

## Occlusion, Symmetry, and Object-Based Attention Reply to [Saiki \(2000\)](#)

**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

[J. Saiki \(2000\)](#) argued that, because the stimuli used by [M. Behrmann, R. S. Zemel, and M. C. Mozer \(1998\)](#) were confounded by symmetry, conclusions about whether amodally completed objects can benefit from object-based attention are unwarranted. Here, the authors examine J. Saiki's claim further and expand on their view of the mechanisms underlying object-based attention, suggesting that perceptual organization is the process whereby features from a single object are selectively attended. In light of this, they claim that heuristics such as symmetry and collinearity play an important role in the facilitation of features from a single object. In support of this claim, they present data from a further experiment using displays that exploit common fate, another grouping heuristic, and show that, under these conditions, the hallmark of object-based attention, a single-object advantage, is obtained for the occluded (amodally completed) shapes.

[Saiki \(2000\)](#) argued that the hypothesis put forward in our earlier study ([Behrmann, Zemel, & Mozer, 1998](#)) that visual attention can be directed to separate segments of an occluded object is not warranted based on our findings. We conducted a series of experiments showing that participants process two features faster when they appear on a single object than when they appear on two different objects. This single-object advantage holds irrespective of whether participants make decisions about individual features of the objects (e.g., whether bumps that appear at two of the four possible ends are the same in number or not) or make decisions about the objects per se (e.g., whether two sets of bumps fall on a single object or not). This single-object advantage holds even when the features of a single object are further apart in space than the features of two different objects. The single-object advantage was demonstrated with two types of stimuli, referred to as X displays ([Behrmann et al., 1998](#), Experiment 1) and V displays ([Behrmann et al., 1998](#), Experiment 2). Examples of these single- and two-object Xs and Vs are shown in the first two rows of [Figure 1](#), with the columns illustrating stimuli that require *same* and *different* responses in the number-of-bumps task. These findings are taken to suggest that features of a single object may be selectively attended and their processing facilitated, and they are consistent with many existing studies on object-based attention (e.g., [Driver & Baylis, 1989](#); [Duncan, 1984](#); [Goldsmith, 1998](#); [Kramer & Jacobson, 1991](#); [Lavie & Driver, 1996](#)).

To determine whether object-based attention also applies to objects that are occluded, we included in our earlier study displays in which participants made judgments about features appearing on the ends of two bars of a single but occluded object (see [Figure 1](#), third row). The major finding both with the X and V displays was that the reaction times (RTs) were equivalent for single and occluded objects with each being significantly faster than in the two-object condition. On the basis of these results, we suggested that attention can be directed to an amodally completed object and its features selectively facilitated as is true in the case of a single, completed object.

In addition to these empirical studies, we also conducted computational simulations using a previously implemented neural network model, Multiple-object Adaptive Grouping of Image Components (MAGIC), that learned to bind together oriented line features into a coherent object ([Mozer, Zemel, Behrmann, & Williams, 1992](#)). MAGIC successfully reproduced our earlier human behavioral data ([Behrmann et al., 1998](#)) and provided a computational account of the mechanism mediating object-based attentional selection. Our argument, instantiated by MAGIC, was that an essential part of object-based selection is a process in which features that belong together are grouped together and are gated through to later stages of processing as a unified set. The gating comes about because features of a single object compete with other features and, through a winner-take-all mechanism, are selected and ultimately enhanced relative to the unselected features.

[Saiki \(2000\)](#) suggested that the empirical data that we obtained in our earlier study ([Behrmann et al., 1998](#)) do not compel this interpretation and may be more easily and simply explained by an artifact in the displays. Specifically, he argued that the central manipulation that we exploited—objecthood—is confounded with symmetry. As is evident from the V displays in [Figure 1](#), both the single and occluded objects are symmetrical about the midline, whereas this is not the case for the two-object condition (nor is it perfectly true for any of the *different* judgments, and the data from the *different* judgments are not as strong as for the *same* judgments in [Behrmann et al., 1998](#)). Because symmetry is such a powerful cue for perceptual organization ([Baylis & Driver, 1995a](#); [Baylis & Driver, 1995b](#); [Rock, 1983](#)), [Saiki](#) suggested that the performance advantage emerges for the single-object and occluded conditions but not for the two-object condition not by virtue of an object-based selection process but by virtue of symmetry. To substantiate his claim, [Saiki](#) first described a replication of the V experiment using the identical stimuli <sup>1</sup>and, thereafter, described a second experiment using modified V displays in which one of the two bars is extended, making the displays asymmetrical. The most important result from this experiment is that, when the displays are asymmetrical, the single-object advantage is not obtained for the occluded object.

In this reply, we first address the specific claims made by [Saiki \(2000\)](#) and explore their implications for theories of object-based attention. We then present data from a new experiment that further support the integral role of grouping processes and perceptual organization in object-based attention. In this experiment, we used displays that involve another principle of perceptual organization, *common fate*, which holds that items that move together tend to be grouped together. The results of this experiment reveal that when two spatially disparate and misaligned bars of an occluded object move together in phase, object-based attention can be directed to the two parts of an occluded object, producing RTs that are equivalent to those obtained for single objects. We discuss the relationship between object-based attention and perceptual organization and argue that these are integrally connected and that, in the process of extracting structure from a display, features of a coherent, unified object come to be processed more quickly. In this sense, there is no clear separation between object-based attention and perceptual organization, and we propose that the facilitated processing of features belonging to a single object is an emergent property of perceptual organization.

