features. The lower (input) layer contains the feature units, arranged in spatiotopic arrays with one array per feature type. The upper layer contains the hidden units that are driven by the input (rather than directly by the environment) and that learn the internal representations necessary for solving the segmentation task. The hidden units help align the phases of the feature units; their response properties are determined through training. There are interlayer connections, but no intralayer connections. Each hidden unit is reciprocally connected to the units in a local spatial region of all feature arrays. We refer to this region as a *patch,* and, in these simulations, a patch has dimensions 4 ? 4. For each patch, there is a corresponding fixed-size pool of hidden units. To achieve uniformity of responses across the image, the pools are arranged in a spatiotopic array in which neighboring pools respond to neighboring patches and the patch-to-pool weights are constrained to be the same at all locations in the array.

Learning in MAGIC

In response to a visual input, the feature units activate the hidden units, which in turn feed back to the feature units. Through a relaxation process, the system settles on an assignment of phases to the features. The learning procedure allows the hidden units to detect local configurations of the image features that have a consistent labeling relative to each other across the training examples. During training, a pair of objects is instantiated with random sizes and positions on the input array, and the target phase of each feature of one object is set at 0°, and the target phases of the other object's features are set at 180°. The initial amplitude of a feature unit is set to 0.1 if its corresponding image feature is present or clamped to 0.0 otherwise. The phases of the feature units are set to random values in the range 0° to 360°. Activity is allowed to flow from the feature units to the hidden units and back to the feature units. The new phase pattern is compared with the target phase pattern, and an error measure is computed. A simple single-step algorithm is used. This involves running the network for a fixed number of iterations and, for each iteration, using a generalization of backpropagation to complex-valued units to adjust the weights so that the feature phase pattern better matches the target phase pattern.

Several hundred trials are required for stable performance, although MAGIC rapidly picks up on the most salient aspects of the domain. For example, when trained on transparent rectangles, some of MAGIC's hidden units will learn the regularity that collinear feature segments generally belong to the same object. Note also that, because a single input feature is connected to several neighboring hidden units, the labeling assigned to a set of input features can be propagated to other input features during the relaxation process. Thus, even though an individual hidden unit can directly affect only the labeling of a small local image patch, it can indirectly affect the labelings of distal features. There are several important points: MAGIC comes to learn, through training, to develop a set of internal representations which will successfully solve the task. Furthermore, even after heuristics are derived through training, these heuristics are not applied to novel displays in a rigid fashion. Instead, MAGIC has derived a host of constraints that simultaneously and flexibly combine to determine the outcome of a particular trial.

A Decision Process

The goal of the computation in MAGIC is to group image features into objects. We hypothesize that a grouping process is an essential component of object-based attention because these grouped features are gated by a selection attention mechanism. In our previous work with MAGIC (Mozer et al., 1992), our concern was more with the nature of the phase alignment. Our concern, here, however, is to provide a framework within which to interpret the empirical findings. Thus, to conduct experiments analogous to those conducted with the human participants, we implemented a simple decision process that worked together with MAGIC and allowed us to gather statistics and performance measures. The decision component of the model is an abstraction of the type of process humans might use in making perceptual decisions, but we are not making any strong theoretical claims about this specific implementation of this process.

Once the features in the display are grouped by MAGIC, the decision process containing two components, one for selection and one for comparison, comes into operation. The first component gates or selects a subset of local image features for further processing. We implemented this component using a histogram of the input features, in which the range of possible phases is divided into bins and each feature is assigned to the bin on the basis of its phase. The features contained in a bin of phases are selected for further processing, and, ideally, the binned features all belong to the same object. A bin must contain a sufficient number of features to be considered, and a random selection is then made from among the bins with enough features.

Once a set of features of a common phase is selected, these features are passed on to a comparison module, which decides whether any of the target features are included in the set. This comparison module is simply a response read-out mechanism and is an approximation of the process humans might use in such a task. To model the experimental results when decisions are made at the feature level, we designated two features of each display as being target features and these features could be on the same (single or occluded) or on two different objects. The comparison module evaluates these target features; once both target features have been seen in the comparison module, the decision process outputs a response. This examination of the two target features by MAGIC is analogous to the local feature judgment task (*same—different* number of bumps), and the time to reach this decision is measured as a function of condition (single, occluded, and two). Although MAGIC is not performing the identical task to that done by the humans and is not counting the number of bumps as in the local feature task, the principles are identical in both the human and model tasks: Features falling in different parts of the image need to be compared, and we measure the speed of this comparison as a function of the condition of the stimulus display. Similarly, in the object-level decision task, both human participants and MAGIC have to decide whether the critical features come from the same object. Again, although the details of the task differed, the essential components were similar and allowed us to compare the performance of the humans with that of the model.

Current Simulations

In previous work Mozer et al. (1992) have explored both the power and the limitations of MAGIC and found that it was capable of learning to segment features in a wide range of images but that its generality was limited by factors such as the input representation. Here, we focus on the extent to which MAGIC demonstrates performance equivalent to that displayed by the humans when presented with displays that have been constructed to resemble those used in the human experiments. The current simulations used the architecture, representation, and learning procedure described earlier. In these simulations, we first trained the network on a variety of images that were roughly similar to the stimuli in the behavioral experiments. We then connected the network with the decision process so that we could obtain a measure of the model's RTs. These RTs were calculated on the basis of the number of iterations required by the system to settle on a response, in which we ascribe a fixed number of additional iterations to the following set of operations: selecting and gating the features in an above-threshold phase bin; passing these features on to the comparison module; and searching (in parallel) for targets among these features. Note that the results do not depend on the details of this computation because it can just be considered as adding a constant number of iterations per selection on top of the network relaxation procedure.

One other point about the decision process concerns its two variables. The first is the number of bins in the phase histogram. Because there were only two objects in the displays we studied here, we found that the system was not sensitive to this value. The simulations we present here used eight bins, carving the phase circle into 45° windows, but the results looked similar for 4, 6, 10, and 12 bins with corresponding degree windows. The second decision process variable is the threshold for selecting the features in one of these bins. This value was determined on the basis of the minimum number of features in the objects in the training set to ensure that any object could be selected. Neither of these parameters is crucial to the mechanism, nor are the results a function of selecting specific values for these parameters.

