Experience-Dependent Perceptual Grouping and Object-Based Attention

Richard S. Zemel
University of Toronto
Toronto, ON M5S 1A4
zemel@cs.toronto.edu

Marlene Behrmann
Carnegie Mellon University
Pittsburgh, PA 15213
behrmann+@cmu.edu

Michael C. Mozer
University of Colorado
Boulder, CO 80309-0430
mozer@cs.colorado.edu

Daphne Bavelier
University of Rochester
Rochester, NY 14627
daphne@bcs.rochester.edu

Abstract

Attention can be allocated to objects as well as locations, even if the objects are partially occluded. A fundamental question concerns the nature of the 'objects' of this attention. Recent studies show that objects can be defined based on Gestalt grouping principles, or based on familiarity. The effects of both grouping and familiarity can be understood in terms of a more general hypothesis: that perceptual experience with particular feature combinations determines whether two features will be integrated as an object of attention. We present data from four experiments showing that recently experienced novel feature combinations gain the object attentional benefit and that this effect is realized by different feature combinations under a range of experimental conditions. These studies indicate that object attention is adaptive and responsive to the statistical structure of the environment.

Acknowledgements: We thank Susan Watt, Jim Nelson, and James Christopher for help with running subjects and analysis, and Mary Peterson for useful discussions, and constructive comments on this manuscript. We also thank Jim Enns and an anonymous reviewer for their helpful comments. Supported by the McDonnell Foundation grant JSMF 95-1 (RSZ), NIMH Program Project grant MH47566-06 (MB and RSZ), NSF award IBN-9873492 (MCM), and a McDonnell Foundation grant (DB).

Introduction

Attention can be directed toward objects as well as toward locations in the visual field, thereby affording preferential processing for the features of a specific object. Evidence for this finding comes from studies which show that it is difficult to attend to two objects simultaneously. For example, when judgements depend on two features in a display, responses are more rapid when both features belong to the same object, even when the objects are spatially superimposed (Duncan, 1984; Baylis & Driver, 1993; Kramer & Watson, 1995). A second source of evidence are studies showing that people find it difficult to ignore features that belong to an attended object (Kramer & Jacobson, 1991; Baylis & Driver, 1992; Yantis, 1992).

A fundamental question concerns the nature of the 'objects' for which this attentional benefit applies. In most experiments demonstrating object-based effects, grouping principles such as continuation (Moore, Yantis, & Vaughan, 1998), collinearity (Lavie & Driver, 1996), similarity (Kramer & Jacobson, 1991; Baylis & Driver, 1992) or common fate (Driver & Baylis, 1989; Behrmann, Zemel, & Mozer, 2000) are sufficient to define the objects in the display. Vecera and Farah (1997) have also shown that object-based attention is stronger for highly familiar shapes (upright letters) than for unfamiliar shapes that benefit from the same grouping principles (upside-down letters).

Current studies have not established whether these two means of 'characterizing' objects—based on generic grouping principles or long-term familiarity—are sufficient to predict when two features will be treated as belonging to the same object. A hypothesis that provides a more general characterization, which accounts for both of these other characterizations, is that perceptual experience with particular feature combinations determines whether or not two features will be integrated as an object of attention. We have previously developed a computational model, named MAGIC, based on this hypothesis (Mozer, Zemel, Behrmann, & Williams, 1992). MAGIC (an acronym for Multiple-object Adaptive Grouping of Image Components) was trained to group features from a set of images containing multiple objects in which each elementary feature was labeled as to which object it belonged. After training, MAGIC successfully segregated features of novel images into separate objects. Examination of the representations derived by MAGIC revealed that the critical aspects were the configurations of image features with a consistent labeling relative to one another. For example, the model discovered that for a T-junction, the features composing the base of the T should be consistently labeled as belonging to one object, and the features composing the top of the T should be labeled as belonging to a different object. In this way, MAGIC embodies the hypothesis that perceptual experience defines which features will be grouped together and which features will not. Under this hypothesis, generic grouping principles emerge based on compiled experience with a variety of feature and object combinations in images.

In this paper, we investigate the role of perceptual experience in object-based attention, examining questions such as whether short-term experience with a novel shape is sufficient to facilitate its being processed as a unitary whole; and the extent to which this experience with a shape may override other cues as to whether two features belong to a common object.

In order to explore these issues, we utilized a paradigm for studying object-based attention developed in previous work (Behrmann, Zemel, & Mozer, 1998). In these experiments, subjects decided whether the number of bumps appearing at two of four possible ends of two overlapping objects (or bars; see Figure 1) were the same (Figure 1a-c) or different (Figure 1d-f). The two features (sets of bumps) appeared on the ends of a single object (Figure 1a and 1d) or on the ends of two different objects (Figure 1b and 1e). Consistent with the *object-cost* hypothesis that it is difficult to attend to two objects simultaneously, subjects' responses were significantly slower to two features of two different objects than to two features of a single object (cf. Duncan, 1984).

These object costs—significant RT differences for responding to features of different objects versus features of a single object—provide an assay to determine when features are grouped into a single object. Instructions to the subjects carefully omitted any mention of objects, making this probe particularly useful because it does not involve any subjective definition of object-hood.

In our earlier experiments, we also included a third type of display in which the bumps were on the occluded object (Figure 1c and 1f) and, again, evaluated whether there was any cost relative to the single object trials. Occlusion is a particularly challenging condition for an object-based account of selection—not only are the features of a single occluded object spatially distant but they are also discontinuous (see Yantis, 1995; Moore, Yantis, & Vaughan, 1998 for other studies of occlusion effects on object-based attention). Interestingly, however, there was no object cost for the occluded trials, reflected in equivalent RTs for the single nonoccluded and the single occluded displays, both of which differedfrom the two object trials. These results held up both in the "X-displays" in which the bars crossed to form an "X", and also in "V-displays" in which the sets of bumps were all at 90 degrees from each other (see, e.g., Figure 1g-i). The evidence from these studies suggests that features of a single object, even if occluded, are grouped together and preferentially processed relative to features of other objects in the scene.

Our experiments established that features of an occluded object enjoy the same processing advantage as features of an unoccluded object relative to features of two different objects: removing explicit continuity as an object-defining cue did not affect object-based attention. An additional experiment also established a boundary condition of this result. When we changed the relation between the two discontinuous fragments of the occluded object so that they no longer formed a plausible single bar (Figure 2b), the object cost reappeared and performance was no longer equivalent to that of a single nonoccluded object.

These earlier experiments thus showed that manipulating the locations of image features could alter their processing, measured relative to processing features of a single versus two different objects. A primary aim of the experiments in this paper was to determine what defines the conditions under which attentional processes treat fragments as belonging to the same object or different objects. One issue concerned why the non-aligned fragments in Figure 2b are not treated as a single occluded object. One response to this question derives from theories proposing general-purpose processes by which image fragments are integrated into objects. For example, the results obtained with both the aligned and non-aligned fragments are consistent with a theory that the particular geometric relations between fragments determines whether they will form objects (Kellman & Shipley, 1991). Under this theory of *relatability*, spatially separated fragments are interpolated when their edges can be connected by a smooth monotonic curve; when the edges are no longer collinear, the fragments are not relatable and do not belong to a single occluded object.

A different hypothesis, consistent with MAGIC, is that these general-purpose mechanisms could emerge from perceptual experience. On this view, experience plays a determining role in perceptual organization.[1] This hypothesis is not incompatible with relatability theory, but in a sense it is

[1]Some people might call these general-purpose processes *bottom-up* and the experience-dependent processes *top-*

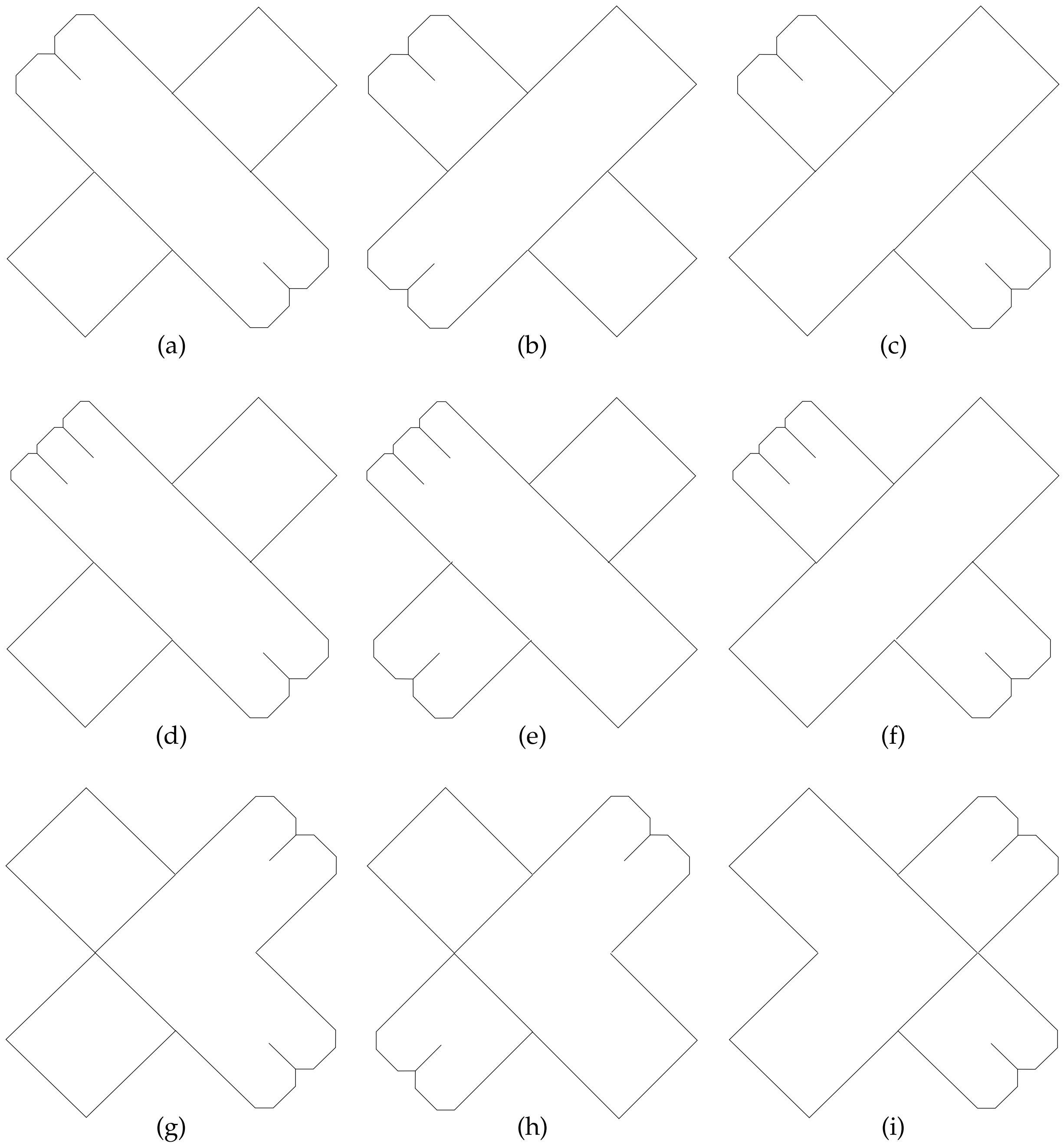


Figure 1: Examples of stimuli used in Behrmann et al. (1998) and here. Subjects had to make same/different judgements based on the number of bumps at two different locations in each figure. The top and third row represent 'same' judgements, while the middle row displays are 'different'. The left-most column depicts a single occluder condition, in which the bumps are on one, occluding object, the middle column shows the two object condition, and the right column shows the single occluded condition. The first two rows are examples of X displays, containing two overlapping bars, while the third row shows examples of V displays, containing overlapping V shapes.