### The Role of Symmetry: Confound or Contributor?

In his commentary on our earlier study ([Behrmann et al., 1998](#)), [Saiki \(2000\)](#), p. 426 not only argued that symmetry may explain the object-based attention results but suggested that it may even have stronger explanatory power than the account that we provided. One criticism that [Saiki](#) made of our interpretation is that, on the basis of previous literature, one might have expected to see a difference in performance between single and occluded objects and this was not observed in our study. He cited the elegant findings from [Sekuler and Palmer's 1992](#) study showing that amodal completion takes roughly 200 ms and, in light of this, argued that there should be a difference between the amodal, occluded objects and the single objects. Because there was no RT difference between these two conditions in our study, he suggested that either amodal completion of the occluded displays must be instantaneous or that the time for amodal completion is not reflected in the RT. The absence of a difference between single and occluded RTs in our study is not that perplexing, however, given the experimental conditions: Displays remained on the screen for an unlimited duration and were not masked, and a trial was terminated at response. These conditions should provide sufficient

time for amodal completion to occur. Also, the base RTs in this experiment are on the order of 700 ms, more than enough time for amodal completion to be consolidated and to be fully incorporated into the participants' RT. A further possibility is that amodal completion can occur in parallel with other early processes and, therefore, need not incur additional time cost relative to single-object trials, hence the absence of any difference in RTs. Given these explanations, it is unnecessary to expect a difference in response times between single and occluded objects in our account, and the absence of such a finding does not undermine our conclusions or our object-based attentional account per se.

Saiki (2000) suggested that further support for the possible superiority of the symmetry account over our earlier account (Behrmann et al., 1998) is that the phenomenological impression of amodal completion is not very strong in the V displays in comparison with the X displays. Consequently, if amodal completion was responsible for the findings, one might have expected a difference between participants' performance on the X versus the V displays. Saiki was correct in noting that there is no difference in participants' performance on the occluded X and V displays in either base RTs or in their differences relative to the single- and two-object conditions. It is important to note that each participant was exposed to single-, occluded, and two-object trials randomly intermixed in a single block of trials, and this might have affected the phenomenological experience of the participants. That participants complete the amodal, occluded V display having been exposed to the single V display is not surprising, and, on questioning at the end of the experiment, participants maintained that this is how they had perceived the occluded Vs (the data from Experiments 1c and 2c in our earlier study, in which participants' phenomenological impressions are probed, also support this view). Thus, within the context of an experiment, the amodal completion of Xs or Vs was sufficiently robust to bring about the obtained results, and participants were not required to make comparisons of the relative robustness of the occlusion for Xs versus Vs. Again, this claim by Saiki does not necessarily undermine our account.

But perhaps most critical for Saiki's (2000) argument is his demonstration that when asymmetrical displays are presented to participants, no single-object advantage is obtained for the occluded display. In this experiment (Experiment 2a), participants performed the number-of-bumps comparison task on symmetric and asymmetric single-, occluded, and two-object displays randomly mixed in a block of trials. The absence of a single-object advantage for the occluded asymmetrical display, relative to the two-object display, is taken as evidence that it is only when objects are symmetrical that this facilitation is observed. It is interesting, however, that for the asymmetrical trials there is also no advantage for the single-object over the two-object condition, as one might have expected. It is unlikely that this basic effect depends on symmetry as there are several instances in the literature in which the displays are not symmetrical and yet this advantage is obtained (e.g., Baylis & Driver, 1993; Duncan, 1984). Why this more basic effect is not obtained with these displays is not clear. It is also the case that for the symmetric displays, there is neither a significant advantage for the single objects over the two objects (again, as might be predicted from the fairly extensive literature on this topic) nor an advantage for the occluded displays over the two-object displays (which is expected both from our earlier results and from Saiki's replication of our results in his Experiment 1). Saiki explained the absence of these effects for symmetrical displays as being a result of reduced salience of symmetry because participants could adopt the symmetry strategy on only half the trials. However, it is not clear how to interpret the findings from this experiment for both symmetrical and asymmetrical displays in the absence of the expected base effects.

We now consider the broader implications of Saiki's (2000) claims. Although his comments were largely, although not entirely, directed at the results from the V displays, we need to consider the full ramifications of his critique in order to understand whether the object-based attentional account is indeed challenged. We do this by examining some predictions that would follow if it were indeed the case that symmetry could account for the full range of existing data. One obvious prediction is that if symmetry was a strong predictor of effects of object-based attention, then we would never expect to obtain an object advantage for displays that are not symmetrical. This is not the case, however. From an examination of the X displays in Figure 1, it is evident that the single- and occluded-object displays are diagonally rather than vertically or horizontally symmetric. We know, however, that diagonal or oblique symmetry is not particularly powerful or robust (Barlow & Reeves, 1979; Rock & Leaman, 1963; Wagemans, Van Cool, & D'Ydewalle, 1992), and naive observers often do not perceive symmetry in these displays. Yet a robust single-object advantage is still obtained with these displays, both for the single and occluded objects. Saiki discounted the results from the X displays, however, as not germane to the issue as they are subject to a second, different confound, that of collinearity or good continuation. As is evident from viewing the X displays, the single- and occluded-object displays but not the two-object condition share collinearity. Indeed, in our earlier study (Behrmann et al., 1998), we used both the X and the V displays to show and to demonstrate the generality of our account and to argue that object-based attention emerges from the benefit afforded a single (or occluded) object by virtue of its perceptual organization. Our interpretation of both the collinearity and symmetry results is that neither is the by-product of an experimental confound. Instead, the fact that one obtains facilitation in performance with an occluded object under conditions of either collinearity or of symmetry suggests that both of these factors play a role in what we want to call *object-based attention* and that it is by virtue of the grouping processes, afforded by these heuristics, that the features of a single (even if occluded) object are enhanced.

