Atten Percept Psychophys (2014) 76:2240–2248
DOI 10.3758/s13414-014-0708-1

Visual statistical learning can drive object-based attentional selection

Libo Zhao · Joshua D. Cosman · Daniel B. Vatterott · Prahlad Gupta · Shaun P. Vecera

Abstract Recent work on statistical learning has demonstrated that environmental regularities can influence aspects of perception, such as familiarity judgments. Here, we ask if statistical co-occurrences accumulated from visual statistical learning could form objects that serve as the units of attention (i.e., object-based attention). Experiment 1 demonstrated that, after observers first viewed pairs of shapes that co-occurred in particular spatial relationships, they were able to recognize the co-occurring pairs, and were faster to discriminate two targets when they appeared within a learned pair ("object") than when the targets appeared between learned pairs, demonstrating an equivalent of an object-based attention effect. Experiment 2 replicated the results of Experiment 1 using a different set of shape pairs, and revealed a negative association between the attention effect and familiarity judgments of the co-occurred pairs. Experiment 3 reports three control experiments that validated the task procedure and ruled out alternative accounts.

Keywords Attention · Statistical learning · Object-based attention · Perceptual organization

Introduction

As Wertheimer (1923/2000) noted, perception is not based on the raw inputs to vision; Instead, visual information is organized into perceptual groups (Palmer, 1999, 2002). Typical visual scenes contain many perceptual groups, and, because of the large number of groups in any visual scene, the wealth of visual information is parsed into behaviorally relevant groups, resulting in object-based attentional selection. When attention is summoned to an object, attention spreads within that object before moving on to other objects (Hollingworth, Maxcey-Richard, & Vecera, 2012; Richard, Lee, & Vecera, 2008).

There is no shortage of demonstrations of object-based attention (see Chen, 2012), and many object-based attentional effects can be explained as attention operating across a grouped array of locations (Vecera, 1994; Vecera & Farah, 1994. The grouped array account proposes that Gestalt grouping principles (e.g., similarity, common fate, closure) package locations and features which then constrain the spread of spatial attention. In one widely-used object-based attention task (Egly, Driver, & Rafal, 1994), attention is summoned to one region of an object; Gestalt grouping cues establish a 'grouped array' and allow spatial attention to spread further or faster within that attended object than to an unattended object.

Accounts of object-based attention have tended to study salient, image-based grouping principles in which locations are organized by physical attributes such as a common color or physical connectedness. There is, however, evidence demonstrating that object-based selection extends beyond Gestalt principles. Specifically, a number of studies have shown that past experience with an object or objects can contribute directly to object-based effects (Li & Logan, 2008; Vecera & Farah, 1997; Zemel, Behrmann, Mozer and Bavelier 2002). For example, Vecera and Farah (1997) found that object-based responses were faster for familiar objects (upright letters) than for less familiar objects (upside-down letters). Similarly, Zemel et al. (2002) demonstrated that, after training, an occluded z-shaped object would produce an object-based effect when occluded, despite the ends of the object being misaligned. Such misalignment would typically abolish completion behind the occluder, as well as object-based selection, but experience can override the absence of completion cues.

L. Zhao · J. D. Cosman · D. B. Vatterott · P. Gupta · S. P. Vecera (✉)
Department of Psychology, University of Iowa, E11 Seashore Hall, Iowa City, IA 52242-1407, USA
e-mail: shaun-vecera@uiowa.edu

Perhaps the most direct evidence for experience-dependent effects on object-based attention based on long-term learning, and that which is most relevant to the current work, comes from Li and Logan (2008), who found that native Chinese readers would show object-based attentional effects for Chinese words composed of two characters. When attention was drawn to one of a pair of Chinese characters, Chinese readers were faster to respond to a target on the other character of the pair when it formed a word with the cued first character than when it did not form a word with the cued character.

Despite these demonstrations of familiarity on object-based attention, the mechanism underlying these effects has received little study. One plausible learning mechanism for the previously reviewed results is visual statistical learning, in which the relationship between co-occurring visual features or objects is learned across space (Fiser & Aslin, 2001) and time (Fiser & Aslin, 2002; Turk-Browne, Jungé and Scholl 2005). In a prototypical visual statistical learning task, observers complete a familiarization task in which they view 'scenes' containing several shapes, and they are instructed to view the displays for a later test. Importantly, the relative spatial relationships between particular shapes are consistent across scenes during familiarization (Fiser & Aslin, 2001). For example, a scene of six shapes might contain three pairs, one pair in which the shapes are always to the left and right of each other, another pair with shapes above and below each other, and a final pith an oblique relationship. After viewing a number of displays, observers were shown two pairs of shapes in sequence, and are asked to discriminate 'old' configurations of shape pairs (i.e., those shown in the same relative positions as in the familiarization session) from new configurations (those shown in different relative positions than the familiarization session). Although the shapes do not appear in precisely the same context in which they were learned, observers nevertheless accurately discriminate the configurations of old pairs from those of new pairs (on the order of 70 % correct). Thus, observers readily learn co-occurrences between pairs of shapes in complex visual scenes, providing a possible substrate for object-based attention, in which visual statistical learning creates the units or groups that guide attention.