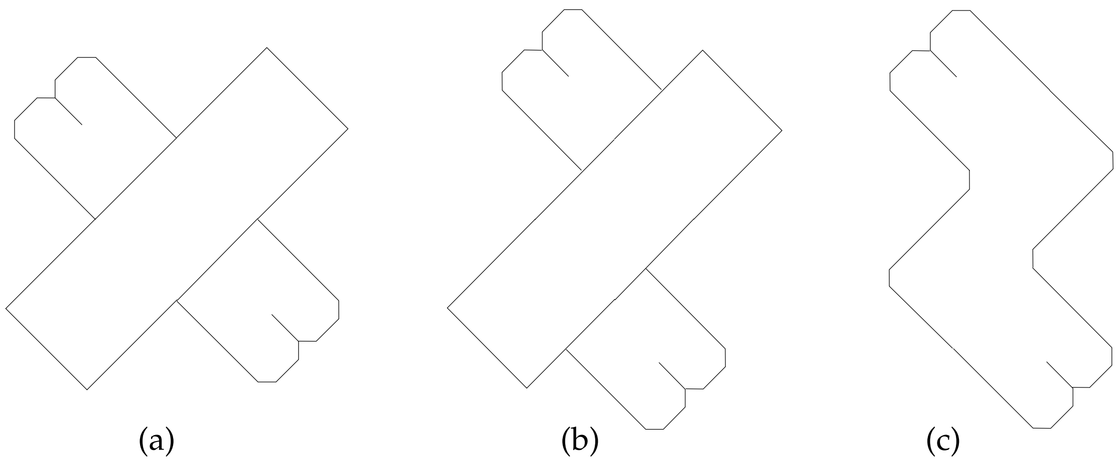


Figure 2: (a) Reaction times on displays where the bumps appear on the two fragments that correspond to the ends of an occluded bar are equivalent to those on unoccluded bars, and significantly shorter than when the bumps appear on two different bars. (b) When the fragments are shifted so they no longer form a plausible occluded bar, the RTs are equivalent to the two bar displays. (c) If perceptual experience plays a determining role in parsing, then subjects exposed to this shape may group the fragments in (b) into a single object.

more fundamental, as it suggests that experience can give rise to the grouping regularities that define relatability. A corollary of this view is that short-term perceptual experience may override the general-purpose, compiled grouping mechanisms. When short-term experience is consistent with longer term regularities then heuristics such as relatability will apply, but in other circumstances they will not. Consider the situation in which subjects are exposed to a shape that could potentially link together the two non-aligned fragments of the occluded object into a plausible object (see Figure 2c). Experience-dependent grouping would then predict that even if this novel shape is rather convoluted and irregularly shaped, the object advantage will apply even when only the non-aligned fragments are visible. We test this prediction in Experiment 1.

To summarize our direction, Experiment 1 demonstrates that short-term experience can affect grouping. The next logical issue concerns the circumstances under which this occurs. One relevant question is whether explicit evidence of occlusion is necessary for object-based attention to apply to the non-contiguous fragments. Is it essential in displays such as Figure 1c and 1f that the occluding bar be present, or can the same effect be produced without it, in a display containing only the fragments? Removing the occluder reduces the chances that *amodal completion*, in which a figure is perceived as complete even though it is not entirely visible because it is covered or occluded by something else (Kanizsa & Gerbino 1982), can be applied to the fragments. An important question is whether experience-dependent grouping is strong enough to operate in the absence of amodal completion, and if so, under what conditions. We explore this question in Experiments 2-4.

down. We prefer the former terms because we do not apply them to describe about the flow of information processing, but rather sources of knowledge. We discuss this issue further in the General Discussion.

Experiment 1: Amodal Completion Arising From Experience

The aim of this experiment was to determine whether exposure to specific displays can alter the perceptual organization of an image. This hypothesis predicts that the grouping of image fragments into objects based on general-purpose principles, such as relatability, uniform connectedness, and good continuation, can be superceded by experience.

In this experiment, subjects were split into two groups, Zee and Fragment. In the first block of trials, subjects in both groups saw displays such as those shown in Figure 3a,b, and they either had to respond to the fully visible object, or to the fragments that would typically not be grouped as belonging to a single object. In the second block of trials, Zee subjects saw displays containing a 'Z' object that linked up these fragments to form a single object (Figure 3c), while Fragment subjects saw the fragments alone without the fully visible object.

The key prediction concerned the third block of trials, in which both groups of subjects again saw displays like Figure 2a,b. The prediction was that because the Fragment subjects' experience will support an interpretation of this display as three separate objects (the two fragments and the central bar), so the fragments would have an object cost, just as they did in the initial block. On the other hand, the Zee subjects who saw the linking object wwould interpret the two fragments as belonging to a single occluded shape, as evidenced by their relatively speeded responses to those fragments.

Method

Participants. A total of thirty-two subjects participated in this experiment. The data of two subjects were excluded from the analysis because of high error rates ($> 10\%$). Subjects were drawn from the Carnegie Mellon University community and were paid $5 for their participation. Subjects ranged in age from 18 to 24 years. All had normal or corrected to normal visual acuity. None of the subjects was aware of the purpose of this study.

Apparatus and materials. This experiment was conducted on a Macintosh IIci computer. Stimuli were presented on a 14-inch color monitor (Basic color monitor: model M1595LL/A) using Psychlab experimental software version 1.0 (Bub & Gum, 1991). The displays were presented as black-and-white line drawings on a white background. Viewing distance was approximately 50 cm. The same viewing distance was used in all the experiments presented here.

There were four types of displays (see Figure 3):

1. (Figure 3a,b) **Ambiguous**. This display could either be interpreted as a rectangular bar occluding a Z-shaped object, or as a rectangular bar with two smaller rectangular ends butted up against it. The rectangular bar was 8.7 cm in length (10.2°) and 2.5 cm in width (2.9°). The two ends were created by taking the two visible fragments of an orthogonal occluded bar of the same dimensions and displacing them by slightly more than the width of the rectangle (3.3°). The lines defining the bumps were 1.25 cm long. This display appeared equally often in four different orientations, as shown in Figure 4. Furthermore, the displays fell into two conditions, based on the locations of the features (bumps):

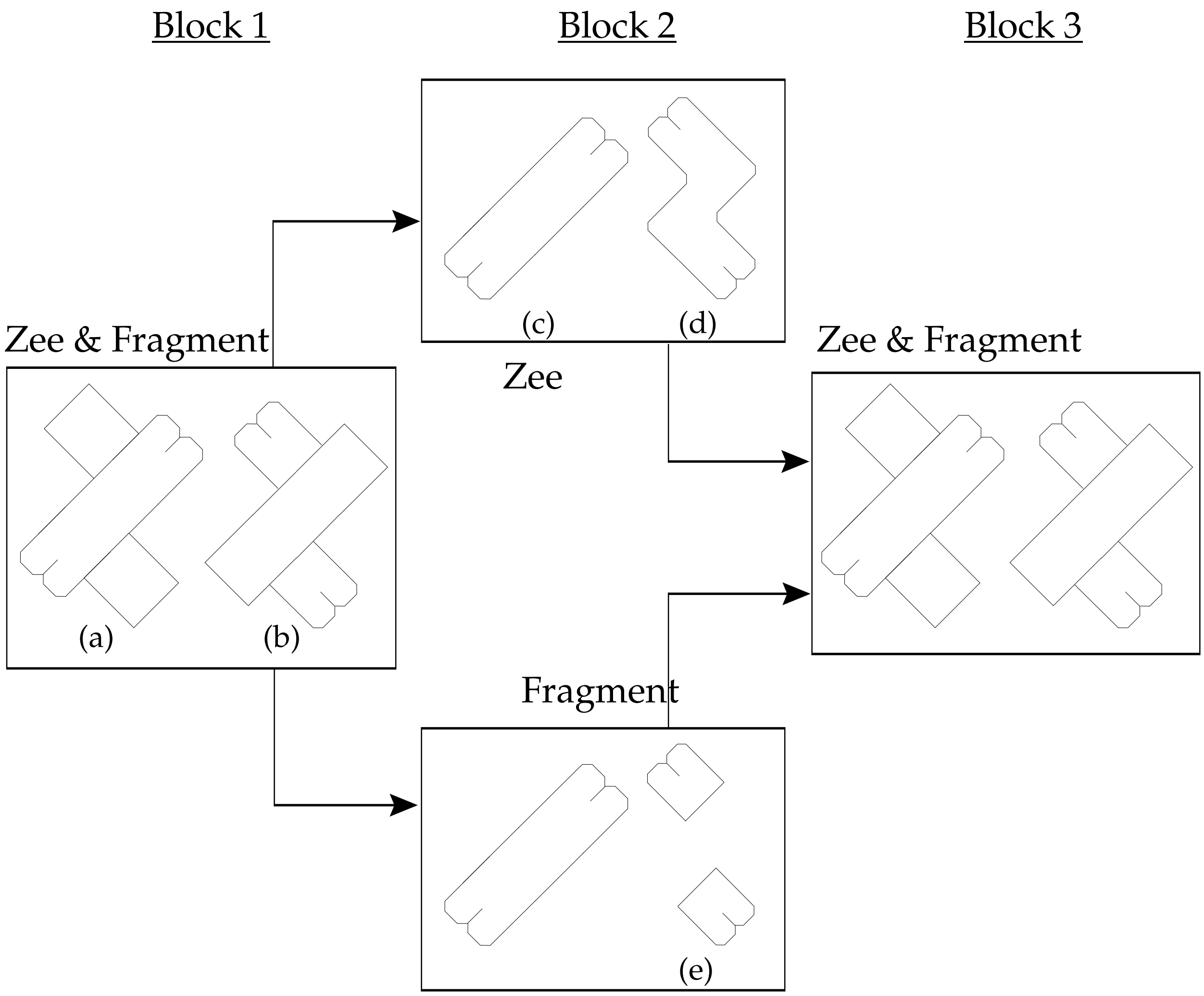


Figure 3: Design of Experiment 1. Both groups of subjects (Zee and Fragment) performed 3 blocks of trials. For both groups, block 1 trials contained Ambiguous displays in which the bumps were either on the (a) Bar or (b) Fragments. Half the trials in block 2 contained displays of a single (c) Bar. For the Zee group, the other trials in block contained the (d) Z shape, while Fragment group subjects saw the (e) Fragments display. Block 3 contained the same stimulus set as block 1 for both subject groups.

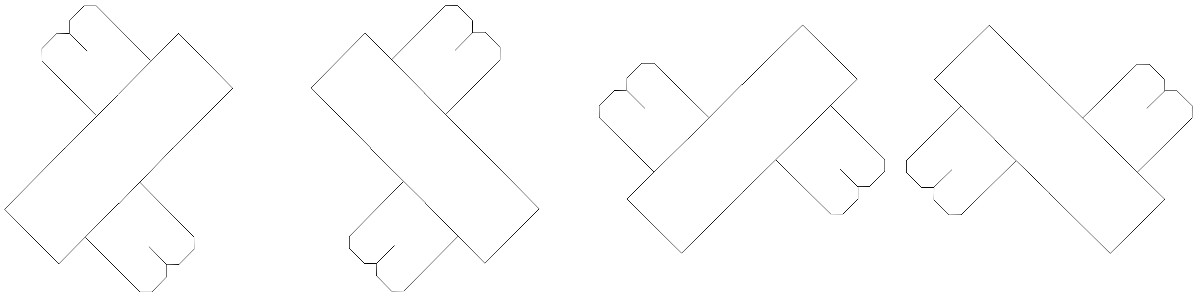


Figure 4: Examples of the four orientations of Ambiguous displays used in Experiment 1.