In sum, Saiki's (2000) assessment of our findings and interpretation is not wrong in general, but it is too narrowly stated; we concur that symmetry is important but argue that a host of other grouping processes are also important and that the same object-based attentional effects can be obtained when the stimuli are organized according to other perceptual principles. To further support our claim, we now present an experiment that demonstrates the contribution of common fate to the single-object advantage, and then we return to our account of the mechanisms mediating object-based attention.

### Experiment: Object-Based Attention in Displays With Common Fate

This experiment was designed to demonstrate that the single-object advantage is obtainable under conditions in which common fate (Wertheimer, 1923/1950), a heuristic of perceptual organization other than collinearity and symmetry, is used. Specifically, we demonstrate that the two spatially discontinuous and misaligned bars of an occluded object, which are normally treated as two separate objects, are bound together by common motion and treated as a single object. This single-object advantage is reflected by performance that is equally good when participants judge features on the ends of the two misaligned bars and when they judge features on a single unoccluded object.

The display we used is termed the Z display (see Figure 2 for examples of single-, occluded, and two-object Z displays). We have demonstrated previously that when participants are shown the occluded Z display with misaligned bars, their response time to judge the number of bumps at the ends of the bars is slower than in the single-object condition, but it is not different from the two-object condition. In that experiment (see Behrmann et al., 1998, Figure 7, for the data), there was a significant 41-ms difference between single and misaligned occluded objects and a 7-ms difference between misaligned occluded and two-object displays for the intermediate displacement condition. In a replication of this experiment, Zemel, Behrmann, Bavelier, and Mozer (2000) showed two groups of participants displays corresponding to the single and misaligned occluded stimuli and found differences of 30.4 ms and 37.9 ms between the two types of stimuli for the two groups. We have suggested that the difference between the single and misaligned occluded Z displays and the equivalence between the latter and two-object displays arise because participants do not perceive the parts of the misaligned display as belonging to the same object, that is, they are not relatable (Kellman & Shipley, 1991, 1992). Note that the absence of a single-object advantage for the misaligned occluded Z trials is not simply an artifact of participant sampling (Behrmann et al., 1998, Experiment 4); when the bars were realigned so that the contours were collinear, turning them into the X displays, the very same participants reproduced the single-object advantage and made feature decisions for single and occluded objects rapidly (761 ms and 759 ms) and significantly faster than in the two-object condition (789 ms).

In the current experiment, we first determined whether we could replicate the absence of a single-object advantage with misaligned Z displays (using slightly different misaligning parameters for the occluded displays from Behrmann et al., 1998 as is explained below). Thereafter, and more important, we explored, within the same participants, whether when these misaligned bars are moved together in phase, participants exploit this common fate principle to bind the bars together such that a single-object advantage emerges (even though there is neither symmetry nor collinearity in the images). If this is so, then, under moving but not static conditions, participants would make equally fast judgments for the single-object and misaligned occluded conditions and both of these would be significantly faster than in the two-object condition.

A similar type of experiment using a display consisting of an oblique bar occluded by a center bar (yielding two disconnected parts) has been used to show that infants perceive a partly hidden object as a connected unit if the ends of the object move together behind an occluder (Spelke, 1990). In these experiments, however, the bars of the occluded rectangle were not misaligned, and so we do not know whether nonrelatable bars can be interpreted as belonging to a single object.

## Method Participants.

Twenty university students (11 men and 9 women, mean age = 20.9 years) volunteered to participate. All participants were recruited from electronic bulletin boards and were paid for their involvement. Volunteers had normal or corrected visual acuity by self-report, and all but 2 were right-handed.

## Apparatus and materials.

The experiment was conducted on a Macintosh Quadra 650 computer. Stimuli were presented on a 15-in. (38.1-cm) color monitor using PsyScope experimental software, Version 1.2.1 (Cohen, MacWhinney, Flatt, & Provost, 1993).

The displays were presented as black-and-white line drawings on a white background. Viewing distance was approximately 50 cm. The displays contained a center rectangle, attached to which were two smaller rectangles, displaced 7.5 cm (8.94°) in opposing directions. The large rectangular bar was 8.7 cm in length (10.2°) and 2.5 cm (2.9°) in width, and the smaller bars were half that length. The display parameters fell in between the intermediate and large displacement conditions that we used in our earlier study (Behrmann et al., 1998, Experiment 4). This was necessary to allow for enough room for the misaligned bars to slide up and down the central bar in the moving condition, but it makes the results difficult to compare directly to either of those two conditions in our earlier study. On each trial, the features (bumps) appeared at any two of the four ends of the two rectangles. The feature ends were divided into two equal parts for the two-bump and into three equal parts for the three-bump displays.