Simulations of Experiment 1

To model the empirical data in Experiment 1a, each input image in the training set for MAGIC contained a pair of overlapping, opaque X rectangles of random sizes and positions. An example of such a display is shown in Figure 9 . A correct segmentation of the X display requires that the label of the features of the single, occluding object be assigned to the same phase. In the case of the occluded object, the feature label must be propagated across the spatial positions occupied by the occluder so that all the features of the occluded object come to have the same phase. After training on 3,500 trials, the system learned to segment novel images successfully. Figure 9 shows an example of MAGIC settling on a correct segmentation of a display that directly matches the occluded X displays used in the previous experiments. As is evident, although the features have random phase assignment initially, by Iteration 25, MAGIC labels all those features that belong to the occluding object in gray and those to the second occluded object in black.

To simulate the RT data from Experiment 1a in which participants made decisions about local features of the objects, we ran the system together with the decision process and collected statistics across a large number of trials. The trials had random initial phases, and the input features corresponded either to the image in Figure 9 or to its complement with the other rectangle on top. The critical target features (on which the comparisons were to be made) appeared equally on the input features of the single object, the occluded object or on the two different objects, and there were 100 trials in each of these three conditions. Trials (fewer than 5%) in which the target features were not selected for comparison were deleted from the analysis. Figure 10 shows the mean and standard error of the number of iterations (as a proxy for RTs) required to detect both target features in the images as a function of condition.

A one-way ANOVA with condition (single, occluded, and two) as a between-subjects variable revealed a significant effect of condition on the number of iterations, $F(2, 297) = 7.6$, $p < .001$. Additional one-way ANOVAs showed no difference between the single and occluded trials, $F(1, 198) = 1.6$, $p > .1$, but there was a significant difference between the single and two trials, $F(1, 198) = 14.6$, $p < .0001$, and between the occluded and two trials, $F(1, 198) = 6.4$, $p < .05$. These findings replicate the single-object advantage and the equivalence of the occluded and the single-object condition observed in the human participants when making local feature decisions as in Experiment 1a.

In this same simulation, we replicated the results of Experiment 1b in which participants were asked whether the two target features were on the same object or on different objects by altering the level of read-out. Similar object-level judgments may be derived from the model simply on the basis of the labels assigned to the target features: If the labels of these two tagged features fall in the same histogram bin, then we say that the model considers them to belong to the same object. For the 100 trials of each of the three conditions, features belonging to the same object appeared in the same bin on 95 of the single trials, on 89 of the occluded trials, and on 3 of the two trials. These simulations demonstrate that MAGIC is able to make global decisions when we use bin assignment as the dependent measure and to correctly segment the displays. The important finding from this simulation is that, when decisions are made on the basis of objects per se, MAGIC, like human participants, is able to segment the images successfully.

The key aspect of its performance that allows MAGIC to segment these images correctly is the existence of local configurations that reliably depict relative labels. An example of one such local configuration is the T-junction formed when an edge of the occluding rectangle intersects with an edge of the occluded rectangle. In such a case, MAGIC succeeds in segmenting these displays by forcing features of the two ends of the occluded object to have opposite phases from features of the occluder and thereby assigns equal phases to all the features of the occluded object. As mentioned previously, however, MAGIC uses a combination of heuristics to reach a segmentation in each trial, and these combinations lead to accurate and robust parsing of the otherwise noisy and potentially ambiguous display.

We have not attempted to simulate all the empirical findings and instead have chosen only those that make the points that are critical for our claim that grouping processes form a central aspect of object-based attention. The first point is that MAGIC can segment an occluded object as well as a single object (the simulation reported earlier). The second point is that MAGIC, like human participants, is sensitive to the perceptual constraints in the display, and, when the evidence is not consistent with the presence of an occluded object, MAGIC does not assign the same labels to these features. We present this simulation next. In principle, however, we believe that MAGIC can account for the full range of empirical data in a fairly straightforward fashion. Because MAGIC learns on the basis of the statistics of its environment, it should be able to handle Vs as well as a mixture of Xs and Vs. Critical local features distinguish the X and the V displays (e.g., a + junction indicates the superposition of two Vs, just as a T-junction indicates the superposition of two bars). Other local features are not unambiguous about the nature of the display (e.g., an L-junction could either indicate the inside junction of an unoccluded V or the corner junction of a bar), yet the labeling of the lines does not depend on knowing the display types. Thus, because there are some unambiguous local features, MAGIC should have no difficulty handling the X and V displays simultaneously.

Simulations of Experiment 4

In the human data, we found that although the object-based selection process was robust and held across the different display types, it was also subject to limitations such as violations of collinearity. An important test of how well MAGIC can reproduce the human performance (and thereby an explanation for the empirical data) is to examine how it performs under similarly difficult conditions. For example, the fact that the two segments of an edge of the occluded object are collinear is a powerful clue that they belong to the same object, and violations of this collinearity are highly suggestive of two different objects. As found in Experiment 4, humans have limited tolerance for misalignments, and, beyond some small threshold, consider the two discontinuous bars of the occluded object to be nonrelatable (also Kellman & Shipley, 1991 , 1992).

In its original instantiation, MAGIC was limited to discovering regularities that occur over restricted spatial distances in the image and was limited to local configurations that were contained within a receptive field of a hidden unit; in the implementation of MAGIC described earlier, these receptive fields were 4 ? 4. This limitation resembles the situation faced by cells in the lower levels of the visual system. For example, cells in striate cortex are estimated to have a receptive field of less than 0.5° of visual angle. The primate visual system overcomes such limitations in several ways, including having long-range horizontal as well as feedback cortical connections. The effect of these connections is that the visual system processes images at multiple spatial scales simultaneously. Fine-scale detectors respond to local image patches, and coarser scale detectors respond to a more global structure, which provides the finer scale detectors with a reasonable starting point and a narrower range of possible values. These coarser scale detectors may then be able to learn the longer range regularities such as those contained in the displays used in Experiment 4.

To model the results of the occluded bar displacement, we extended MAGIC to consider coarser resolution features. We altered the input images to correspond to coarser resolution images, in which each feature corresponded to a larger edge segment (for a similar computational device based on hierarchical decomposition of objects at different spatial scales, see Mozer et al., 1992 , p. 662). At this coarser scale, the objects appeared smaller, although clearly this was a computational device rather than a claim about object size scaling. Training the system on smaller objects allows the network to discover coarse scale features, such as those that bind features of the occluded object across the occluder. When the edges of the occluder are misaligned, as in Experiment 4, the system should learn to assign a different phase to the relevant features of the two disparate bars of the occluded objects; it should label the two ends as separate objects. The consequence of this is that when the system finally settles, there should be three rather than two differently labeled objects in the display.