   (a) *Connected-Bumps*: The bumps appeared on the opposite end of the single coherent bar.
   (b) *Disconnected-Bumps*: The bumps appeared on the two fragments.

2. (Figure 3c) **Bar**. This display was created by removing all but the bar from the Ambiguous displays. The Bar appeared in two different orientations, and only the Connected-Bumps condition was relevant to this display.

3. (Figure 3d) **Z**. This display was created by removing the bar from the Ambiguous displays and replacing it with contours connecting the two remaining fragments. The contours were slightly narrower than the Bar: 2.0 cm wide as opposed to 2.5 cm. This display also appeared equally often in the four different orientations equivalent to those shown in Figure 4. Only the Connected-Bumps condition was possible in this display.

4. (Figure 3e) **Fragments**. This display was created by removing the bar from the Ambiguous displays and simply adding a single line to each of the two remaining fragments to form separate rectangular ends. The Fragments appeared in four orientations, and only the Disconnected-Bumps was relevant to this display.

In all trials of this experiment, and in every experiment described below, subjects performed the same/different number-of-bumps decision. On each trial, bumps appeared at the extremities of either the Fragments or the Bar. Each set of bumps was either in a *two-bump* or *three-bump* configuration: the end was divided into two equal parts for the two-bump and into three equal parts for the three-bump displays. There was an equal number of 'same' and 'different' judgments in each of the two conditions. On 'same' trials, there were either two (known as a 2-2 trial) or three bumps (known as a 3-3 trial) and there were an equal number of 2-2 and 3-3 'same' trials. On 'different' trials, there were always two bumps at one location and three bumps on the other and the locations of the 2 and 3 bumps were evenly counterbalanced.

The subject's task was simply to decide whether the number of bumps at the two locations was the same or different. Responses were indicated with the [Z] or [/] keys with the left and right index fingers on the standard QWERTY keyboard. The assignment of keys to 'same' or 'different' responses was counterbalanced across subjects.

Design. Subjects performed 3 experimental blocks of trials. Blocks 1 and 3 were equivalent for the two groups; the difference occurred in block 2. In blocks 1 and 3, all subjects made same/different judgements to Ambiguous displays. Feature location (Connected-Bumps vs. Disconnected-Bumps) was crossed with bump-number combinations (2-2, 2-3, 3-2, and 3-3) and orientation (the 4 shown in Figure 4), yielding a base set of 32 trials which was replicated 8 times for a total of 256 trials in block 1. A practice block consisting of 16 trials—a randomly-selected half of this base set—was completed before block 1 to accustom subjects to the display and response keys. These data were not analyzed.

In block 2, the subjects' experience was manipulated. Both groups saw Bar displays on half of the trials in block 2. On the other trials, Zee group subjects saw Z displays while Fragment group subjects saw Fragments displays. Block 2 also involved 256 trials, consisting of 8 replications of the basic crossing of trial type (Bar or Z for Zee group, Bar or Fragments for Fragment group), bump-number combination and orientation. A break of a few minutes was given between blocks.

The experiment contained one between-subjects variable (Zee or Fragment group) and two within-subjects variables: block (1 and 3); and feature location (Disconnected-Bumps and Connected-Bumps). We expected that both groups would improve in their overall reaction times in block 3 compared with block 1 due to a general practice effect. More importantly, we predicted that if learning is mediated by exposure to a linking object, we would find a 3-way interaction: the Zee group would process the Disconnected-Bumps of the Ambiguous displays more quickly than the Fragment group, but only in block 3, once they have been exposed to the disambiguating display, and not in block 1. This prediction can be re-expressed in terms of object costs, i.e., RT differences for features of different objects versus features of a single object. The prediction is that the obect costs would be observed in block 1 for the Ambiguous Disconnected-Bumps trials for both groups; and they would disappear in block 3 for the Zee group, but not for the Fragment group. Thus the critical predictions here are: (1) a block by feature location effect for the Zee group, but not the Fragment group; and (2) a block by feature location effect for the Zee group, but not the Fragment group.

Procedure. Each trial proceeded as follows: A fixation point appeared for 1 second (sec) followed by a 500 millisecond (msec) delay. Thereafter, the display appeared and remained on the screen until a response was made. An inter-trial interval of 1 sec occurred following the response and prior to the next trial. The same procedure was followed in all experiments presented in this paper.

Treatment of results. The data from the practice trials were discarded from the analysis. Error trials were excluded from the reaction time (RT) analysis. First an analysis was conducted, crossing group, feature location, block, judgement, and orientation. The mean RT and errors for each crossing of group, feature location, block, judgement, and orientation were calculated for each subject and were then subject to analyses of variance. Reaction times that exceeded two standard deviations above or below a subject's mean were also excluded from the analysis. Further analyses were then conducted on these data, as discussed below.

Results and Discussion

The overall error rate for this experiment was low, comprising 2.2% of the total number of trials. A further 3.5% of the data was excluded as exceeding the two standard deviations cutoff.

The analysis of variance conducted on the error data did not reveal significant effects of any of the factors nor any interactions so the error data were not subject to further analysis.

The analysis of variance conducted on the reaction time data revealed no difference as a function

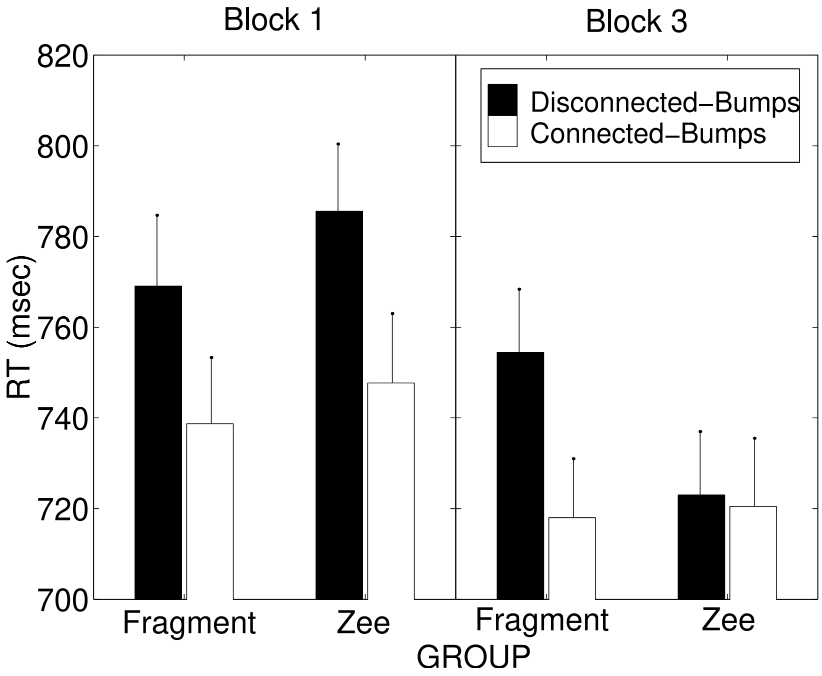


Figure 5: Mean reaction times and standard error bars in blocks 1 and 3 for the two subject groups as a function of feature location (on the Bar or on the two Fragments) for the Ambiguous displays in Experiment 1.

of orientation ($F(1,28) = 2.04, p = .14$). With respect to judgement, as is typically the case in the same-different paradigm (Nickerson, 1965), same judgements were found to be significantly faster than different judgements ($F(1,28) = 5.51, p < .05$). Importantly, no significant interaction was found between these two variables, judgement and orientation, and no interaction between either of them and the other factors. Consequently, the RT data were pooled across judgement and orientation for subsequent analyses.

The RT data contained a significant main effect for block ($F(2,56) = 17.55, p < .001$), revealing the overall effect of practice as the experiment progresses. There was also a significant main effect of feature location ($F(1,28) = 20.82, p < .001$), with a difference of 24 msec between the RTs for the Connected-Bumps and Disconnected-Bumps conditions. This effect corresponds to the overall object cost. The RT data contained no effect for group ($F(1,28) = .01, p = .93$). No pair-wise interactions were found between the variables. The critical finding with respect to our hypothesis was the sole interaction between the factors: a significant three-way interaction between block, group, and feature location ($F(2,56) = 3.32, p = .04$). This three-way interaction indicates that the object cost varies between the blocks as a function of group.

Given this three-way interaction, we then conducted separate analyses within each block of the experiment, crossing one between-subjects variable (Zee or Fragment group) and one within-subjects variable (Disconnected-Bumps or Connected-Bumps feature location). Note that the second block was included for the sake of completeness. The critical data concerns blocks 1 and 3, since the stimulus sets were the same across the groups in these blocks. Figure 5 shows the mean RTs for these variables (group and feature location) in the two critical blocks.

The block 1 RT data contained a significant effect of feature location ($F(1, 28) = 17.81, p < .001$; difference of 34 msec), but no effect of group ($F(1, 28) = .43, p = .515$) and no interaction between the two variables. This analysis reveals a clear object cost, suggesting that the bumps on the fragments are treated as being on separate objects.

In block 2, the RT data again contained a significant effect of feature location ($F(1, 28) = 7.47, p = .01$; 19 msec difference), but no effect of group ($F(1, 28) = .06, p = .8$). There was an interaction between the two variables ($F(1, 28) = 4.79, p = .04$), which reflects the different stimuli seen by the two groups in one of the feature location conditions. In the Connected-Bumps condition, which in this case involved a unitary bar, the reaction times of the two groups were similar (710 ms for the Zee group; 700 ms for the Fragment subjects). The Zee group responded similarly on the Disconnected-Bumps (713 ms), which in this case contained the unitary Z-shaped display. The Fragment group was considerably slower on the Disconnected-Bumps (731 ms), revealing an object cost for the Fragment stimuli.

The RT data from block 3 also contains a significant effect of feature location ($F(1, 28) = 7.7, p < .01$; 19 msec difference), but no effect of group ($F(1, 28) = .63, p = .43$). Most importantly, a significant interaction again exists between these variables in block 3 ($F(1, 28) = 5.85, p = .02$), with the Fragment group showing significantly shorter RTs to Connected-Bumps than Disconnected-Bumps and the Zee group exhibiting equivalently fast RTs to the two stimulus types. Note that in block 3, there is no difference between the two groups on the Connected-Bumps (2 msec) and the effect arises solely from the Disconnected-Bumps (32 msec difference).

Finally, to examine the breakdown of the data within the two groups, we conducted ANOVAs for each of the two groups separately, using block and feature location as within-subject variables. We only included RT data from blocks 1 and 3 to focus the analysis on trials involving physically identical display sets. For the Fragment group, a clear object effect is evidenced by the significant effect of feature location. These subjects were consistently faster when the bumps appeared on the single object (Connected-Bumps) than on two separate objects (Disconnected-Bumps): $F(1, 14) = 25.99, p < .001$. There was also a general speed-up between blocks 1 and 3 (15 ms for Disconnected-Bumps, 21 ms for Connected-Bumps), but this difference did not reach significance ($F(1, 14) = 3.75, p = .07$). Critically, no interaction existed between these two variables ($F(1, 14) = .15, p = .71$). This implies that the object cost (poorer performance on Disconnected-Bumps than Connected-Bumps) in blocks 1 and 3 were equivalent: 30 ms for block 1, 36 ms for block 3.