The displays fell into three different conditions: (a) single (or unoccluded) object, in which the two sets of bumps appeared at each end of the single center rectangular bar; (b) two objects, in which one of the two sets of bumps appeared at one end of the central rectangle and the other at one end of a noncentral rectangle; and (c) occluded object, in which the two sets of bumps appeared at each end of the displaced noncentral rectangular bars. Examples of the displays appear in Figure 2, with the columns and rows illustrating the judgments and different conditions, respectively.

On *same* trials, there were either two bumps at each of two ends or three bumps at each of two ends, and there were an equal number of each of these *same* trials. On *different* trials, there were always two bumps on one end and three bumps on the other, and the locations of the two and three bumps were evenly counterbalanced. Participants indicated their response on a button box that interfaces with PsyScope using their two index fingers; for half the participants, *same* was indicated with the left hand and *different* with the right, and the mapping was reversed for the other half.

There were two major display types: static and moving. In the static trials, the displays appeared as they do in Figure 2 and remained on the screen for an unlimited duration until participants made their feature judgments. In the moving trials, the display appeared on the screen for 200 ms and then was immediately replaced with a display in which the displaced smaller rectangle ends shifted 2.5 cm (2.98°) in the same direction along the single object. After an additional 200 ms, the initial object would replace the adjusted object for 200 ms, and this sequence would continue until the participant responded. The arrangement yielded the perception of apparent motion such that the occluded, displaced rectangular bars slid up and down in phase along the nondisplaced rectangle. Participants could make their decisions at any point during the display, but because of the speed of the movement, several iterations of phase change occurred before any response could be emitted.

We predicted that response speed would be relatively fast for the bumps on the center rectangular single object and slow for the two-object condition and that this would be so irrespective of whether the display was static or moving. The critical condition is the occluded display in which the two misaligned bars either move together up and down or remain static. The prediction was that in the moving condition, the RTs will be equivalent to those of the single-object condition by virtue of common fate, but in the static displays, they will be more like the two-object condition.

## 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 conducted in six blocks of 96 trials each, and each block contained all appropriate crossings of the variables that were randomized within a block. This design allowed for a total of 576 trials, 288 in each of the static and moving blocks. The first three blocks contained only static displays, followed by three blocks of moving displays. This order could not be counterbalanced because the effect of experience with the moving displays might have altered the participants' interpretation of the static displays (Zemel et al., 2000). The participants had a short break between the blocks. Before the first static and the first moving block, participants completed 24 practice trials (static or moving only depending on the experimental trials to follow), which contained all possible trial types presented in the upcoming block.

## Design.

The design was entirely within subjects, with the independent variables being condition (single, two, or occluded), judgment (*same* or *different*), and display type (moving or static). There were an equal number of trials drawn from each of the three conditions, an equal number of *same* and *different* trials, and an equal number of moving and static trials.

## Results and Discussion

The data from the practice trials were discarded from the analysis. The remaining data were then collapsed across the six experimental blocks, and the error trials were excluded from the RT analysis. The mean RT and mean error for each crossing of judgment, condition, and feature activity were calculated for each participant and were then subjected to analyses of variance (ANOVAs) with Tukey post hoc tests set at  $p < .05$  to examine pairwise comparisons in the case of significant interactions.

The three-way repeated measures ANOVA with error rate as the dependent measure revealed no significant effects with means of 2.5%, 2.29%, and 2.44% errors for the single-, occluded, and two-object displays, respectively, and no difference in error rate across static and moving displays or across *same* and *different* trials (all  $F$ 's  $< 1$ ). No interactions were observed either. In contrast, the ANOVA on the correct RTs revealed a significant three-way interaction between condition, judgment, and display type,  $F(2, 38) = 6.8$ ,  $MSE = 623.4$ ,  $p < .01$ . The results of this interaction are plotted in Figure 3 separately for *same* and *different* trials with the means for the single-, occluded, and two-object conditions shown as a function of static versus moving display type. The two-way ANOVA of condition by display type, conducted on the *same* and *different* responses separately, revealed a significant interaction in both cases: *same*,  $F(2, 38) = 5.6$ ,  $MSE = 795.8$ ,  $p < .01$ ; *different*,  $F(2, 38) = 3.3$ ,  $MSE = 412.9$ ,  $p < .04$ . As is evident from Figure 3, however, we do not see the predicted pattern of data for the *different* trials but only for the *same* trials. The failure to obtain replicable findings with *different* judgment trials is well recognized already in our earlier study (Behrmann et al., 1998) and by Saiki (2000). Exactly why this is so remains a puzzling but separate issue. In light of this, we focus only on the data from the *same* trials, but before we do so, we present the rest of the data from the three-way analysis.

The two-way interaction between judgment and condition was significant,  $F(2, 38) = 7.5$ ,  $MSE = 871.4$ ,  $p < .001$ , reflecting the smaller increase in RT between single and occluded trials for *same* but not *different* judgments. The remaining two-way interaction between judgment and display type,  $F(1, 19) = 3.9$ ,  $MSE = 1,824.3$ ,  $p = .06$ , was marginally significant. The main effect of judgment was significant,  $F(1, 19) = 12.5$ ,  $MSE = 3,000.3$ ,  $p < .01$ , reflecting the well-known advantage for *same* over *different* judgments (Nickerson, 1965). There was also a significant main effect of condition,  $F(2, 38) = 23.5$ ,  $MSE = 1,392.4$ ,  $p < .0001$ , highlighting the overall advantage of 17 ms for single objects over occluded objects and of 14 ms for occluded objects over two

objects. There was no main effect of display type, revealing equivalent base RTs overall for the static and motion displays. No other effects were significant.