Although statistical learning is the likely source of many object-based attentional findings (e.g., Li & Logan, 2008), most previous research has examined how image-based grouping information influences learning, not the converse. Baker and colleagues (2004) examined the effect of attention and perceptual grouping by connectedness on statistical learning. In their most relevant experiments, observers attended to one location occupied by a shape and learned to associate that shape with a response. Another shape appeared at an unattended location; this shape did not predict the response, but it did indicate influence RTs to the attended shape. This influence occurred only when the unattended shape was physically connected to the attended shape, indicating that statistical learning is influenced by perceptual grouping (specifically, connectedness).

Another demonstration that directly links statistical learning and object-based attention comes from Vickery and Jiang (2009). In that study, observers were first shown displays containing several shapes with pairs of these shapes co-occurring inside a common region; observers were later tested in the absence of the common region information. During testing, observers were faster to search for repeating colors when those repetitions occurred within a single learned pair (within a group) compared to repetitions across different pairs (across groups). This result shows that statistical co-occurrence plus an extrinsic grouping cue can lead to grouping effects that persist when the grouping information is removed. However, these findings do not address whether learning the co-occurrences of specific visual features can itself act as a grouping cue that drives object-based attention effects. There is evidence, however, that statistical learning can influence the deployment of spatial attention: Visual search is biased toward locations that contain a predictable sequence of shapes (Zhao, Al-Aidroos, & Turk-Browne, 2013), although the testing procedure in this latter study could reflect attentional capture by the novel search array, which was not part of the predictable sequence of shapes.

In sum, although there is a rich literature on the relationships among attention, perceptual grouping, and visual statistical learning, unanswered is the question of whether shape co-occurrences themselves can produce perceptual groups that guide attention. In the current work, we examine how statistical learning acts to group shapes in the absence of stimulus-based evidence, such as connectedness or common region, forming the basis of object-based attention.

We also examined a corollary that would arise if co-occurring visual elements were to demonstrate an object-based attention effect: What might be the relationship between such an attentional effect and the ability to judge the familiarity of co-occurring pairings? Some unpublished findings discussed in Vickery and Jiang (2009) showed that training methods that produced superior familiarity judgments of the co-occurring shape pairs failed to yield evidence of perceptual grouping, while training methods that produced poor familiarity judgments of those pairs led to strong effects of perceptual grouping. These findings suggest that object-based attentional effects and familiarity judgments both arise during a visual statistical learning task but may be negatively associated. Consistent with this, a recent fMRI study on visual statistical learning showed that both the medial temporal lobe and the striatum were sensitive to statistical structures, and that the striatum showed increased activation for co-occurring shape triplets no matter whether they were recognized or not (Turk-Browne, Scholl, Chun, & Johnson, 2009; also see Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014). These findings suggest that different kinds of learning, mediated by different neural structures, may occur in parallel during a visual statistical learning task, and may be unrelated to each other. Thus, to explore the potential relationship between object-based attention and familiarity judgments, the present study included both an object-based attention test and a measure of familiarity of the co-occurring visual elements.

To preview, the present study yielded two findings. First, we demonstrated in Experiment 1 that shape pairs that co-occurred in the absence of any perceptual cues in an initial training phase later exhibited an object-based attentional benefit in addition to being recognizable. Second, in Experiment 2 we showed that the ability to recognize the co-occurring pairings was negatively associated with the object-based attention effect. Finally, in three control experiments (Experiments 3A–C) we validated the task procedures and ruled out alternative accounts.

Experiment 1

The shapes and training scenes used in this experiment were similar to those used in other statistical learning tasks (e.g., Fiser & Aslin, 2001; see Fig. 1). During a familiarization phase observers were exposed to structured scenes that consisted of two pairs of shapes (4 shapes total) that were randomly chosen from a larger set (see Fig. 1a). The two shapes in a pair always co-occurred in a fixed spatial configuration (horizontal or vertical) throughout the familiarization phase (see Fig. 1b). As can be seen, the shapes in the training scenes were unconnected to each other and image-based cues did not explicitly mark the boundaries of the base pairs. Thus, statistical co-occurrences of specific shapes across scenes were the only information that could be used to bind the shapes into groupings.