For the Zee group, there was a much larger drop in RTs from block 1 to 3 for Disconnected-Bumps than Connected-Bumps (63 ms vs. 27 ms). In this group, the 2-way interaction between feature location and block is significant: $F(1, 14) = 6.7, p = .02$. Considered individually, feature location was not significant ($F(1, 14) = 4.1, p = .06$); block, however, was highly significant ($F(1, 14) = 17.2, p < .001$). The trend towards an effect of feature location comes about because of the difference in block 1. The other effects (the significant 2-way interaction and effect of block) can be traced primarily to the same cause: the speed-up for the Fragment-Bump condition between blocks 1 and 3. This speed-up wiped out the object cost (effect of feature-location) in block 3 that was present in block 1 (38 ms vs 2 ms).

Two main conclusions may be drawn from these results. First, the results replicate the finding in Behrmann et al. (1998) that displaced fragments are not treated as an occluded object in the Ambiguous displays. This is manifested in the significant effect of feature location in block 1, where subjects are faster on Connected-Bumps than Disconnected-Bumps, suggesting that the fragments are not being perceived as a single object. This finding is consistent with the principle of relatability (Kellman & Shipley, 1992), which predicts that the contours of the two fragments will not be interpolated because of the misaligned geometric relationship between them.

The second conclusion is the more interesting one: exposure to a novel object that biases the interpretation of the display changed the processing of the ambiguous visual input. The object costs in blocks 1 to 3 in the Fragment group subjects are almost identical, indicating that viewing the displays in block 2 (Bar and Fragments) had a similar effect on the processing of the feature-location conditions. For the Zee group, however, viewing the block 2 displays (Bar and Z) had a differential effect on the Disconnected-Bumps and Connected-Bumps conditions. For this group of subjects, the RT means are almost identical for these two feature locations in block 3 (723 vs. 721 msec), indicating an object-based effect in the Ambiguous displays that is as strong for the Fragments as for the fully visible Bar.

Experiment 2: V $\rightarrow$ Ends Transfer

The results of Experiment 1 demonstrate that exposure to a novel shape that links together feature fragments affects the later processing of displays in which fragments may be interpreted as part of a single object. This suggests that subjects perform a form of amodal completion given knowledge of an object that can link the fragments in a display. Note that in these stimuli, the occluding bar was present along with the fragments in the ambiguous display, and subjects then presumably interpolated the presence of the Z object below the occluding central bar. This result is particularly interesting given that the fragments have a terminating edge suggesting that they are closed and likely to be objects unto themselves. Despite this, subjects still come to treat the fragments as part of a single object once they had been exposed to the Z object. A natural next question concerns the necessity of occlusion: Does this completion require the presence of an occluding object?

Consider for example Bregman's well-known B displays, where one recognizes a smattering of edge fragments as a set of block-letter Bs once the occluding blobs are added to the image (see Figure 6). The question is whether conditions exist under which the fragments of the Bs may be sufficient to allow for completion without the presence of the occluding blobs. In the context of our experience-dependent grouping hypothesis, this question concerns the strength of the experience effects: Could repeated experience with B shapes influence perceptual organization such that the effect of the object can be detected in an attention task, even without the occluding blobs? Experiment 2 was designed to address this question.

To examine these issues, we devised a series of experiments using the same methodology as Experiment 1, and similar stimuli to our earlier experiments. As in Experiment 1, subjects first saw a block of trials with an ambiguous display, then a block using a full display that favored a particular interpretation of the ambiguous display, followed by another block of the ambiguous display. In Experiment 2, the stimuli were derived from the V displays (Figure 1g-i), for which we found an object-based attention effect (Behrmann et al., 1998). The central prediction in the new experiment was that exposure to V displays would affect the processing of displays in which there was no occlusion information in the image, and the fragments could be consistent with many different shape configurations. These *Ends* displays (see Figure 7) were similar to the Fragments displays

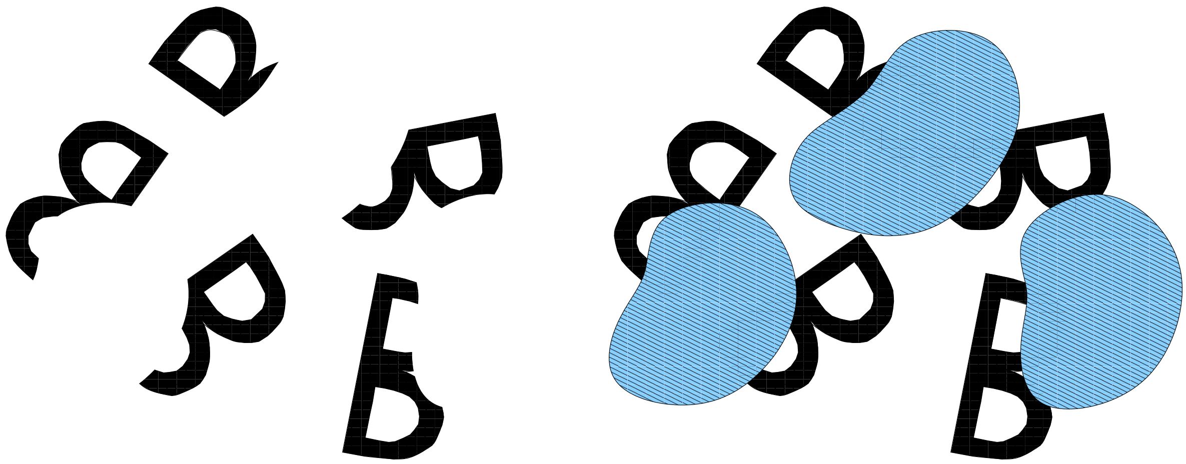


Figure 6: Evidence of occlusion facilitates completion of the occluded shapes on the right, while they are more difficult to perceive on the left (after Bregman, 1981).

in Experiment 1, except that the fragments here corresponded to the ends of two overlapping V shapes (the V display). We used the Ends displays to examine the influence of viewing V displays on subjects' performance.

Method

Participants. Eighteen subjects, nine male and nine female, between 18 and 23 years of age were recruited from the undergraduate subject pool at the University of Arizona. All subjects had normal or corrected visual acuity by self report, and were unaware of the purpose of the experiment.

Apparatus and materials. The apparatus was the same as in Experiment 1, except that we used a 13-inch color monitor.

The Ends displays contained four rectangles of 2.5 cm x 1.5 cm, oriented at 45 degrees (see Figure 7). The Ends were made by removing the center of a display which contained two V's lying atop one another (see Figure 1g-i). The diagonal extent of this display matched the dimensions of the bar in the displays of Experiment 1 (8.7 cm long by 2.5 cm wide). The horizontal line drawn from the midpoint of one rectangular End to the midpoint of the horizontally-aligned other End was 6.5 cm. On each trial, the features (bumps) appeared on two of the four Ends, in either a two- or three-bump configuration. The bump configurations were the same as in the previous experiment.

The displays used in this experiment fell into two conditions, based on feature location:

1. Diagonal-Bumps (Figure 7a-b): the bumps appeared on the diagonally opposite ends. For the V displays, this configuration corresponds to the bumps lying on two different objects.
2. Vertical-Bumps (Figure 7c-d): the bumps appeared either on the end pairs on the right or left hand side of the display. Here the object relationship is reversed. For the V displays, this configuration corresponds to a single object.

There was an equal number of 'same' and 'different' judgments in each of the two conditions, as in the previous experiment, and the locations of the bumps were evenly counterbalanced (diagonal-left or right for Diagonal-Bumps, vertical-left or right for Vertical-Bumps). The V displays had one other degree of freedom, orientation: whether the left or right-facing V was on top. This variable was also counter-balanced. The total number of displays for the Ends was 16; this number was doubled to equal the number of V displays.

As in Experiment 1, the subject's task was simply to decide whether the number of bumps on the two ends was the same or different. Responses were indicated with the [Z] or [/] keys with the left and right index fingers on the standard QWERTY keyboard. The assignment of keys to 'same' or 'different' responses was counterbalanced across subjects. Reaction times (RT) to make the decision was recorded in milliseconds and accuracy was noted.

Design. The experiment was run in 3 blocks. In the first block, the subjects saw only the Ends. In the next block, the subjects saw only V displays. These two blocks constituted the *Initial* epoch for these two displays respectively. In the final block (the *Test* epoch), Ends and V displays were randomly intermixed. Note that subjects did not see any examples of the V displays before block 2. The design was entirely within-subject, with the relevant independent variables being feature location (Diagonal-Bumps, Vertical-Bumps), display (Vs or Ends), and epoch (Initial, Test). Judgement (same, different) was another independent variable, as was orientation for the V displays. The design is summarized in Figure 7.

The three experimental blocks each consisted of 192 trials, with a few minutes break between each block. Trials were randomized within a block. Prior to starting the experiment, subjects were given 32 practice trials, two of each of the Ends trials. Timing and response measurements were the same as in the previous experiment.

Results and discussion

The errors constituted a small proportion (2.3%) of the trials. As in Experiment 1, trials exceeding the two standard deviation cutoff were removed, resulting in the removal of an additional 2.8% of the trials. As in Experiment 1, an analysis of variance with error rates as the dependent measure and epoch (Initial, Test); feature location (Diagonal-Bumps, Vertical-Bumps); display (Ends, V); and judgement (Same; Different) was conducted. Initial epoch refers to block 1 for Ends and block 2 for Vs, and Test refers to block 3 in which Ends and Vs are shown. This ANOVA on error rates revealed no significant factors nor interactions.

We conducted a similar analysis of variance with correct RTs as the dependent measure, crossing epoch, feature location, display, and judgement. Same judgements were again significantly faster than Different ($F(1,17) = 4.7, p < .05$), but no significant interactions were found between this and the other variables. A separate analyses of variance conducted crossing orientation with epoch and feature location on the V displays (since the Ends displays have only one orientation) revealed no significant interactions (all $F < 1$). Thus we pooled the data across orientation and judgement for subsequent analyses. The mean RT data for the three primary factors are shown in Figure 8.

The ANOVA revealed that feature location ($F(1,17) = 10.66, p < .01$; RT difference of 27 msec) and epoch ($F(1,17) = 7.76, p = .01$; 18 msec) were significant, but display type was not ($F(1,17) =$

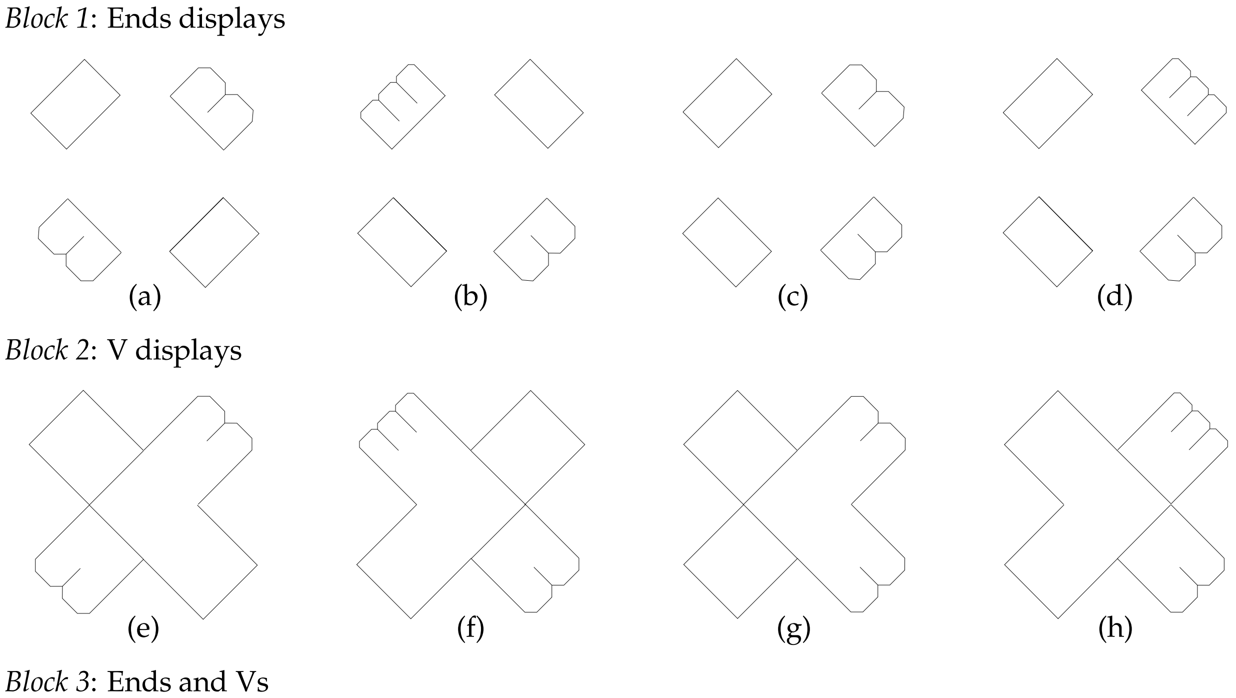


Figure 7: Design and stimuli used in Experiment 2. In block 1, subjects performed same/different judgements to Ends displays, examples of which are shown here (a-d). These same Ends displays were also used in Experiment 3. In block 2, subjects performed the same task on V displays (e-h). Displays a, b, e and f are examples of Diagonal-Bumps displays, while c, d, g, and h are Vertical-Bumps; a, c, e, and g are same judgements, b, d, f, and h are different. In block 3, both types of displays, Ends and Vs, were then mixed.