We now consider only the *same* trials. The first important result is that, in the static condition, the RTs for the occluded trials (641 ms) are significantly slower than for the single-object trials (613 ms), replicating our previous findings (Behrmann et al., 1998; Zemel et al., 2000). In contrast, in the motion displays, there is no significant difference between the occluded (618 ms) and single (626 ms) trials, reflecting the single-object advantage. We should also note that there is no statistically significant difference between the static and moving single-object condition and that the main effect comes about from the change in the moving occluded condition. This is the pattern we had predicted. The second issue concerns the relationship of the occluded condition and the two-object condition. Although there was a difference between the two-object condition (658 ms) and the single and occluded conditions for the moving displays, as predicted, this difference also held for the static displays (two-object condition, 683 ms). This latter finding was not expected. Interestingly, this difference in static displays also held in Experiment 4 of our earlier study (Behrmann et al., 1998) for the large displays (808 ms, occluded, vs. 829 ms, two object) but not for the intermediate displays (803 ms, occluded, vs. 810 ms, two object), and the displays used here fall exactly in between these two display types in the extent of the displacement. Exactly why the misaligned occluded displays appear to be closer to the single-object than the two-object displays in the static condition here and in the large displays in our earlier study is unclear. One possible explanation is that the static two-object condition is disproportionately long for some unexplained reason. A second possibility is that the appearance of bumps on both of the occluded bars aids the participants in perceptual organization and hence facilitates the perception of the occluded display relative to the two-object condition in which bumps appear on only one rectangle (cf. occluded and two-object displays in Figure 2) even for static displays. This facilitation for the occluded object might be most apparent in larger displays as RTs are longest here (see Figure 7 in Behrmann et al., 1998, and Figure 3 here). This latter possibility is consistent with our account that perceptual organization can enhance perception in several different ways by using whatever heuristics might be applicable. What is most relevant, however, is that the occluded motion condition is not only equivalent to the single-object motion condition but is also significantly faster than the occluded static condition.

## General Discussion

In one experiment reported here as well as in two previous experiments and replications of these (Behrmann et al., 1998; Zemel et al., 2000), we have shown that the advantage for judging features that appear on a single object versus two different objects (the single-object advantage) is observed under a variety of display and task conditions. More important, this advantage applies equally for occluded objects that require amodal completion and for single, complete objects. Furthermore, this advantage is obtained when the displays are symmetrical, when they contain collinearity, and when they exploit common fate or motion as a grouping principle. This advantage is assumed to reflect object-based attention, that is, the enhancement afforded features of a single object that is selected for preferential processing.

Before discussing the mechanism of object-based attention further, it is worth noting that the terms *object-based* and *object-based attention* have been used in different ways in the literature, and each term has consequently acquired several connotations over time. The term *object-based* as we are using it here differs from the stricter usage in which *object-based* is equated with the use of spatially invariant, object-centered representations (Marr, 1982). The latter interpretation argues that the representations mediating object-based facilitation are invariant over translation and rotation and therefore denies the possibility that spatial representations might moderate an object-based effect (Kramer, Weber, & Watson, 1997; Vecera, 1994, 1997). Our use of the term *object-based* does not preclude spatial variables (and one might even consider proximity as simply another powerful grouping heuristic) from participating in object-based attention, and so we use the term simply to indicate that properties of the object structure are built into the representation that is attended.

The term *object-based attention* is also confusing, and it too has acquired multiple meanings. This unfortunate situation may have arisen in part because the mechanism by which the features of a single object are enhanced remains undetermined. One usage of the term *object-based attention* refers to the finding that spatial attention spreads or radiates along or within the boundaries delimited by an object. On this view, the attention deployed to objects is primarily spatially mediated in that activation in one location enhances activation in contiguous locations within the spatial region delineated by an object. For example, many studies (Abrams & Law, in press; Egly, Driver, & Rafal, 1994; Gibson & Egeth, 1994; Vecera & Farah, 1994) have shown that participants can detect a probe or make target discriminations better when the probe or target appears at a location that is validly cued. Interestingly, although performance is not as good for targets appearing in uncued locations, it is significantly better in the invalid on-object location (when the probe or target appears at an uncued location within the confines of the cued object) than in the invalid but spatially equidistant location on a second, uncued object. In these studies, features of the object per se are not being judged; rather, there is facilitation for processing information that appears within the spatial boundaries of an object to which attention has been directed (even when there is amodal completion; Moore, Yantis, & Vaughan, 1998). Although this is often interpreted as a spatial spread of attention, this phenomenon might also be described as arising from processing a common region, another Gestalt principle (Palmer, 1999; Palmer & Rock, 1994).