We trained the system to assign a common label to the features of the occluder when the two ends were aligned but to assign different labels when they were not. Figure 11 shows an example of MAGIC correctly segmenting a test image with the displaced bars, as the random initial phases eventually were divided into three pools (light gray, gray, and black) corresponding to the occluder and the two nonaligned ends of the occluded shape. This three-way segmentation is equivalent to the displacement displays in Experiment 4, in which occluded trials were no longer assigned the status of a single object, with the result that three separate objects were considered to be present in the image.

To collect the statistical data, we repeatedly ran the system with the decision process with different random initial phases assigned to the input features of the image shown in Figure 11 (the displaced condition) or its complement with the other rectangle on top until 300 correct trials had been completed. The ($n = 23$) trials in which the system did not settle to a proper three-way segmentation were not included in the statistical analysis. Figure 12 shows the mean and standard error of the number of iterations required to find both target features in these images as a function of *single,* occluded, and two conditions. As was the case with the normal participants, the time required to make decisions on the display was longer in this displacement condition than was the case in the simple standard condition (compare Figure 2 with Figure 7 for human data and Figure 10 with Figure 12 for MAGIC).

A one-way ANOVA showed a significant difference across the three conditions, F (2, 297) = 11.2, p < .0001. Pairwise comparisons showed that the single-object advantage held in the difference between the single and two condition, F (1, 198) = 33.1, p < .0001. Now, in contrast to the results in the simulation of the standard condition (see Figure 9), the occluded condition was no longer equivalent to the single condition, F (1, 198) = 10.1, $p < .005$, and was instead equivalent to the two condition, F (1, 198) = 1.39, $p > .1$. MAGIC thus behaved in a similar way to the participants in Experiment 4, showing that, when the parallel lines were misplaced, the bars of the occluded object were no longer assigned the same labels. Although there were no direct correspondences between the displays of small, intermediate, and large misalignment, as used in Experiment 4, and the displays used here, the critical result was that, when the edges of the bar were misaligned, both human participants and MAGIC no longer treated the two bars as belonging to the same amodal object.

Taken together, the results of the simulations produced data that were very similar to those obtained in the empirical studies with the human participants. MAGIC was able to segment displays, even those that were occluded, and correctly assigned the phase labels to the features of the noncontiguous bars. Furthermore, the number of iterations required for the segmentation was less for the single (occluded or not) displays relative to the two-object display in the standard condition. Additionally, MAGIC was sensitive to the perceptual properties of the display, and, when the ends of the occluded object were out of alignment, MAGIC no longer interpreted the features as deriving from the same object (i.e., did not interpolate across the intervening occluded space). Aside from replicating the human data in MAGIC, the computations embodied in the system provide a feasible explanation for how object-based attention might arise. The claim is that, through feature grouping (which is based on the internal representations developed through perceptual experience), a set of elements that belong together come to cohere and to be selectively gated. Preferential processing is afforded to these bound features, and further processing such as local judgments or object decisions that involve these grouped features is facilitated relative to nonselected or ungrouped features.

General Discussion

A fundamental problem facing the visual system is how to deal with the overwhelming amount of information that is present in a multiobject visual scene. A long-standing proposal has been that spatially contiguous regions are preferentially selected, thereby reducing the complexity of the display and facilitating the processing of information from a discrete physical region of the visual environment. Whereas this space- or location-based mechanism might suffice under some conditions, in many real-life situations, objects appear in front of one another and cannot be segregated by spatial region alone. An alternative selection process by which objects, rather than physical locations, may be selected has been proposed (for early proposals, see Duncan, 1984 ; Neisser, 1967), and considerable recent evidence supports the independent existence of such a process. Despite the robustness of the empirical data on this object-based attention, the conditions under which this mechanism operates have not been explored in much detail. The goal of this article was twofold: (a) to present a comprehensive set of empirical data showing the conditions under which this object-based process operates and illustrate some boundary conditions and (b) to suggest a way in which this object-based selection occurs through investigation of a computational model, MAGIC, that performs object segmentation via feature grouping.

In a series of experiments, we found that participants were able to make both local feature judgments and object-level judgments faster for single objects than for two different objects. These findings are consistent with the data on object-based selection (e.g., Baylis & Driver, 1993 ; Duncan, 1984 ; Egly, Driver, & Rafal, 1994 ; Kramer & Watson, 1995 ; Vecera & Farah, 1994). A particularly stringent situation in which such an object-based mechanism is required to operate is one in which features or contours that presumably belong together are fragmented so that elements of a single object project to the retina from nonproximal locations. This arises under conditions of occlusion and illusory contours, with the former being far more common in natural scenes. Yet, despite the fragmented input, humans experience little difficulty in determining the unity and boundaries of these amodal or incomplete objects (Kanisza, 1979). In these experiments, we extended the basic single-object advantage result to demonstrate that an object-based selection mechanism also operates on occluded objects, affording it and its features preferential enhancement in a manner equivalent to that which occurs for a unitary, modal object. The benefit that accrues to a single object, whether occluded or not, is robust and general and is found even under conditions in which different display configurations appear to be randomly intermixed.

This object-based selection procedure, however, also is particular in its operation and does not consider all input as constituting a single object. When the perceptual support for features forming a single object is weakened, as in the case in which occluded bars are displaced, the object-based advantage is not obtained (although with minimal displacement, the two occluded bars are also not yet treated as two objects and retain an intermediate status). As the two bars of the occluded object are increasingly displaced, participants no longer show the superiority in decision time for the occluded object, with the RTs now being equivalent to the two- rather than the single-object condition. The implication of this result is that, under conditions of intermediate and large displacements, participants function as though there were three objects present in the display (two independent occluded bars and one single occluder). Thus, this procedure admits into its domain only those displays whose features are consistent with a single object, and only those features are then preferentially processed.

That partly occluded objects benefit from object-based attention suggests that completion occurs relatively early on in processing and is not, as is sometimes claimed, the result of high-level inference or problem solving (see Gregory, 1970 ; Scholl & Leslie, 1998). This finding also converges with the growing agreement that segmentation and grouping operates early on in visual processing, perhaps preattentively and in parallel across the display (Enns & Rensink, 1996 ; Rensink & Enns, in press), and can even affect low-level processes such as motion and orientation perception. For example, Shimojo and Nakayama (1990) showed that, in bistable displays, when an occluding surface was interpreted as being present, it blocked the position of a moving target and hence affected the correspondence solving process for apparent motion (but see Sekuler & Sekuler, 1993 , for a discussion of the level of motion tested). In the case of orientation, Watanabe (1995) found that the McCollough effect, generally thought to occur early in the visual pathway (Humphrey, Goodale, & Gurnsey, 1991), is still elicited even with occluded, perceptually discontinuous edges. Finally, Valdes-Sosa, Bobes, Rodriguez, and Pinilla (1998) have identified early potentials (N1 and P1) starting roughly 100 ms after stimulus onset that are suppressed in two-object but not in single-object displays.