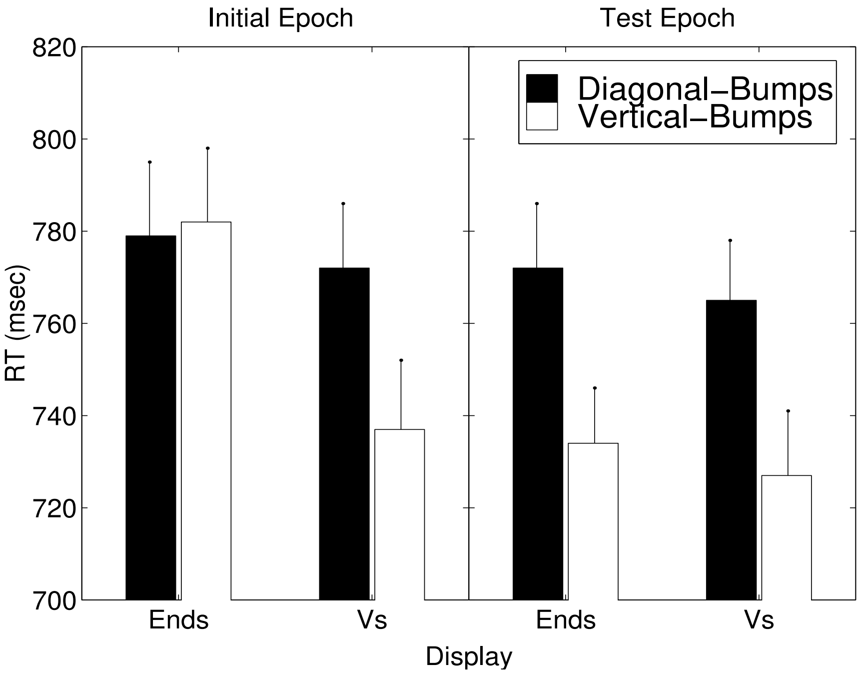


Figure 8: Mean RTs as a function of display (Ends, Vs), epoch (Initial, Test), and feature location (Diagonal-Bumps, Vertical-Bumps) in Experiment 2. The first panel of the display shows the mean RTs to Ends and Vs during blocks 1 and 2 respectively. The second panel shows the mean RTs to these same displays in block 3, which consisted of both display types. Note that Diagonal corresponds to the bumps being on separate objects, while Vertical corresponds to them being on a single object.

$.45, p = .5$). No pairwise interactions were significant, but, importantly, the three-way interaction between the variables was ($F(1, 17) = 6.03, p = .02$). Consequently, we conducted separate ANOVAs within each block of the experiment.

The first result from the study was that, for Ends displays in the Initial epoch, RTs for Diagonal-Bumps and Vertical-Bumps were no different: $F(1, 17) = .07, p = .8$. This indicates that for the Ends displays, subjects did not treat the ends separated vertically differently from those separated diagonally. Given that these are discontinuous, spatially separated features, the virtually equivalent mean RTs (difference of 4 msec) is not surprising. Secondly, for V displays in the Initial epoch, RTs for the Vertical (single-object) condition were significantly shorter than the Diagonal (two-object) condition: $F(1, 17) = 7.6, p = .01$. The RT difference was 35 ms. The analysis of the Test epoch revealed a highly significant effect of feature location ($F(1, 17) = 16.36, p = .001$), no effect of display ($F(1, 17) = 0.06, p = .8$), and no interaction between them ($F(1, 17) = 0.01, p = .9$). Thus the object effect was found for both displays in this epoch. For the Vs, the object cost was approximately equivalent in the two epochs: a difference of 37 ms in mean RTs in the Test epoch versus 35 ms in the Initial one.

The epoch analysis results revealed a replication of the basic object effect for the Vs found in Behrmann et al. (1998). To determine the extent to which the data really do replicate the results of Behrmann et al., (1998), in which the RTs were equivalent for features when they fell on the same, occluding object as when they fell on the same, occluded object, we separated out the data in the single-object condition. In the context of our design, this refers to the Vertical-Bump stimuli. An analysis of the effect of occlusion—whether the V with the bumps was complete or occluded—revealed no significant effect ($F(1, 17) = .43, p = .52$). This shows that our replication of the earlier results extends to the occluder/occluded distinction.

The critical result with respect to our hypothesis was the object effect for the Ends displays in the Test epoch. Whereas feature location was not significant for the Ends in the Initial epoch—a 4 ms difference in the wrong direction—mean reaction times for same-object features were 38 ms shorter than different-object in the Test epoch ($F(1, 17) = 7.46, p = .01$).

This last result, which corresponds to the significant three-way interaction between display, feature location, and epoch, clearly shows that exposure to the V displays in block 2 induced an object effect in the Ends displays in the following block, *within a single group of subjects within one testing session*. The equivalence of the diagonal and vertical feature locations for these subjects in block 1 was undone by their perceptual experience during block 2. Thus, even though the Ends are ambiguous in and of themselves, they were treated as equivalent to the single-object and two-object conditions of the V displays by virtue of subjects' experience with the V displays.

This finding is analogous to the biasing effect that the Z displays had on the Zee group subjects in Experiment 1. In this case, however, the effect of the novel object on the grouping of fragments was observed even without the presence of an occluding shape. This result is surprising, because experience with the V displays induced ends to be grouped together even though all perceptual information in the test display indicates that they are *not* related, i.e., the four Ends are each closed rectangles.

One possible account for this finding is that the short-term effect of experience with the object is so strong that it can over-ride evidence that the Ends are closed self-contained objects, and produce a percept of the linking V shape. This account would suggest that the Bs in Figure 6a would be perceived following some exposure to the B shapes. An alternative explanation for the results of this experiment is that the object attention process is imprecise, and can be tricked into linking the Ends even when the image evidence is not consistent with the interpretation of the two fragments as belonging to a single object. Whereas the first account is driven by a whole-object matching process, this second account relies on a limited analysis of local features. We return to this issue in the General Discussion.

In any case, the finding provides strong evidence that subjects' perceptual experience alters subsequent visual processing: the effect of experience is sufficiently robust so that even though information in the image might be interpreted to the contrary, that the bumps are not part of a larger object by virtue of the closing bar, they are still grouped together.

Experiment 3: X $\rightarrow$ Ends Transfer

The previous experiment confirmed the hypothesis that image fragments may become treated or interpreted in a particular way depending on the specific experience of the subject. Exposing a subject to images that link particular fragments into objects affects their subsequent organization of these fragments, both in the presence of and absence of occlusion. In the occlusion-present Ambiguous displays of Experiment 1, the grouping of features on the fragments was neutral, potentially interpretable as belonging to a single occluded object or two separate objects. Following experience with the Z displays, subjects grouped the Disconnected-Bumps together. In the occlusion-absent Ends displays of Experiment 2, prior to any exposure to the V displays, subjects did not show any RT differences to Vertical-Bumps or Diagonal-Bumps. Here however, after experience with the V displays, cues indicating that the ends were unrelated were overcome, and the vertically aligned features of the Ends display were responded to faster than the features that were aligned diagonally.

One important issue concerns whether subjects are somehow predisposed to grouping vertically aligned features, and the slightest experience with shapes in which features group vertically is sufficient to induce the vertical bias in the neutral Ends displays. A stronger demonstration of the effect of experience would utilize a different display, in which the features are grouped diagonally instead of vertically, and show that this reverses the grouping of features in the same neutral Ends displays. We addressed this issue in Experiment 3 by using the X displays as the disambiguating displays instead of the V displays. This leads to the directly opposite prediction: vertically aligned bumps now belong to two different objects while diagonally aligned bumps belong to the same object, so subjects should now show an object advantage for Diagonal-Bumps on the ends as opposed to the advantage obtained for Vertical-Bump ends in Experiment 2.

Given that we know from Experiment 2 that initially there is no difference between Vertical-Bumps and Diagonal-Bumps in the Ends displays, we started this experiment by exposing subjects to the Xs directly, without probing the Ends alone first.

Method

Participants. Twenty-four subjects, half male and half female, between 18 and 25 years of age were recruited from the undergraduate subject pool at the University of Toronto. All subjects were right-handed and had normal or corrected visual acuity by self report.

Apparatus and materials. The same apparatus as used in Experiment 1 was used here. The displays included the Ends used in Experiment 2 as well as the full X displays, shown in Figure 1a-f. The dimensions of the X displays were identical to those of the Vs and Ends display as these displays were constructed from the full X displays. As in the previous experiment, the displays fall into two feature conditions, depending on whether the two sets of bumps fall on the diagonal (Diagonal-Bumps) or on the vertical left or right (Vertical-Bumps). Note that here as opposed to the previous experiment, Diagonal-Bumps corresponds to a single object (in the full X display), while Vertical-Bumps corresponds to two different objects. The task and response measures were the same as in the previous experiments.

Procedure. The experiment was run in 4 blocks with a few minutes break between each block. The first two blocks contained X displays, and the next two contained Ends displays. Each block consisted of 128 trials: four replications of the full set of X displays, or eight replications of the Ends displays. Trials were randomized within a block. At the beginning of the experiment, subjects were shown printouts containing examples of the X displays and were instructed to make same/different judgements on the number of bumps. Each trial proceeded as in the previous two experiments. Prior to starting the experiment, subjects were given 32 practice trials, one of each of the X displays.

Design. The design was entirely within-subject, with the statistically independent variables being display type (Xs, Ends), feature location (Diagonal-Bumps, Vertical-Bumps), judgement (same, different). Orientation was another independent variable for the X displays.

Results and discussion

The errors constituted a small proportion (2.5%) of the trials. As in the two previous experiments, trials exceeding the two standard deviation cutoff were removed, resulting in the removal of an additional 1.9% of the trials. An analysis of variance on the error rates crossing display type (Ends, X) and feature location (Diagonal, Vertical) found no significant effects nor interactions. Also, separate analyses of variance on reaction time data conducted crossing judgement (and orientation in the case of the Xs) with the main factors—display type and feature location—revealed no significant interactions (all $F < 1$), so the data were pooled across orientations and judgements for subsequent analyses.