A different sense in which *object-based attention* is used is in reference to the findings by Treisman and colleagues (Treisman, 1992; Treisman & De Schepper, 1996), who used the review paradigm. In this paradigm, RTs are faster for naming a letter when it appears in the probe inside the same shape in which it appeared in the prime (or preview) relative to naming it when it appears in a different shape in the prime and in the probe conditions. The explanation for this result is that the initially associated shape and letter are stored as a temporary episodic object file and, when this is reactivated, the speed with which the items are processed is enhanced. This advantage for previewed items is also referred to sometimes as a consequence of object-based attention. We should note that a similar interpretation is also given to the results of visual indexing studies by Pylyshyn and colleagues in which participants track up to 5 independently moving objects in a field of 10 other identical objects and can indicate whether a subsequent probe occurred on a target or nontarget object (Pylyshyn & Storm, 1988).

A final usage of the term—and the one that we emphasize—is that features belonging to a single perceptual object are enhanced by virtue of perceptual organization processes. Our claim is that object-based attention or the ability to selectively attend to features of a single object is an emergent property of perceptual organization. Through segmentation and grouping processes, the perceptual organization of an image (even when the elements are spatiotemporally fragmented) affords faster processing of features that belong together than features that do not belong together. Moreover, features that are grouped together compete against other feature groupings, causing one to beat out its competitors. Over time, the features that belong together become integrated and form a perceptual object. We have shown that, through experience, MAGIC comes to discover heuristics (including, e.g., symmetry and collinearity, although it would undoubtedly exploit common fate in moving displays, too) that reliably indicate how features are grouped into units, and it then uses these heuristics to parse the display into discrete chunks. By virtue of this competition, the features of individual chunks are then processed as a unit in a winner-take-all fashion. In sum, this view suggests that object-based attention is a dynamic process in which elemental features cooperate and are bound together and the attentional facilitation is an emergent property of this binding.

Although we have emphasized differences in usages of the term *object-based attention*, these alternatives need not be mutually exclusive. Thus, the conception of perceptual organization processes that facilitate single-object features is not mutually exclusive with that of a spread of spatial attention within the confines of the object. Similarly, perceptual organization clearly plays a role in deriving the initial association between the letter and shape in the object file experiments. The emphasis of the third view, however, and the one that is most relevant for understanding Saiki's (2000) objections, is on the extent to which object-based (Gestalt) structure is built into the representations that are selectively enhanced (see also Goldsmith, 1998).

The idea that Gestalt principles or perceptual grouping heuristics play an important role in attentional processing is not novel, and a similar claim has been made in many previous studies. For example, in one classic study Kramer and Jacobson (1991) exploited similarity (of color or texture) in a task in which participants made attribute judgments of features of a central line that was either connected or not connected with other lines in the image. Notwithstanding the fact that the other lines were irrelevant for the task, participants' ability to judge the attribute of the central line was influenced by whether the irrelevant information could be grouped with the central line. When a target appeared on the same perceptual unit as distractors, response compatibility effects increased (facilitated performance if the lines shared a dimension such as the color or texture or inhibited it if they did not) relative to when targets and distractors appeared on different perceptual units. These findings led Kramer and Jacobson to conclude that structural aspects of the display can amplify costs and benefits in a response compatibility paradigm and, moreover, that these effects are contingent on the quantitative metric of grouping strength; elements are organized along a continuum of strength, and when elements are strongly grouped, stronger costs and benefits are obtained relative to when the grouping is weaker. Many other studies have replicated

these basic effects (e.g., [Bacon & Egeth, 1991](#) ; [Baylis & Driver, 1992](#) ; [Fox, 1998](#) ; [Lavie & Driver, 1996](#)).

The grouping of image elements and its impact on visual selection is also noted in the neuropsychological literature. Patients with hemispatial neglect who might otherwise fail to detect a contralateral item might be able to report that same item with greater frequency when the left and right items can be grouped together by similarity or even element connectedness. For example, following a right-hemisphere lesion, patients appear able to respond to the left-sided item when it is symmetrical with the one on the right ( [Driver, Baylis, & Rafal, 1992](#) ; [Ward, Goodrich, & Driver, 1994](#) ) or when the left and right items share color, brightness, or collinearity ( [Driver & Halligan, 1991](#) ; [Gilchrist, Humphreys, & Riddoch, 1996](#) ; [Rorden, Mattingley, Karnath, & Driver, 1997](#) ). This benefit from grouping is also obtained with incomplete objects as the left-sided item is detected better when it can be grouped with the right by an illusory contour ( [Driver & Halligan, 1991](#) ; [Mattingley, David, & Driver, 1997](#) ) or under occlusion ( [Humphreys, 1998](#) ).

If it is indeed the case that object-based attention and perceptual organization are so integrally bound, two unanswered questions remain. The first concerns what types of elements are subject to organization and the second, how one formally defines what constitutes an object. Not all elements appear to have been created equal, with some operating as dimensions of features and some as parts of a whole; for example, although a box's tilt and brightness are clearly two dimensions of a single object, parts of a single object (like the occluded bars) are not two dimensions of the same object and yet both dimensions and parts show facilitation relative to a two-object condition. One way of understanding this general facilitation is to adopt a model that codes information at different levels of a hierarchy and that can facilitate the processing of parts at these different levels ( [Mozer, 1999](#) ). In a recent empirical demonstration of this, [Vecera, Behrmann, and McGoldrick \(2000\)](#) reported that participants show an object-based advantage simultaneously for features of a single object as well as for features of a single part of a single object; participants were able to report two attributes from a part of a single object better than two attributes coming from two different parts of a single object. These findings suggest that elements can be attended to at different levels of a structured hierarchy of perceptual objects and that, at each level, all features that are coded as properties of the same part or whole are facilitated in tandem (see also [Goldsmith, 1998](#) , Experiment 4; see also [Humphreys et al., 2000](#) , for a different view).