In this article the central behavioral results demonstrating the object-based advantage for nonoccluded and occluded objects were obtained using a novel paradigm for the study of object-based attention and, simultaneously, for the study of occlusion. Regarding occlusion, through a more direct probe of the participants' phenomenological experiences with the occluded displays, we obtained evidence that they represented the occluded object as though it were a completed shape. This confirms the suitability of these displays for the study of occlusion. Across a variety of experiments, this paradigm produced robust and replicable results of object-based facilitation, and both accuracy and RT data were obtained. The paradigm also avoids some of the difficulties typically associated with previous object-based experiments. In Duncan's (1984) original experiments, some of the judgments involved local properties of an object (e.g., line texture), whereas others involved more global properties (e.g., box size); also, the two objects–a box and a line–were much different and possibly could have been segregated by spatial frequency differences (Watt, 1988).

In each display of our paradigm, participants made the identical judgments on a pair of objects. The fact that the expected pattern of object advantage and cost was observed in this task in which there were no obvious cues about how the objects might be separated lends further support to the object-based hypothesis and is consistent with the findings from Baylis and Driver (1993) and Lavie and Driver (1996) , in which potential artifacts were well controlled. In addition, as is evident from Experiments 1a and 1b, the same paradigm can be used with different task instructions to probe the outcome at the local feature level or at the object response level. We obtained consistent results with both these versions of the task and replicated the findings of Vecera and Farah (1997 ; Vecera (1993) , who had their participants make object-level decisions in an object segmentation task.

A crucial aspect of our results that has emerged from this new paradigm was that object-based attention also operates on occluded objects. This finding suggests that the features of occluded objects are bound together before or as part of the operation of attentional selection. That perceptual completion occurs early in processing is supported by the recent studies of Moore, Yantis, and Vaughan (1998) , who showed that the object-based benefit also accrues in the case of subjective contours. This conclusion raises the second primary issue in this article, which concerns the underlying mechanisms of object-based attention. We have suggested here that an essential component of object-based attention is the segmentation of the image into objects and have proposed that a model such as MAGIC provides one feasible mechanism underlying object-based attention. We have shown that MAGIC, an adaptive grouping system, produced behavior remarkably similar to that observed in humans: MAGIC learned to assign features that belonged to a single object to a particular label and did so competently even in the case of the occluded objects. This result held both when local feature judgments were required as well as when the output of the more global object segmentation procedure was assessed. Furthermore, MAGIC performed in a manner that paralleled that of the human participants under conditions of collinearity violation. When the field size in MAGIC was increased or, equivalently, resolution was decreased to be more comparable to the dimensions of the human visual system, performance was again similar to the human participants: MAGIC no longer considered the occluded bars as belonging to the same object and assigned three different phases to the features present in the display.

The mechanism by which object segmentation and parsing is achieved in MAGIC is through the assignment of features of a single object to the same phase. Although this phase-based scheme is appealing because it provides a way in which to represent a dynamic, temporally synchronous process in a static continuous representation and because it has a neurally plausible correlate in the form of coupled oscillators (for a recent review, see Singer & Gray, 1995), we are not arguing that it is through this exact mechanism that humans perform this task. The key principle concerns the nature of the representations developed by MAGIC over the course of the training regime through experience with a set of randomly generated geometric images. It is these same representations that might well mediate human performance and phase is just one way of labeling these representations.

One useful representation reflected in MAGIC's hidden units assigns the two features of a T-junction to two different objects by setting them out of phase with each other (Mozer et al., 1992). This works well to segregate the occluded from the occluding object. A second valuable underlying representation assigns the same phase to two features that belong to the same object (e.g., collinearity). The representations or grouping rules discovered by MAGIC are consistent with Gestalt principles of perceptual organization. What is particularly important, however, is that during segmentation, multiple heuristics operate simultaneously, and their joint constraints determine the eventual response of the network. For example, both the collinear and T-junction representations are useful when assigning a common phase or label to the two ends of the occluded objects: The T-junctions set both of the occluded bars out of phase from the occluding object, and the collinearity (across the occluded boundary) is consistent with both these end bars belonging to a single object. Similarly, collinearity alone would not suffice in the intermediate displacement condition in Experiment 4 (e.g., see Figure 6 c), in which the two edges of two different objects lined up, but these "illusory" collinear lines should not be grouped together and are actually out of alignment. If collinearity alone were operating, these disparate bars would have been bound together. That these two bars are not grouped together or considered as part of the same object suggests that neither the human participants nor MAGIC use collinearity as the exclusive clue to the segmentation of the image. Thus, more than one form of representation likely comes into play during grouping, and the converging evidence from the T-junction and collinearity, inter alia, works to segregate the overlapping objects.

There are currently several explanations for how amodal or partially incomplete objects are bound together by humans. One view is that the objects are segregated by virtue of assigning them to two different depth planes: The top object is perceived as being nearer and the back or occluded object is seen as being farther away. This disparity is crucial for segmentation, and the presence of a T-junction provides strong evidence for depth (Nakayama & Shimojo, 1990 ; Nakayama, Shimojo, & Silverman, 1989 ; a similar point was made by Helmholtz, 1910/1962). Recent neurophysiological evidence suggests that veridically manipulating the depth planes of an object and its occluder has direct consequences for the underlying neuronal responses. Baylis (1998) , for example, found that the face-selective responsivity of cells in inferotemporal cortex was greatly reduced when an occluder fell in the same plane as the face (coplanar occlusion). The selectivity reappeared, however, when the occluder and the face occupied different depth planes with the occluder in front, suggesting that depth helps in the parsing and segmenting of the face when the image is fragmented through occlusion.