An analysis of variance with mean correct RTs as the dependent measure and display type and feature location was conducted. The RT data are illustrated in Figure 9.

Because there was no initial epoch in this experiment, the important comparison was between the Ends and Xs after the subjects had been exposed to the Xs. The primary result from this study was the highly significant effect of feature location, $F(1, 23) = 19.34, p < .001$. This finding, together with the lack of any significant interaction between feature location and display type, $F(1, 23) = .233, p = .64$, indicates that the object effect held identically for both the X displays and the Ends displays: the Diagonal-Bumps (lying on a single bar) were processed more quickly than the Vertical-Bumps (lying on separate bars).

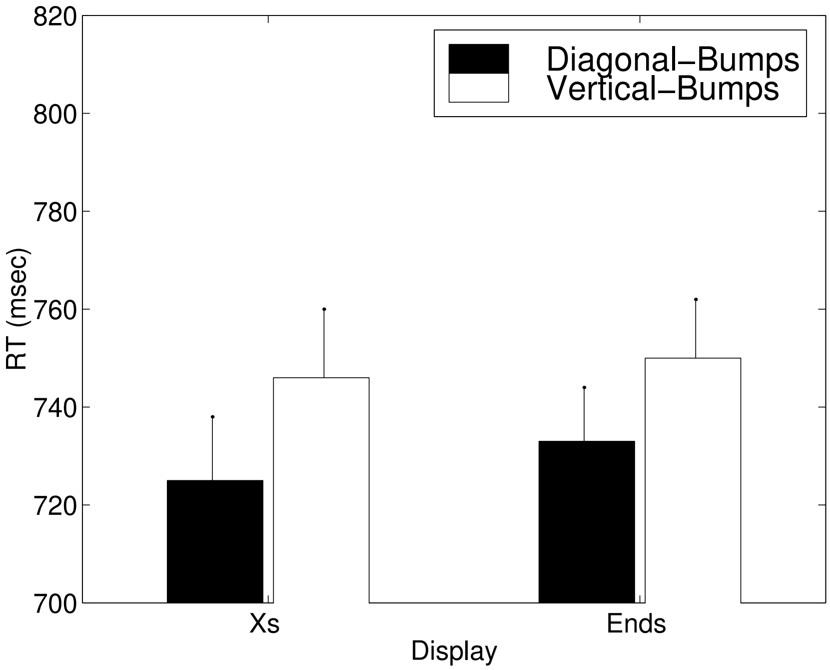


Figure 9: Mean reaction times as a function of display (Xs or Ends) and feature location (Diagonal or Vertical feature locations) for Experiment 3. Subjects in Experiment 3 performed the same/different number-of-bumps task on two blocks of the X displays followed by two blocks of Ends displays. Note that here the Diagonal-Bumps lie on a single object, while the Vertical-Bumps lie on different objects. In Experiment 2, for subjects who had not yet seen either X or V displays, mean RTs on the Ends displays were 778 and 782 msec for Diagonal-Bumps and Vertical-Bumps, respectively. Thus the object effect apparent here for the Ends display emerges after exposure to the Xs.

To summarize, the results of this experiment were straightforward. The difference between the single object and two object condition in the X display was replicated. In addition, this difference also applied to the Ends displays: after being exposed to X displays, subjects responded relatively quickly to the Ends displays which corresponded to the single object in the full X displays and less quickly to the Ends displays which corresponded to the two object condition in the full X displays. We know from Experiment 2 that naive subjects who did not have experience with full X displays did not treat the Diagonal or Vertical bumped Ends differentially. Thus, even though the Ends conditions were ambiguous in and of themselves, they came to be treated as equivalent to the single and two conditions of the X displays by virtue of subjects' experience with these X displays.

Taken together, Experiments 2 and 3 provide complementary results illustrating the ability of perceptual experience with an object to affect a feature judgement task where the displays contain only fragments. In both cases, the fragments were closed shapes, so the images not only lacked information about occlusion but contained contradictory evidence against an occlusion interpretation, yet the results demonstrate that they were treated as parts of an object due to experience.

Experiment 4: Location Specificity of Perceptual Experience

The experiments presented above show that exposure to displays containing an object can induce an object effect in ambiguous displays that in and of themselves do not contain any evidence of an occlusion relationship. Subjects exposed to X displays respond faster to diagonal ends than to vertical ones in the Ends displays, corresponding to the object effect in the X displays. In contrast, subjects exposed to the V displays, responded to the vertically aligned ends faster than the diagonal ends. The Ends display, then, is initially ambiguous but is later parsed according to the subject's recent perceptual experience.

An important question then concerns the nature of the representations that are activated as a function of experience with a particular display. A number of possibilities exist. We can characterize them based on the relevant reference frame for the active representations:

1. **Environmental**. One possibility is that particular pairs of locations *on the screen* obtain a processing advantage due to experience with a specific shape. In our experiments, the stimuli were of a constant size, and were displayed in a consistent location on the screen. Specific objects then could induce groupings of screen locations where features of that object appear. Clearly, at least pairs of positions must be primed by the object, because the X and V displays involved the same set of individual positions, and only differed with respect to which pairs of positions belonged to the same objects. On this view, the Ends displays do not need to activate any object representations in order to produce the object effects observed in Experiments 2 and 3.

2. **Viewer-based**. Another possibility is that the perceptual organization process is mediated by shape representations, but these representations are highly viewpoint specific, i.e., tied to particular locations on the screen. This view is consistent with recent theories about the relationship between object-based and spatial attention (e.g, Goldsmith, 1998; Mozer et al., 1992; Vecera, 1994; see General Discussion), in that the objects activate particular spatial locations, and attention is then allocated to those spatial positions. On this view, exposure to the V (or

X) displays primes particular groupings of spatial locations that correspond to the objects, and this priming is then apparent in the Ends displays.

3. **Object-based**. A third possibility is that the grouping process utilizes shape representations that are not viewpoint specific. On this view, experience with a particular shape does not only effectively prime the grouping of its features in the specific spatial locations in which they appear. In addition, the experience may also prime grouping of features in other locations that correspond to that same object in a different position, scale, or orientation than the original one. Or it may prime some more abstract representation of the object, one not tied to any particular exemplar.

The aim of this experiment was to test the hypothesis contained in the third view described above: that the experience-induced object effect is not absolutely position-specific, but rather generalizes to instantiations of an object that are never actually presented in any full-object displays.

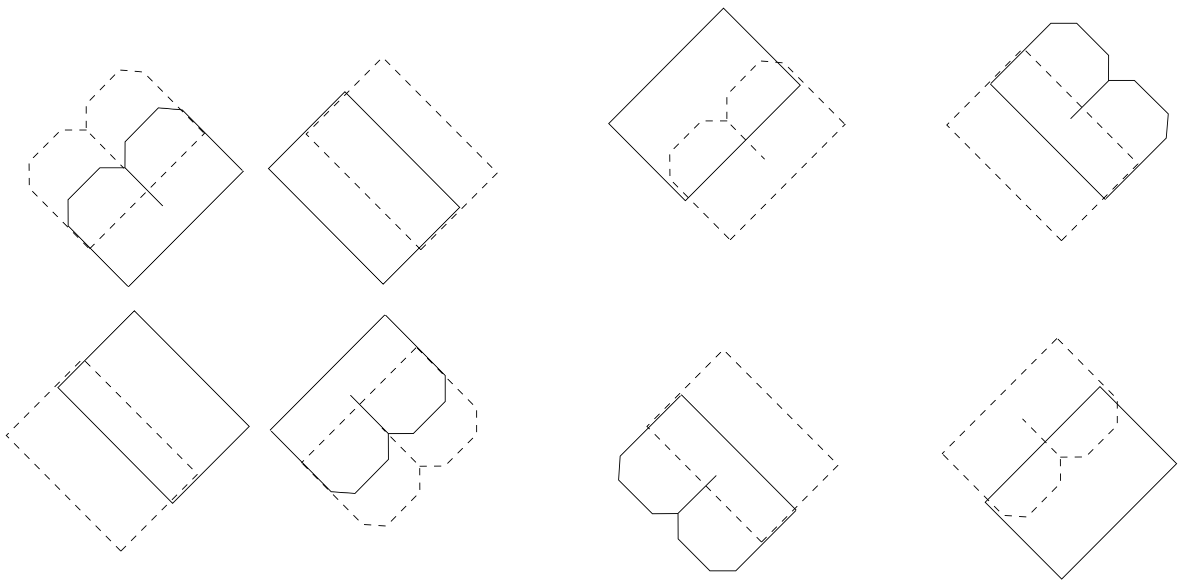
We tested this by manipulating the relative positions within the Ends displays, while not changing the X displays. We modified the design used in Experiment 3—in which the subjects saw a block of X displays, followed by a block of Ends—in two ways:

- We added an initial pretest block containing Ends displays. This addressed a limitation in the design of Experiment 3, in which, unlike Experiments 1 and 2, there was no baseline block which could be used to assess the effects of experience with a completed object. Nonetheless, any transfer to the Ends based on perceptual experience with the X leads to the same prediction as in Experiment 3: the Ends-Diagonal will be faster in block 3 than Ends-Vertical, since the diagonal bumps correspond to a single object in the disambiguating full object display.
- The features (bumps) in the Ends displays were in different locations than the X displays. There were two types of Ends displays, one in which the distance between the features was larger and the other smaller than the X displays, so the feature locations did not match between the Ends and X. The critical question here is whether subjects still obtain the benefit of block 2 such that Ends-Diagonal will be faster in block 3 than Ends-Vertical even when the feature locations in the X and Ends displays do not correspond.

Note that the manipulation of the feature locations in the Ends displays did not explore the full range of possible viewpoint variations, including location, orientation, and size. Without explicitly considering manipulations along each of these dimensions, this approach still allowed an exploration of the basic issue of whether the object benefit from the X displays applied to the Ends displays even when the feature locations did not match.

Method

Participants. Twenty-four subjects, ten male and fourteen female, between 18 and 23 years of age were recruited from the undergraduate subject pool at the University of Arizona. All subjects had normal or corrected visual acuity by self report, and were unaware of the purpose of the experiment.



(a) (b)

Figure 10: Examples of modified Ends displays used in Experiments 4. (a) Ends-Small. (b) Ends-Large. The dashed lines in both displays depict the original Ends stimuli (these are the same size in (a) and (b)).

Apparatus and materials. The same apparatus used in Experiment 2 was used here. The displays included the X displays used in Experiment 3 as well as modified versions of the Ends displays used in Experiments 2 and 3. Two variations of the original displays were created by shifting the locations of the Ends: in the Ends-Small set, the Ends were all moved towards the center of the display by 1 cm (1.2°); in the Ends-Large set, the Ends were moved away from the display center by the same amount (see Figure 10). As a result of this manipulation, the bumps in the Ends displays were not in the same locations as in the full, disambiguating displays, unlike the previous experiments.

Procedure. At the beginning of the experiment, subjects were shown printouts containing examples of the Large and Small Ends displays and were instructed to make same/different judgements on the number of bumps. Each trial proceeded as in the previous experiments. As before, the subject's task was simply to decide whether the number of bumps on the two ends was the same or different. The experiment was run in 3 blocks with a few minutes break between each block. Trials were randomized within a block. Prior to starting the experiment, subjects were given 32 practice trials, one of each of the full set of Ends displays.

Design. The design was entirely within-subject. The important independent variables are exactly as in Experiment 2, except that there is an additional variable (Large, Small) for the Ends displays. As in Experiment 2, the experiment was conducted in a series of three blocks. In the first block, the subjects saw only the Ends, with Large and Small randomly inter-mixed. In the next block, the subjects saw only X displays. The X displays were intermediate in position, and matched neither the Large nor Small displays. The final block consisted solely of Ends trials, again with Large and Small intermixed.