The second issue concerns what constitutes an object. We have applied the term *object-based attention* to stimuli such as the X, V, and Z displays, none of which are likely to have been encountered by the participants before the experiment. Are these to be considered legitimate objects with the same status as objects that are familiar and known to participants? We have suggested that any group of elements that cohere together is a legitimate perceptual entity and that the effects we obtain are not contingent on our knowledge of these as "real objects" in the world. Thus, perceptual organization defines an object by virtue of what features go with what other features. Although segmentation may be aided by knowledge of real objects relative to coherent but unfamiliar objects ( [Vecera & Farah, 1997](#) ), it is not restricted to known stimuli. We have also suggested that perceptual learning and experience play an important role in determining what features are grouped together and have demonstrated that MAGIC, a tabula rasa initially, comes to discover grouping principles that can be applied robustly and in the absence of lexical or semantic knowledge simply through extracting the statistical regularities of co-occurring features. As the participant (or MAGIC) encounters the stimuli repeatedly, the effects grow in strength, but, we argue, even novel objects are subject to perceptual organization and object-based attention because their features are grouped into a single unit.

In conclusion, we have argued that perceptual organization and attentional selection are deeply intertwined and that object-based facilitation is a flexible and dynamic process operating at multiple spatial scales and over familiar and unfamiliar objects. Perhaps our view is best expressed by [Duncan \(1984\)](#) , whose study triggered many of the current issues in which we are interested: Duncan suggested that in order to address further questions "the study of visual attention and of perceptual organization must proceed together" (p. 502). We argue that it is not possible to do otherwise.

## References

- Abrams, R. & Law, M. B. (2000). Object-based visual attention with endogenous orienting. *Perception & Psychophysics*, 62 , 818-833. [P](#)
- Bacon, W. F. & Egeth, H. E. (1991). Local processes in pre-attentive feature detection. *Journal of Experimental Psychology: Human Perception and Performance*, 17 , 77-90. [P](#)
- Barlow, H. B. & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, 19 , 783-793. [P](#)
- Baylis, G. C. & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics*, 51 , 145-162. [P](#)
- 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. [P](#)
- Baylis, G. C. & Driver, J. (1995a). One-sided edge assignment in vision: I. Figure—ground segmentation and attention to objects. *Current Directions in Psychological Science*, 4 , 140-146. [P](#)
- Baylis, G. C. & Driver, J. (1995b). Obligatory edge-assignment in vision: The role of figure- and part-segmentation in symmetry detection. *Journal of Experimental Psychology: Human Perception and Performance*, 21 , 1323-1342. [P](#)
- Behrmann, M., Zemel, R. S. & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal subjects and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*, 24 , 1011-1036. [P](#)
- Cohen, J. D., MacWhinney, B., Flatt, M. & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers*, 25 , 257-271. [P](#)
- Driver, J. & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance*, 15 , 448-456. [P](#)
- Driver, J., Baylis, G. C. & Rafal, R. D. (1992). Preserved figure—ground segregation and symmetry perception in visual neglect. *Nature*, 360 , 73-75. [P](#)
- Driver, J. & Halligan, P. W. (1991). Can visual neglect operate in object-centered coordinates? An affirmative study. *Cognitive Neuropsychology*, 8 , 475-496. [P](#)
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113 , 501-517. [P](#)
- Egley, R., Driver, J. & Rafal, R. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123 , 161-177. [P](#)
- Fox, E. (1998). Perceptual grouping and visual selective attention. *Perception & Psychophysics*, 60 , 1004-1102. [P](#)
- Gibson, B. S. & Egeth, H. (1994). Inhibition of return to object-based and environment-based locations. *Perception & Psychophysics*, 55 , 323-339. [P](#)
- Gilchrist, I. D., Humphreys, G. W. & Riddoch, M. J. (1996). Grouping and extinction: Evidence for low-level modulation of visual selection. *Cognitive Neuropsychology*, 13 , 1223-1249. [P](#)
- Goldsmith, M. (1998). What's in a location? Comparing object-based and space-based models of feature integration in visual search. *Journal of Experimental Psychology: General*, 127 , 189-219. [P](#)
- Humphreys, G. W. (1998). Neural representation of objects in space: A dual coding account. *Philosophical Transactions of the Royal Society of London, Series B*, 353 , 1341-1351.
- Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A. & Klempen, N. (2000). Fractionating the binding process: Neuropsychological evidence distinguishing binding of form from binding of surface features. *Vision Research*, 40 , 1569-1596.
- Kellman, P. J. & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23 , 141-221. [P](#)
- Kellman, P. J. & Shipley, T. F. (1992). Perceiving objects across gaps in space and time. *Current Directions in Psychological Science*, 1 , 193-199. [P](#)
- Kramer, A. F. & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception &*

*Psychophysics*, 50 , 267-284. [P](#)

Kramer, A. F., Weber, T. A. & Watson, S. E. (1997). Object-based attentional selection—Grouped-arrays or spatially-invariant representations? Comment on Vecera and Farah (1994). *Journal of Experimental Psychology: General*, 126 , 3-13. [P](#)

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

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

Mattingley, J. B., David, G. & Driver, J. (1997, January 31). Pre-attentive filling in of visual surfaces in parietal extinction. *Science*, 275 , 671-674.