On the basis of these data, one might conclude that object-based attention for single objects is not necessarily related to feature grouping, as we have claimed, but is a direct consequence of segregation in depth by a spatial-based 3-D attentional mechanism. In our studies, both the single and the occluded condition involved dividing attention within the same depth plane, but the two-object condition involved dividing attention across different depth planes. In fact, Posner (see the footnote in Duncan, 1984) proposed this depth effect as the explanation for Duncan's original findings of an object-based advantage. There has been a spate of recent work that has carefully examined whether a spatial attention spotlight indeed operates in three dimensions. Although there is not total agreement on this issue, the central finding suggests that, under conditions similar to those used in our experiments, spatial selection operates on a representation that does not include depth information (for different types of displays and results, see Downing & Pinker, 1985 ; Gawryszewski, Riggio, Rizzolatti, & Umilta, 1987 ; Hoffman & Mueller, 1994). For example, when benefits of spatial cuing are examined for targets occupying two locations in the same depth plane or at locations in two different depth planes that are equidistant to those in the same depth plane, attentional cuing benefits were observed only for the former but not for the latter displays (Ghirardelli & Folk, 1996 ; see also Iavecchia & Folk, 1994 ; Zimba & Tellinghuisen, 1990). These findings suggest that a 3-D spatial mechanism does not suffice as an explanation for our findings. Importantly, in its current formulation, MAGIC also does not rely on the recovery of or assumptions about depth information. Instead, the claim is that by virtue of the representations developed through experience with a set of perceptual displays, MAGIC has captured some of the statistical regularities contained in the image and it then makes use of this knowledge to group features that belong together. Those features are then preferentially enhanced or gated, giving rise to the single-object advantage without any reliance on depth cues.

Our hypothesis, then, is that object segmentation, such as that instantiated in MAGIC, serves to chunk the display into discrete objects on the basis of whatever heuristics are adopted. The product of this segmentation or feature grouping is selected and preferentially enhanced (potentially through a competitive process such as suggested by Desimone, and, & Duncan, 1995 , and Duncan, 1996) for later analysis. This view proposes that object-based attention is a dynamic process in which elemental features are bound together and then enjoy an attentional advantage. Taken together, the findings from the human participants and from MAGIC converge with much of the recent data suggesting that object-based selection is a robust and reliable process and that this process may operate even under difficult perceptual conditions such as occlusion.

A final issue that has not been discussed much concerns the development of the representations that underlie feature grouping. A primary focus in our computational work was to suggest that, through experience with particular displays from which regularities and consistencies were extracted, grouping was driven in a bottom-up fashion. What constitutes an object for this mechanism, then, is not necessarily the presence of a top-down label, categorizing it into a known or familiar shape (although additional top-down knowledge can be advantageous and assist segmentation; Peterson, 1994 ; Vecera, 1993). Instead, for our purposes, an object is simply a set of features that has structure or regularity by virtue of being organized into the same configuration over multiple occurrences. Our computational model is thus not just proof for an object-based attention mechanism but instead allows us to derive strong predictions for human performance in other tasks. The first set of predictions concerns the adaptive nature of the grouping. If it is indeed the case that the representations that develop from perceptual experience can be used as the basis for object selection and enhancement, an obvious and testable prediction is that, through learning, participants may be taught to group together features that might normally be considered as belonging to different objects. For example, exposing participants to the displays in Experiment 4 in which the occluded bars were displaced but using color or common motion to indicate that the bars actually belonged to one object and thereby to ensure relatability may have led the participants to group together the occluded bars even when they were maximally displaced (Spelke, 1990). Another method of achieving this effect is to present evidence that the displaced occluded bars do in fact belong to the same object by reversing the occlusion relations in the display and having the misaligned bars be part of an unusually shaped object (see Figure 13 a for an example of such a display). Our prediction was that after some exposure to this new type of stimulus, we would find that participants treated the "displaced ends" of this object when occluded as legitimate components of a single object and that the RTs to make decisions about the bumps in this display would be equivalent to those for a single object. Preliminary data suggested that this was indeed the case (Zemel, Behrmann, Mozer, & Bavelier, 1998).

Another set of predictions concerns the generalization of the knowledge embodied in the representations that have been derived from experience on a novel set of images. For example, we predicted that segmentation based on learned local continuity and occlusion for the X displays would serve to bootstrap additional grouping principles in the display. Consider the displays in Figure 13 b and 13c, in which only the ends of the objects are present. In such a situation, if participants perform *same—different* judgments on the number of bumps, we would predict equivalent RTs for both displays (i.e., irrespective of which of the four corners are occupied by the bumps). Because there is no object to mediate the grouping, the end sections will be interpreted as two independent objects. However, if these same participants are then presented with the occluded bars from the Xs in Experiment 1, make local feature judgments on these, and are then retested on these ends-alone displays, we predict that the single-object advantage will emerge in Figure 13 b but not for Figure 13 c; the pairs of bumps at the opposite diagonal ends will benefit from this advantage after exposure to the overlapping X displays and will come to be treated as two ends of a single object. Because opposite ends of the bars will always be grouped together (via the primary grouping principles), a secondary grouping principle will be learned to link the opposite ends. In fact, a process similar to this might be mediating illusory conjunctions.

A final prediction is that if we follow the same method on a separate group of participants, exposing them to the ends-alone first, then to the occluded Vs from Experiment 2 and then test them again on the ends alone, we will see the single-object advantage emerge for the set of ends that are consistent with the single-object V displays (see Figure 13 c) rather than with the X displays (see Figure 13 b): The bumps on the horizontally adjacent end sections will produce faster RTs than the bumps on the diagonal ends. Our model predicts that these types of perceptual learning effects will emerge as specific grouping strategies are learned because of repeated exposure to a particular class of stimuli. Indeed, evidence for the power of perceptual learning and the functional plasticity of the perceptual mechanism, even during the early stages of visual processing, has been repeatedly demonstrated (e.g., Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994 ; Sagi & Tanne, 1994) and reflects the effect of specific changes that are dependent on the particular perceptual experience of the participant. Experiments examining the transfer and generalization of knowledge to novel displays have yielded results consistent with the predictions we have laid out here (Zemel et al., 1998).

References

Baylis, G. (1998). *Effects of partial image occlusion on the face-selective responses of cells in macaque temporal lobe.* (Manuscript submitted for publication)

Baylis, G. & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors.(*Perception & Psychophysics, 51,* 145—162.)

Baylis, G. C. (1994). Visual attention and objects: Two-object cost with equal convexity.(*Journal of Experimental Psychology: Human Perception and Performance, 20,* 208—212.)

Baylis, G. C. & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location.(*Journal of Experimental Psychology: Human Perception and Performance, 19,* 451—470.)

Behrmann, M., Vecera, S. & McGoldrick, J. (1998). *Selective attention to objects and their parts: Enhancement at multiple levels of a hierarchical representation.* (Manuscript submitted for publication)

Broadbent, D. (1982). Task combination and selective intake information.(*Acta Psychologica, 50,* 253—290.)