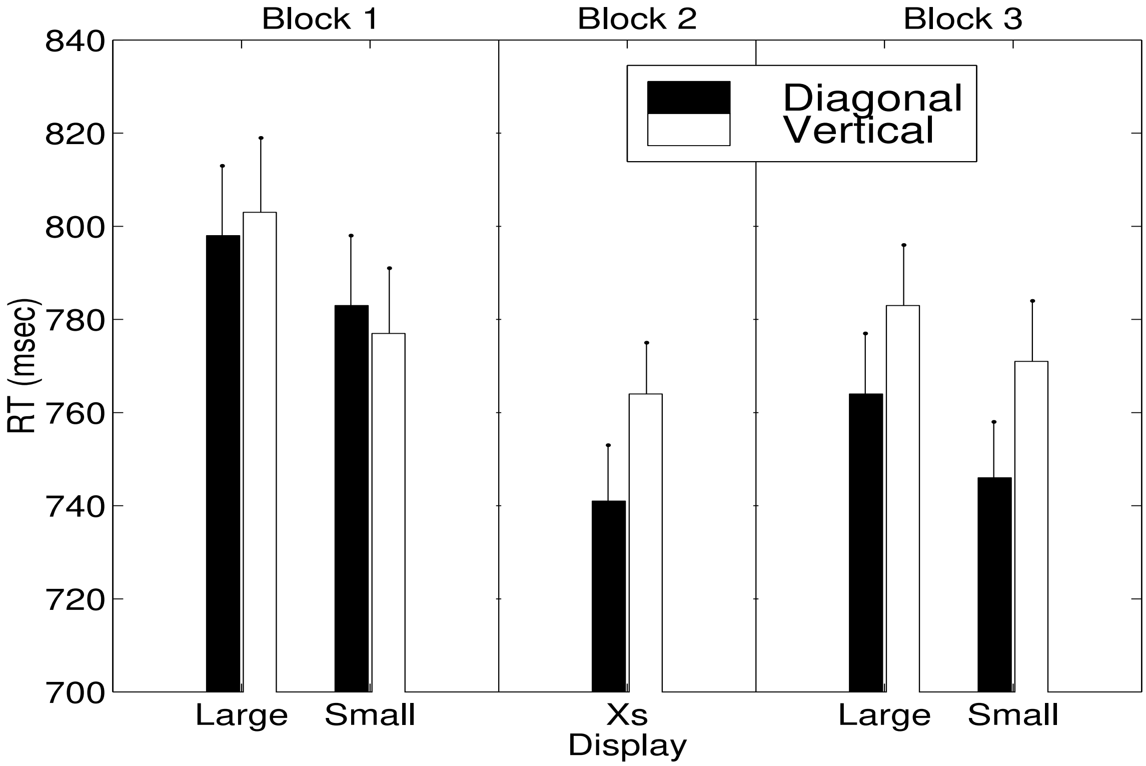


Figure 11: Mean RTs as a function of display (Ends-Large, Ends-Small, X), Block (1, 2, 3), and feature location (Diagonal, Vertical) in Experiment 4. Note that Diagonal corresponds to a single object while Vertical corresponds to the two object condition.

Results and discussion

The errors constituted 3.4% of the trials. As in the previous experiments, trials exceeding the two standard deviation cutoff were removed, resulting in the removal of an additional 2.3% of the trials. We found no significant effects nor interactions in an analysis of variance with percent error as the dependent measure and feature location (Diagonal, Vertical), size (Small, Large), epoch (Initial, Test), and judgement (Same, Different). This ANOVA was conducted on the data from the first and third block (Initial and Test epoch) of the experiment, since these blocks utilized the same stimulus sets.

We conducted a similar ANOVA with mean RT as the dependent measure. The mean correct RT data are illustrated in Figure 11. There was no significant effect of judgement in this experiment, ($F(1, 23) = .94, p = .23$), and no interactions with any other factor, so the data was pooled across this variable. Each of the other three factors considered individually were highly significant: feature location ($F(1, 23) = 18.3, p < .001$), epoch ($F(1, 23) = 40.04, p < .001$), and size ($F(1, 23) = 29.1, p < .001$). The only interaction between these variables is a pairwise interaction between feature location and epoch ($F(1, 23) = 12.6, p = .001$). This interaction is critical to our hypothesis, as it reflects the object cost induced by the intervening block of X displays.

Given this interaction, we conducted separate analyses of variance on correct RTs crossing feature location (Diagonal, Vertical) and size (Large, Small) for the blocks 1 and 3 of the experiment, and another ANOVA crossing feature location and orientation for block 2.

The first result of these within-block studies was the lack of a significant difference for feature location in block 1: $F(1, 23) = 0.01, p = 0.92$. RTs were shorter for Small displays than Large ones, and this difference was significant: $F(1, 23) = 39.14, p < .001$. There was no interaction between the feature location and size variables: $F(1, 23) = 1.16, p = 0.29$. The mean RT difference between the Diagonal and Vertical feature locations was 8 msec for the Large, and 9 msec for the Small, in opposite directions. This result replicated and extended the results of block 1 of Experiment 2, showing that prior to exposure to a disambiguating display, the different feature locations of any Ends display (regular, large, or small), were treated equally.

In block 2, we replicated the object-cost for the X displays (Behrmann et al., 1998), as Diagonal-Bumps were processed significantly faster than Vertical-Bumps: $F(1, 23) = 8.44, p = 0.008$. The mean RT difference between these two conditions was 28 msec. Orientation was not significant for these displays, and there was no interaction between orientation and feature location.

The crucial data involved block 3. These results revealed a significant degree of generalization of the object cost to the different Ends displays. As in block 1, Ends-Small displays were processed significantly faster than Ends-Large (24 msec): $F(1, 23) = 23.4, p < 0.001$. More importantly, for both Ends displays, Diagonal-Bumps were processed faster than Vertical-Bumps ($F(1, 23) = 3.8, p = 0.064$). In block 3, the mean RT difference between Diagonal and Vertical was 22 and 31 msec for the Large and Small displays, respectively. When the data for the two displays were considered separately, the difference for Ends-Large approached significance ($F(1, 23) = 3.2, p = 0.077$), while the difference for Ends-Small was significant ($F(1, 23) = 4.3, p = 0.05$). The salient aspect of these results is that the difference between the single-object (Diagonal-Bumps) and two-object (Vertical-Bumps) conditions was approximately equal in blocks 2 and 3: 28 msec versus 22 and 31 msec.

This study demonstrates that the effects of perceptual learning in this task generalize to some degree to other screen locations. Exposure to a block of X displays led to faster processing of the diagonal Ends—the Ends pairs consistent with the objects in the X displays—in the subsequent block even though the exact feature locations did not match in the two blocks.

Note that we utilized relatively small manipulations of feature location in this experiment, as the feature locations only shifted by 1.2 degrees between the X and either Ends display. However, with respect to previous research on perceptual learning, e.g., orientation discrimination and Vernier acuity, which often show extraordinary sensitivity to retinal location, these actually consitute fairly large shifts.

The results of this final experiment argue against the first two of the three alternatives presented above, that the learning is screen location specific or tied to location-specific shape representations. The learning does not appear to be occurring in the earliest stages of neocortical visual information processing. We return to this issue below.

General Discussion

Numerous studies have shown that the parsing of a visual scene is an important factor affecting the distribution of attention. Spatial proximity clearly plays an important role in this parsing process (Posner, Snyder, & Davidson, 1980; Tsal & Lavie, 1988). Many other studies have shown that attention can be directed to objects rather than to contiguous regions of the visual field (Duncan, 1984; Egly, Driver, & Rafal, 1994; Kramer & Jacobson, 1991). Earlier studies (discussed below) have shown effects of attention being applied to objects that are defined based on various factors, including generic grouping principles, highly familiar shapes, and task instructions.

The results presented in this paper extend these findings to apply to recently viewed novel shapes. The primary result in these experiments is that object attention benefits—shorter reaction times to features appearing on a single object—are obtained for newly-learned objects. We have previously found a benefit for features appearing on visible parts of an occluded object (Behrmann, Zemel, & Mozer, 1998). The results reported here are in some sense more surprising: these benefits apply even when the visible parts correspond to objects to which standard completion heuristics, e.g., relatability, do not apply. Relatively brief exposure to a novel, odd-shaped linking object suffice to induce object-based attention to fragments that can be interpreted as the visible parts of that object under occlusion. This is a particularly important result as it attests to the dynamic and flexible operation of the perceptual system, adapting as a function of experience.

A second primary finding is that these experience-dependent object benefits can apply to fragments even without any evidence of occlusion. The effects of experience were strong enough to overcome evidence that the fragments were separate objects (i.e., the presence of terminators in each End). Also, the results of this study show that uniform connectedness (Palmer & Rock, 1994) is not necessary for object attention. Whereas Kramer and Watson (1995) found no object effects when an object's uniform connectedness was disrupted by a new region of different color or texture, the findings here demonstrate a robust object effect.

Factors determining perceptual organization

A central issue in the area of object attention is what defines an 'object' of attention. Several studies have reinforced the pivotal role of perceptual organization. This role has been demonstrated in distractor studies, which show that it is difficult to ignore information that belongs to the same object or group as task-relevant information. For example, by virtue of common fate, identification of a central target was more affected by distant distractors that moved in the same direction as the target than by nearby static distractors (Driver & Baylis, 1989). Similarly, the response-compatibility effect (enhancement from similar, and inhibition from dissimilar distractors) was reduced when the target and distractors were embedded in or grouped with different objects compared to when they were grouped on the same object (Kramer & Jacobson, 1991); for other examples, see (Baylis & Driver, 1992; Bundesen, 1990).

However, generic grouping principles do not suffice to define the objects of attention. Past experience or familiarity appears to play a role as well. For example, when subjects decided whether two "x"s appear on the same or on two different superimposed letters or non-letters, performance was superior on letters than on nonletters (Vecera, 1993), and subjects were faster on upright letters than upside-down letters (Vecera & Farah, 1997). Joseph and Nakayama (1999) showed that the experience of seeing a full object, such as a rectangular bar, before partial occlusion influences the way in which objects are represented after occlusion has occurred. This work is similar to the studies presented here, in that recent experience is shown to determine the representation of occluded shapes. Our experiments examine different aspects of this issue: their clever manipulation allows for examination of trial-to-trial changes, while our methodology permits probes of both within- and between-object effects within a single trial.

While these studies demonstrate that attention can also be allocated preferentially to familiar shapes, other studies have shown that task instructions can be used to suggest a particular parsing of the scene. Yantis (1992) asked subjects to track five out of ten randomly moving dots, and to indicate, after all the dots had stopped moving, whether or not a particular dot was a member of the target set. He found that subjects who were encouraged to group the target dots as a higher order form or ″object″ performed better in the early phases of the experiment than those who saw the same stimuli but did not receive such encouragement. In a more directly relevant study, Chen (1998) found an object effect when subjects were instructed to view a display as two separate objects, but no effect when the instructions suggested a single-object interpretation, for the identical stimulus configuration. Baylis and Driver (1993) also used task instructions to get subjects to interpret the same displays in different ways. Our results extend these findings. The earlier studies showed that *explicit* instructions influence image processing; ours suggests that one can view perceptual experience as a form of *implicit* instruction about how to parse an image.