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

Mozer, M. C. (1999). A principle for the unsupervised hierarchical decomposition of visual scenes. (In M. S. Kearns, S. A. Solla, & D. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11, pp. 52—58). Cambridge, MA: MIT Press.)

Mozer, M. C., Zemel, R., Behrmann, M. & Williams, C. (1992). Learning to segment by dynamic feature binding. *Neural Computation*, 4 , 647-672.

Nickerson, R. S. (1965). Response times for "same" and "different" judgments. *Perceptual and Motor Skills*, 20 , 15-18. [P](#)

Palmer, S. E. (1999). *Vision science: From photons to phenomenology*. (Cambridge, MA: MIT Press)

Palmer, S. E. & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, 1 , 29-55. [P](#)

Pylyshyn, Z. & Storm, R. W. (1988). Tracking multiple independent objects: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3 , 1-19.

Rock, I. (1983). *The logic of perception*. (Cambridge, MA: MIT Press)

Rock, I. & Leaman, R. (1963). An experimental analysis of visual symmetry. *Acta Psychologica*, 21 , 171-183. [P](#)

Rorden, C., Mattingley, J. B., Karnath, H.-O. & Driver, J. (1997). Visual extinction and prior entry: Impaired perception of temporal order with intact motion perception after unilateral parietal damage. *Neuropsychologia*, 35 , 421-433. [P](#)

Saiki, J. (2000). Occlusion, symmetry, and object-based attention: Comment on Behrmann, Zemel, & Mozer (1998). *Journal of Experimental Psychology: Human Perception and Performance*, 26 , 424-433. [P](#)

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14 , 29-56. [P](#)

Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, 47 , 862-875. [P](#)

Treisman, A. & De Schepper, B. (1996). Object tokens, attention, and visual memory. (In T. Inui & J. L. McClelland (Eds.), *Attention and performance 16: Information integration in perception and communication* (pp. 15—45). Cambridge, MA: MIT Press.)

Vecera, S. P. (1994). Grouped locations and object-based attention: Comment on Egly, Driver, and Rafal (1994). *Journal of Experimental Psychology: General*, 123 , 316-320. [P](#)

Vecera, S. P. (1997). Grouped arrays versus object-based representations: Reply to Kramer et al. (1997). *Journal of Experimental Psychology: General*, 126 , 14-18. [P](#)

Vecera, S. P., Behrmann, M. & McGoldrick, J. (2000). Selective attention to parts of an object. *Psychonomic Bulletin & Review*, ,

Vecera, S. P. & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: Human Perception and Performance*, 123 , 1-14.

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

Wagemans, J., Van Cool, L. & D'Ydewalle, G. (1992). Orientation effects and component processes in symmetry detection. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 44A , 475-508. [P](#)

Ward, R., Goodrich, S. & Driver, J. (1994). Grouping reduces visual extinction: Neuropsychological evidence for weight-linkage in visual selection. *Visual Cognition*, 1 , 101-129. [P](#)

Wertheimer, M. (1950). Untersuchungen zur Lehre von der Gestalt, II([Investigations on the theory of the Gestalt]) *Psychologische Forschung*, 4 , 301-350.

Zemel, R. S., Behrmann, M., Bavelier, D. & Mozer, M. C. (2000). *Experience-dependent perceptual grouping and object-based attention*. (Manuscript submitted for publication)

## 1

Note that the results obtained by Saiki are not exactly identical to those in our earlier study ([Behrmann et al., 1998](#), Experiment 2a). Whereas we found faster RTs for *same* than *different* trials as well as an overall advantage for single and occluded objects over two objects and no interaction between them, Saiki found no main effects, and a difference between single and occluded versus two objects emerged only for the *same* judgments. Also, Saiki obtained similar results in error rates (which were higher than in [Behrmann et al., 1998](#)), whereas we found no significant effects in the error data.

This work was supported by National Institute of Mental Health Grant MH57466. We thank Rutie Kimchi, David Plaut, and Mary Peterson for their helpful comments and Thomas McKeef 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](mailto:behrmann+@cmu.edu)

Received: June 9, 1999

Revised: October 1, 1999

Accepted: October 1, 1999

Figure 1. Examples of (a) X and (b) V displays used by [Behrmann, Zemel, and Mozer \(1998\)](#); columns indicate *same* and *different* trials, and rows reflect the single object, two-object, and occluded displays.

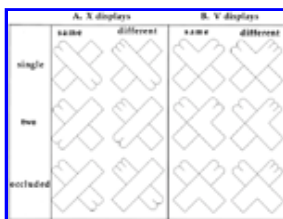


Figure 2. The misaligned Z displays; each of the two columns indicate *same* and *different* trials, and the rows reflect the single-, two-object, and occluded displays.

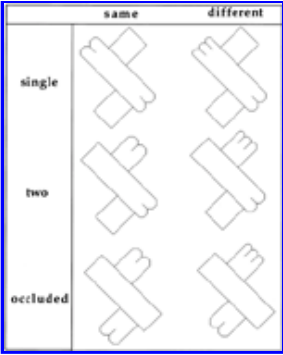


Figure 3. Mean reaction time (RT)  $\pm$  1 SE across participants for single-, occluded, and two-object Z displays for static and motion trials, plotted separately for *same* and *different* judgments.