Bub, D. & Gum, T. (1991). *Psychlab experimental software.* (Montreal: McGill University)

Day, R. H. & Halford, A. P. (1994). On apparent misalignment of collinear edges and boundaries.(*Perception & Psychophysics, 56,* 517—524.)

Desimone, R. & Duncan, J. (1995). Neural mechanisms of selective visual attention.(*Annual Review of Neuroscience, 18,* 193—222.)

Downing, C. J. & Pinker, S. (1985). The spatial structure of visual attention.(In M. Posner & O. Martin (Eds.), Attention and performance XI (pp. 171—187). Hillsdale, NJ: Erlbaum.)

Driver, J. & Baylis, G. (1989). Movement and visual attention: The spotlight metaphor breaks down.(*Journal of Experimental Psychology: Human Perception and Performance, 17,* 561—570.)

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

Duncan, J. (1993). Similarity between concurrent visual discriminations: Dimensions and objects.(*Perception & Psychophysics, 54,* 425—430.)

Duncan, J. (1996). Co-operating brain systems in selective perception and action.(In T. Innui & J. L. McClelland (Eds.), Attention and performance XVI (pp. 549—578). Cambridge, MA: MIT Press.)

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboek, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex?(*Biological Cybernetics, 60,* 121—130.)

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.)

Egly, R., Rafal, R., Driver, J. & Starrveld, Y. (1994). Covert orienting in the split-brain reveals hemispheric specialization for object-based attention.(*Psychological Science, 5,* 380—383.)

Enns, J. T. & Rensink, R. A. (1996). An object completion process in early vision.(In A. G. Gale (Ed.), *Visual search III: Proceedings of the Third International Conference on Visual Search.* London: Taylor & Francis.)

Eriksen, B. A. & Eriksen, C. W. (1974). Effect of noise letters upon the identification of target letter in a nonsearch task.(*Perception & Psychophysics, 16,* 155—160.)

Eriksen, C. W. & Yeh, Y.-Y. (1985). Allocation of attention in the visual field.(*Journal of Experimental Psychology: Human Perception and Performance, 11,* 583—587.)

Gawryszewski, L. D. G., Riggio, L., Rizzolatti, G. & Umilta, C. (1987). Movements of attention in the three spatial dimensions and the meaning of "neutral" cues.(*Neuropsychologia, 25,* 19—29.)

Gerbino, W. & Salmaso, D. (1987). The effect of amodal completion on visual matching.(*Acta Psychologica, 65,* 25—46.)

Ghirardelli, T. G. & Folk, C. L. (1996). Spatial cuing in a stereoscopic display: Evidence for a "depth-blind" attentional spotlight.(*Psychonomic Bulletin and Review, 3,* 81—86.)

Goebel, R. (1993). Perceiving complex visual scenes: An oscillator neural network mode that integrates selective attention, perceptual organization and invariant recognition.(In C. L. Giles, S. J. Hanson, & J. D. Cowan (Eds.), Advances in neural information processing systems (Vol. 5, pp. 903—910). San Mateo, CA: Morgan Kaufmann.)

Gray, C. M., Koenig, P., Engel, A. K. & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties.(*Nature (London), 338,* 334—337.)

Gregory, R. L. (1970). *The intelligent eye.* (New York: McGraw-Hill)

Helmholtz, H. (1962). *Treatise on physiological optics.* (New York: Dover. (Original work published 1910)

Hoffman, J. E. & Mueller, S. (1994, November). *An in-depth look at visual attention.* (Paper presented at the 35th Annual Meeting of the PsychonomicSociety, St. Louis, MO.)

Hummel, J. & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition.(*Psychological Review, 99,* 3, 480—517.)

Humphrey, G. K., Goodale, M. A. & Gurnsey, R. (1991). Orientation discrimination in a visual form agnosic: Evidence from the McCullough effect.(*Psychological Science, 2,* 331—335.)

Humphreys, G. W., Romani, C., Olson, A., Riddoch, M. J. & Duncan, J. (1994). Nonspatial extinction following lesions of the parietal lobe in humans.(*Nature, 372,* 357—359.)

Iavecchia, H. P. & Folk, C. L. (1994). Shifting visual attention in stereographic displays: A time course analysis.(*Human Factors, 36,* 606—618.)

Kanade, T. (1981). Recovery of three-dimensional shape of an object from a single view.(*Artificial Intelligence, 17,* 409—460.)

Kanisza, G. (1979). *Organization in vision: Essays in Gestalt perception.* (New York: Praeger)

Kanisza, G. & Gerbino, W. (1982). Amodal completion: Seeing or thinking? (In J. Beck (Ed.), *Organization and representation in perception.* Hillsdale, NJ: Erlbaum.)

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M. & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill.(*Science, 265,* 679—681.)

Kellman, P. J. & Shipley, T. F. (1991). A theory of visual interpolation in object perception.(*Cognitive Psychology, 23,* 141—221.)

Kellman, P. & Shipley, T. F. (1992). Perceiving objects across gaps in space and time.(*Current Directions in Psychological Science, 1,* 193—199.)

Kellman, P. & Spelke, E. S. (1983). Perception of partly occluded objects in infancy.(*Cognitive Psychology, 15,* 483—524.)

Koriat, A. (1994). Object-based apparent motion.(*Perception and Psychophysics, 56,* 392—404.)

Kramer, A. F. & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing.(*Perception and Psychophysics, 50,* 267—284.)

Kramer, A. F. & Watson, S. E. (1995). Object-based visual selection and the principle of uniform connectedness.(In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), Converging operations in the study of visual attention (pp. 395—414). Washington, DC: American Psychological Association.)

Kramer, A. F., Weber, T. A. & Watson, S. E. (1997). Object-based attentional selection: Grouped-arrays or spatially-invariant representations.(*Journal of Experimental Psychology: General, 126,* 3—13.)

Lappin, J. S. (1967). Attention in the identification of stimuli in complex displays.(*Journal of Experimental Psychology, 75,* 321—328.)

Lavie, N. & Driver, J. (1996). On the spatial extent of attention in object-based visual selection.(*Perception & Psychophysics, 58,* 1238—1251.)

Lowe, D. G. & Binford, T. O. (1982). Segmentation and aggregation: An approach to figure-ground phenomena.(*Proceedings of the Defense Advanced Research Program ISU Workshop* (Palo Alto, CA), pp. 167—178.)

Lumer, E. & Huberman, B. A. (1992). Binding hierarchies: A basis for dynamic perceptual grouping.(*Neural Computation, 4,* 341—355.)