One possible formulation of these different effects on perceptual organization is top-down versus bottom-up. Generic grouping principles could constitute bottom-up factors, while experience and instructions could be viewed as top-down factors. However, we prefer to avoid these terms, as they have different meanings in different contexts. In terms of information processing, "bottom-up" may include information that is traditionally thought of as "top-down." For example, domain knowledge, such as familiar relations between image features, can easily be integrated into "bottom-up" connections in a neural network. This is what MAGIC does. So the same effect, such as the experiments in this paper, can be viewed as either bottom-up or top-down depending on one's theoretical perspective and interpretation of the terms. In addition, perceptual experience can play a pivotal role in both directions. Therefore we have avoided this terminology.

Considered in conjunction with the earlier results on object-based attention, our findings highlight the role of perceptual organization in the allocation of attention. In all of the experiments presented here, we contrasted subjects' responses to the same stimulus pattern as a function of their perceptual experience. In every study, the recent experience had a significant influence on their organization of the displays, as reflected in their responses. Short-term shape familiarity, as well as long-term familiarity and generic grouping principles, affect the scene organization and attentional allocation. The influence of these processes will vary under different conditions. For example, recent results show that short-term experience with a shape does not always over-ride the influence of standard grouping principles (Pratt & Sekuler, 2001). The interaction of these various factors, and their relative impact on perceptual organization require further study.

Underlying mechanisms

What mechanisms underlie the results described here, and the growing body of object attention findings? Here we focus on two issues: mechanisms that concern the relationship between object and spatial attention; and the representation of familiar shapes.

For some time, space and object attention had been considered to be mutually exclusive alternatives (Kanwisher & Driver, 1992). More recently, attempts have been made to reconcile the two forms of attention. Egly and colleagues (Egly et al., 1994) have argued that both processes co-exist. For example, they reported that a cost in RT and accuracy is incurred when attention is shifted between a cue and a target both when the target appeared at a second location in the cued object (within-object, object attention) or at an equidistant location but in a different object (between-object, spatial attention). Data favoring the simultaneous operation of space- and object-based processes also come from a study by Umilta et al. (1995) who cued a vertex of a cube which either remained stationary or rotated. Subjects not only showed facilitation when the target appeared in the same spatial or retinal location as the cue (stationary) but also when the target appeared in a different retinal location but in the equivalent object-defined location as the cue (rotated condition). Similar findings are revealed in studies on inhibition of return in both location- and object-based coordinates (e.g., Gibson & Egeth, 1994; Tipper & Weaver, 1996).

Relatively few accounts integrating object- and space-based attention have been formulated. One proposal is a *two-stage feedforward* model in which spatial attention follows object-based attention. Under this account, visual routines identify regions of salience or coherence in the visual field pre-attentively and in parallel. These regions are then subjected to further analysis by focal spatial attention processing (Julesz, 1981; Koch & Ullman, 1985; Neisser, 1967; Treisman, 1982; 1988). This view of has been proposed to account for numerous findings in the visual search literature as well as findings in which grouping, based on feature similarity or proximity, occurs early, in parallel and independent of spatial attention (Driver, Baylis, & Rafal, 1992; Marshall & Halligan, 1994; Moore & Egeth, 1997). One characterization of this view posits that object-based effects occur because space-based attention automatically spreads from the local, task-relevant region to the entire extent of the perceptual group encompassing that task-relevant region. In line with this proposal, it has recently been shown that object-based attentional effects disappear when the spatial extent that visual attention has to cover is equated in the within- and between-object conditions (Davis, Driver, Pavani, & Sheperd, 2000). These results indicate that object-based attention may be explained, at least in part, by the greater spatial extent space-based attention has to cover when judging parts of two objects, rather than by a fixed inability of visual attention to be focused on more than one object at a time.

An alternative to this feedforward scheme is one in which object- and space-based processes operate in parallel and mutually influence each other (Farah, 1990; Humphreys & Riddoch, 1991; Humphreys et al., 1996). The interaction may occur through a topographically organized *grouped array*, which represents the currently active bottom-up input from the environment as well as top-down activation from matching higher-level descriptions. Through this explicit array, spatiotopic information and grouping information are both present and simultaneously influence visual processing. Vecera (1994) claimed that such an array-like representation must exist; using the Egly et al. (1994) paradigm, he showed not only that there is a cost associated with shifting attention within and between objects but also that the cost of shifting attention between objects increased as the spatial distance between the objects increased. Similarly, as is usually the case in the distractor paradigm, Kramer and Jacobson (1991) showed that the response compatibility effects were diminished when the spatial distance between the grouped features was increased. Taken together, these findings suggest that both space- and object- selection are operative in parallel.

Within this parallel account of attention, the issue of how the grouped array operates is still open. The original proposal was that a combination of generic grouping principles and shape-specific information act to label the array locations (Vecera & Farah, 1994; Kramer et al., 1997). Geometric properties of the display, such as the relatability of fragments (Kellman & Shipley, 1992), would be fundamental elements of the grouping component.

With respect to the influence of familiar shapes, the standard conception is that this involves representations of whole objects. Within this view, several possibilities exist. An object could be: (a) exact exemplar, specific to particular spatial locations and orientations; (b) fuzzy exemplar, specifying a particular shape, but less specific in its spatial instantiation; or (c) spatially invariant object representation. The results of the experiments presented here do not bear on the spatially invariant representation hypothesis, because the objects generally occupied the same spatial position. Limited evidence exists for this view: the results of Vecera and Farah's (1994) study implicated spatially-invariant object representations, but other studies have not found evidence for them (for further discussion, see Kramer et al, 1997). Experiment 4 in this paper provides evidence that the object effect can transfer to different feature locations, which makes the exact exemplar representation unlikely. Instead, these results are consistent with the fuzzy exemplar representation, as there was some spatial overlap between the learned feature locations and the generalized locations. In addition, the fact that the degree of transfer was greater to the Small-Ends (seen in the significant difference in the feature location conditions for Small-Ends versus a trend for Large-Ends) is also consistent with the fuzzy exemplar, under the assumption that the object attention benefit extends to all locations encompassed by the viewed exemplar.

An alternative conception of the shape-specific component of the grouped array involves learned configurations of local features rather than whole objects. This mechanism is consistent with the computational model, MAGIC (Mozer et al., 1992). Under this interpretation, the disambiguating displays primarily serve to facilitate the grouping of particular pairs of Ends in the displays; and this grouping then applies to the ambiguous displays. Our results provide stronger evidence for this account, based on the finding that the grouping of the Ends appears to operate even in the presence of terminators, which provide evidence that the Ends are complete objects themselves. This feature-based representation can also account for the results of Experiment 4, assuming a feature-based analog of the fuzzy exemplar model proposed for the whole-object representation.

We note, however, that the present studies cannot support or disconfirm the MAGIC model. If we had not found that recent experience influences processing in this task, then it would be possible that such an influence could be produced simply by a longer learning period. The effects we did find are consistent with MAGIC, yet the specific feature-based grouping mechanisms that underlie MAGIC are not the only possible explanation. Indeed, a framework that allows for the rapid formation of novel objects, and their influence on perceptual organization, may also be able to account for the data we present here.

Finally, an important point is that all of these different mechanisms can be learned from statistical structure in the environment. Whole objects, or particular local feature configurations, can both be extracted based on experience with various feature combinations in images. The evidence for experience-dependence provided in this paper further indicates that such higher-order statistical regularities play a critical role in visual perception.

Placed in a larger historical context, our results are therefore consistent with the Brunswick school of perception, favoring the influence of statistically-learned rules of visual ecology more than purely stimulus-driven Gestalt principles (Palmer, 1999). Note that the dichotomy between these two schools is not as sharp as commonly believed. Despite the standard nativist view of Gestalt principles, even pioneers of this school posited a role for perceptual experience in grouping. Wertheimer (1923) describes how a stimulus that would normally be grouped in one way may be perceived with a different grouping given a particular recent perceptual experience. Our results provide direct evidence for this proposal.

References

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

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

Behrmann, M., Zemel, R. S., and Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4):1011–1036.

Behrmann, M., Zemel, R. S., and Mozer, M. C. (2000). Occlusion, symmetry, and object-based attention: Reply to Saiki (2000). *Journal of Experimental Psychology: Human Perception and Performance*, 26(4):1497–1505.

Ben-Av, M. B., Sagi, D., and Braun, J. (1992). Visual attention and perceptual grouping. *Perception and Psychophysics*, 52:277–294.

Bregman, A. S. (1981). Asking the 'What for' question in auditory perception. In Kubovy, M. and Pomerantz, J. R., editors, *Perceptual Organization*, pages 99–118. Lawrence Erlbaum Associates, Hillsdale, NJ.

Bub, D. and Gum, T. (1991). Psychlab experimental software. Technical report, McGill University, Montreal.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97:523–547.

Chen, Z. (1998). Switching attention within and between objects: The role of subjective organization. *Canadian Journal of Experimental Psychology*, 52:7–16.

Davis, G., Driver, J., Pavani, F., and Shepherd, A. (2000). Reappraising tne apparent costs of attending to two separate visual objects. *Vision Research*, 40:1323–1332.

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

Driver, J., Baylis, G., and Rafal, R. D. (1992). Preserved figure-ground segregation and symmetry perception in visual neglect. *Nature*, 360:73–76.

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

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

Farah, M. J. (1990). *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision*. MIT Press, Cambridge, MA.

Gibson, B. S. and Egeth, H. (1994). Inhibition of return to object-based and environment-based locations. *Perception and Psychophysics*, 55:323–339.

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.

Humphreys, G. W., Olson, A., Romani, C., and Riddoch, M. J. (1993). Competitive mechanisms of selection by space and object: A neuropsychological approach. In Kramer, A. F., Coles, M. G. H., and Logan, G., editors, *Converging operations in the study of visual selective attention*. American Psychological Association, Washington, DC.

Humphreys, G. W. and Riddoch, M. J. (1991). Interactions between object and space systems revealed through neuropsychology. In Meyer, D. E. and Kornblum, S., editors, *Attention and Performance XIV*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Joseph, R. and Nakayama, K. (1999). Amodal representation depends on the object seen before partial occlusion. *Vision Research*, 39:283–292.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, 290:91–97.

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

Kanwisher, N. and Driver, J. (1992). Objects, attributes and visual attention. *Current Directions in Psychological Science*, 1:26–31.

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

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

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227.

Kramer, A. F. and 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. and Watson, S. E. (1995). Object-based visual selection and the principle of uniform connectedness. In *Converging operations in the study of visual attention*, pages 395–414. American Psychological Association.

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

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

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

Marshall, J. C. and Halligan, P. W. (1994). The yin and yang of visuo-spatial neglect: A case study. *Neuropsychology*, 32:1036–1057.

Moore, C. M. and Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, 23:339–352.

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

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

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

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

Palmer, S. and Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review*, 1:29–55.

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA.

Pratt, J. and Sekuler, A. B. (2001). The effects of occlusion and past experience on the allocation of object-based attention. *Psychonomic Bulletin & Review*, In Press.

Shipley, T. F. and 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.

Tipper, S. P. and Weaver, B. (1996). The medium of attention: Location-based, object-centered or scene-based. In *Visual Attention*. Oxford University Press, Oxford, UK.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8:194–214.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 40A(2):201–237.

Tsal, Y. and Lavie, N. (1988). Attending to color and shape: The special role of location in selective visual processing. *Perception and Psychophysics*, 44:15–21.

Umilta, C., Castiello, U., Fontana, M., and Vestri, A. (1995). Object-centred orienting of attention. *Visual Cognition*, 2:165–182.

Vecera, S. P. (1993). Object knowledge influences visual image segmentation. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, pages 1040–1045, Hillsdale, NJ. Erlbaum.

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

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

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24:295–340.

Yantis, S. (1995). Perceived continuity of occluded visual objects. *Psychological Science*, 6:182–186.