Marr, D. (1977). Analysis of occluding contour.(*Proceedings of the Royal Society of London, B197,* 441—475.)

Moore, C., Yantis, S. & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion.(*Psychological Science, 9,* 104—110.)

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.)

Nakayama, K. & Shimojo, S. (1990). Toward a neural understanding of visual surface representation.(In T. Sejnowski, E. R. Kandel, C. F. Stevens, & J. D. Watson (Eds.), *Cold Spring Harbor Symposium on Quantitative Biology, 55,* 911—924.)

Nakayama, K., Shimojo, S. & Ramachandran, V. S. (1990). Transparency: Relation to depth, subjective contours and neon color spreading.(*Perception, 19,* 497—513.)

Nakayama, K., Shimojo, S. & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping and the recognition of occluded objects.(*Perception, 18,* 55—68.)

Neisser, U. (1967). *Cognitive psychology.* (New York: Appleton-Century-Crofts)

Nelson, J. I. (1995). Binding in the visual system.(In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 157—159). Cambridge, MA: MIT Press.)

Nickerson, R. F. (1965). Response times for "same"-"different" judgements.(*Perceptual and Motor Skills, 20,* 15—18.)

Peterson, M. A. (1994). Object recognition processes can and do operate before figure-ground organization.(*Current Directions in Psychological Science, 3,* 105—111.)

Pomerantz, J. & Kubovy, M. (1981). Perceptual organization: An overview.(In M. Kubovy & J. R. Pomerantz (Eds.), Perceptual organization (pp. 423—456). Hillsdale, NJ: Erlbaum.)

Posner, M. I. (1980). Orienting of attention.(*Quarterly Journal of Experimental Psychology, 32,* 3—25.)

Posner, M. I., Snyder, C. R. R. & Davidson, B. J. (1980). Attention and the detection of signals.(*Journal of Experimental Psychology: General, 109,* 160—174.)

Prinzmetal, W. (1981). Principles of feature integration in visual perception.(*Perception & Psychophysics, 30,* 330—340.)

Ramachandran, V. S. (1992). Filling in gaps in perception: I.(*Current Directions in Psychological Science, 1,* 199—205.)

Rensink, R. A. & Enns, J. T. (in press). Early completion of occluded objects.(*Vision Research* .)

Rock, I. & Gutman, D. (1981). The effect of inattention on form perception.(*Journal of Experimental Psychology: Human Perception and Performance, 7,* 275—285.)

Sagi, D. & Tanne, D. (1994). Perceptual learning: learning to see.(*Current Opinion in Neurobiology, 4,* 195—199.)

Schmidt, H. (1985). *The development of Gestalt perception.* (Unpublished doctoral dissertation, University of Pennsylvania,Philadelphia.)

Scholl, B. & Leslie, A. M. (1998). Explaining the infant's object concept: Beyond the perception/cognition dichotomy.(In E. Lepore & Z. Pylyshyn (Eds.), Rutgers lectures on cognitive science. Oxford, England: Blackwell.)

Sekuler, A. B. & Palmer, S. E. (1992). Perception of partly occluded objects: A microgenetic analysis.(*Journal of Experimental Psychology: General, 121,* 1, 95—111.)

Sekuler, A. B., Palmer, S. E. & Flynn, C. (1994). Local and global processes in visual completion.(*Psychological Science, 5,* 260—267.)

Sekuler, A. B. & Sekuler, R. (1993). Representational development of direction in motion perception: A fragile process.(*Perception, 22,* 899—915.)

Shimojo, S. & Nakayama, K. (1990). Amodal representation of occluded surfaces: Role of invisible stimuli in apparent motion correspondence.(*Perception, 22,* 899—915.)

Shimojo, S., Silverman, G. H. & Nakayama, K. (1989). Occlusion and the solution of the aperture problem for motion.(*Vision Research, 29,* 619—626.)

Shipley, T. F. & Kellman, P. J. (1992). Perception of partially occluded objects and illusory figures: Evidence for an identity hypothesis.(*Journal of Experimental Psychology: Human Perception and Performance, 18,* 106—120.)

Singer, W. & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis.(*Annual Review of Neuroscience, 18,* 555—586.)

Spelke, E. S. (1990). Principles of object perception.(*Cognitive Science, 14,* 29—56.)

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V. & Pinilla, T. (1998). Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials.(*Journal of Cognitive Neuroscience, 10,* 137—151.)

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

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

Vecera, S. P. & Farah, M. J. (1997). Is image segmentation a bottom-up or an interactive process?(*Perception & Psychophysics, 59,* 1280—1296.)

Vecera, S. P., Strayer, D. & Chamberlain, S. (1996, November). *ERP indices of object-based attention.* (Poster presented at the 37th Meeting of the Psychonomic Society,Chicago.)

Watanabe, T. (1995). Orientation and color processing for partially occluded objects.(*Vision Research, 35,* 647—655.)

Watt, R. (1988). *Visual processing: Computational, psychological, and cognitive research.* (Hillsdale, NJ: Erlbaum)
Wertheimer, M. (1955). Laws of organization in perceptual forms.(In W. D. Ellis (Ed. and Trans.), A sourcebook of Gestalt psychology (pp. 1—11). London: Routledge & Kegan Paul. (Original work published 1923))
Witkin, A. P. & Tenenbaum, J. M. (1983). On the role of structure in vision.(In J. Beck, B. Hope, & A. Rosenfeld (Eds.), Human and machine vision (pp. 481—513). New York: Academic Press.)
Yantis, S. (1995). Perceived continuity of occluded visual objects.(*Psychological Science, 6,* 182—186.)
Yantis, S. & Moore, C. (1995, November). *Spread of visual attention behind an occluding surface.* (Paper presented at the 36th Annual Meeting of the PsychonomicSociety, Los Angeles.)
Zemel, R. S., Behrmann, M., Mozer, M. C. & Bavelier, D. (1998, April). *Experience-dependent perceptual grouping and object-based attention.* (Paper presented at the 1998 annual meeting of the CognitiveNeuroscience Society, San Francisco.)
Zemel, R. S., Williams, C. K. I. & Mozer, M. C. (1995). Lending direction to neural networks.(*Neural Network, 8,* 503—512.)
Zimba, L. & Tellinghuisen, D. J. (1990, November). *The covert orienting of attention to stereoscopic targets.* (Poster presented at the 31st Annual Meeting of the PsychonomicSociety, New Orleans, LA.)

1

We thank Allison Sekuler for suggesting this experiment to us.

2

It also was suggested to us that even though collinearity was removed in this experiment, symmetry was introduced. It is unlikely that symmetry, however, was driving this effect. As we show in Experiment 2b, the occluded object was treated as being functionally equivalent to the single object. Also, as mentioned in the Discussion section (see Figures 13 b and 13c), when one uses a symmetrical display that does not have an explicit representation of objects, one does not see the single-object advantage.

3

We thank Steve Yantis for this suggestion.

Figure 1. Examples of X displays from the six conditions of Experiment 1. The left and right columns indicate same and different judgments, respectively, and the rows from top to bottom indicate the single-, two-object, and occluded conditions, respectively.

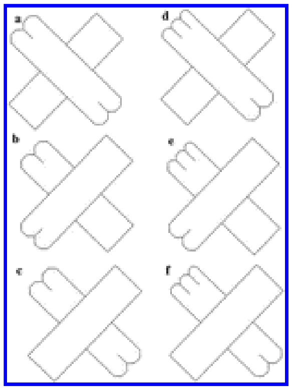

Figure 2. Mean of median reaction times and standard error bars (with error rates in parentheses) for single-, occluded, and two-object conditions as a function of judgment for X displays.

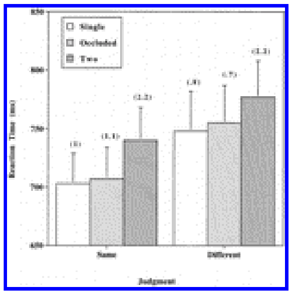

Figure 3. Examples of V displays from the six conditions of Experiment 2. The left and right columns indicate same and different judgments, respectively, and the rows from top to bottom indicate the single-object, two-object, and occluded conditions, respectively.

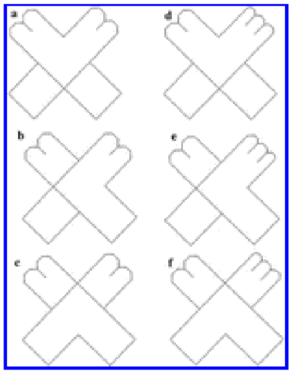

Figure 4. Mean of median reaction times and standard error bars (with error rates in parentheses) for the single-, occluded, and two-object conditions as a function of judgment for V displays.

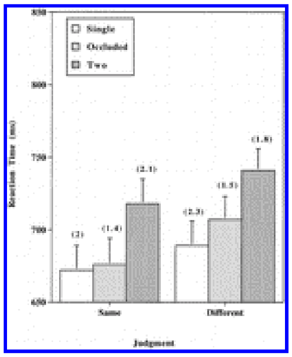

Figure 5. Mean of median reaction times and standard error bars for X (left panel) and V (right panel) displays as a function of condition (single, occluded, and two object) and judgment for Experiment 3. Error rates are in parentheses.

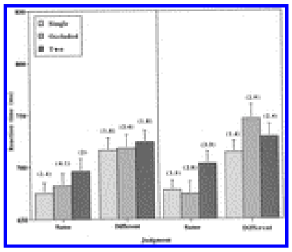

Figure 6. Examples of displays of the single-, occluded, and two-object conditions from Experiment 4: (a) the standard form; (b—d) increasing displacements of the two bars of the occluded object.

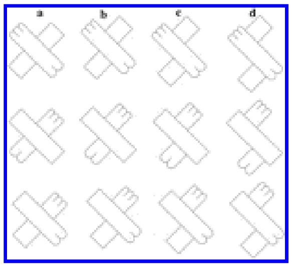

Figure 7. Mean of median reaction times and standard error bars (with error rates in parentheses) for the single-, occluded, and two-object conditions as a function of increasing displacement of the occluded bars for Experiment 4.

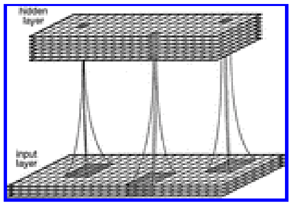

Figure 8. The architecture of MAGIC. The lower (input) layer contains the feature units; the upper layer contains the hidden units. Each layer is arranged in a spatiotopic array with several different feature types at each position in the array. Each plane in the feature layer corresponds to a different feature type. The shaded hidden units are reciprocally connected to all features in the corresponding shaded area of the feature layer. The lines between layers represent projections in both directions.

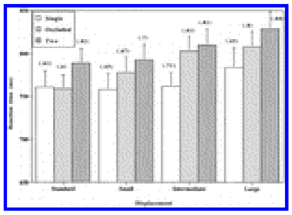

Figure 9. Example of MAGIC segmenting an X display corresponding to the displays used in Experiment 1. The iteration refers to the number of times activity flowed from feature units to the hidden units and back. The phase value of a feature is represented by a level on a gray-scale continuum, as shown at the bottom of the display. The cyclic phase continuum is approximated only by a linear gray-level continuum, but the basic information is conveyed nonetheless.

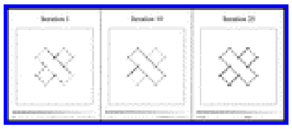

Figure 10. Mean number of iterations and standard errors for the single-, occluded, and two-object conditions for MAGIC making local feature decisions on the standard displays.

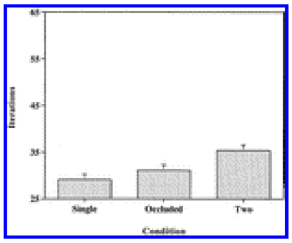

Figure 11. Example of MAGIC segmenting displays in which the two occluded bars are displaced. As is evident at Iteration 60, there are three objects present in the displays: the two ends of the occluded bar and the modal, occluding object. The presence of three separate objects is also reflected on the gray-scale continuum.

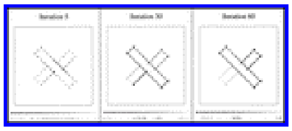

Figure 12. Mean number of iterations and standard errors for single-, occluded, and two-object conditions for MAGIC making local feature decisions on displacement displays in which the bars of the occluded object are no longer collinear.

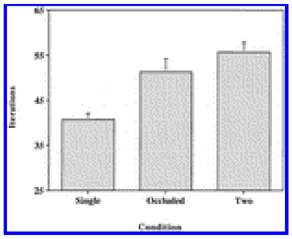

Figure 13. Examples of displays that participants would be tested on in experiments evaluating perceptual learning and object-based attention. (a) Displacement display in which occluder-occluded relations are reversed. (b) and (c) Ends-only display containing only the corners of the X display for feature judgments, where (b) corresponds to the single and occluded displays and (c) corresponds to the two-object display.