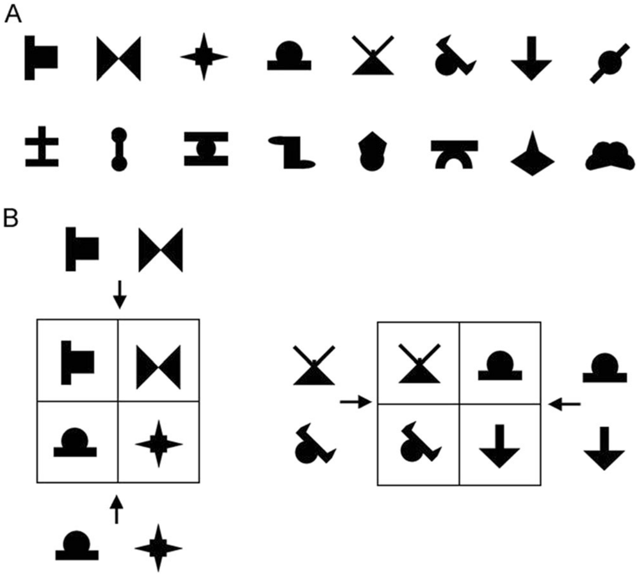

Fig. 1 Shapes and examples of exposure scenes used in this study. **a** The complete set of 16 shapes used in this experiment. **b** The example scenes at the exposure phase, the *left* one consists of two horizontal base pairs, and the *right* one consists of two vertical base pairs

Afterward, we used an object-based attention task to test if attention obeyed the units or groups established by the shapes' co-occurrences. In this task, shown in Fig. 2, observers reported whether two bars on the sides of the grid were the same color or different colors. If the co-occurring shape pairs were treated as a single unit by attention, a faster response was expected when the bars fell on the axis (horizontal or vertical) on which objects appeared during the familiarization phase (Fig. 2a) compared to when they did not (Fig. 2b). Observers also performed a familiarity judgment test similar to that used by Fiser and Aslin (2001), in order to demonstrate that the learned shape pairs could be distinguished from randomly grouped pairs.

Method

Subjects Thirty-five University of Iowa undergraduates participated for course credit and reported normal or corrected-to-normal visual acuity.

Stimuli Sixteen novel black shapes were used, twelve adapted from Fiser & Aslin (2001) and four newly created (Fig. 1, top panel). The size of these shapes varied slightly, but each measured approximately 1.33° of visual angle on the longer dimension (either length or width). Unknown to the observers, the shapes were organized into eight base pairs whose two shape elements always co-occurred in a fixed spatial relation. Four base pairs were horizontally oriented and four were vertically oriented. Scenes were created by randomly choosing two base-pairs of the same spatial orientation and then placing them side by side in a 2 × 2 grid (Fig. 1, lower panels). The grid measured 4.18° square. The scenes appeared on a 15" CRT monitor powered by a Macintosh Mini computer, using MATLAB and the Psychophysics Toolbox (Brainard, 1997).

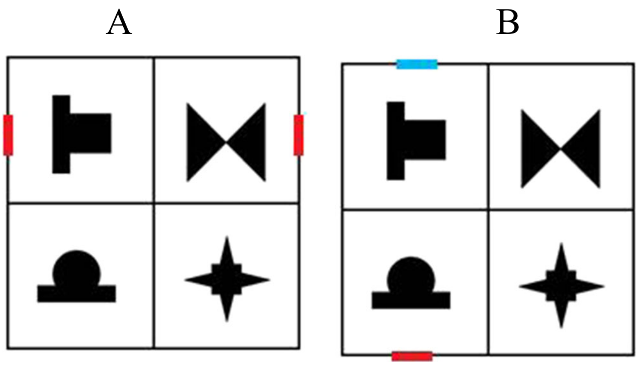

Fig. 2 Examples scenes used in the object-based attention test. **a** The two *bars* are on the base pairs, **b** they are on non-base pairs

Task & procedure There were three phases in the experiment: the familiarization phase, the test phase, and the object-based attention phase. During the familiarization phase, observers viewed 384 scenes broken into blocks of 64 trials. Scenes appeared for 2 s, with a 1-s pause between adjacent scenes. There was a 1-min break between each block. We informed observers that they should attend to the scenes in order to answer questions about them later in the experiment.

The test phase occurred after a 1-min break following the familiarization phase. During this test phase, observers performed a two-alternative forced choice between a base pair and a non-base pair that were shown in two successive displays. After a 1-s fixation point, each pair of shapes appeared for 2 s with a 1-s interval between them. Observers indicated which one of the two pairs (the first or second) was more familiar by pressing a key on the computer keyboard ("N" or "B"). Non-base pairs were composed of two shapes drawn from two different base pairs; observers had never viewed the specific non-base pair configuration, although observers had viewed the individual shapes in this non-base pair. Four horizontal and four vertical non-base pairs were generated randomly for each observer. In these eight non-base pairs, each of the 16 shapes occurred once. The next trial began 1 s after observers pressed one of the two keys. Across the sixteen trials, each base pair appeared twice, and the order of base pairs and the non-base pairs was counterbalanced such that for half of the trials, the first pair was a base pair, and for the other half, the first pair was a non-base pair.

The final phase, the object-based attentional task, began after a 1-min break. In this phase, scenes identical to those in the familiarization phase appeared with the addition of two colored segments on the frame surrounding the shapes (see Fig. 2). In total, there were 384 such scenes that were broken into blocks of 64 trials. There was a one-minute break between adjacent blocks. In each trial, two parallel bars were presented on the edges of the grid. Each bar extended 1.0° and could be either red or blue. Observers were asked to report whether the two bars were of the same color or different colors by pressing a key on the computer keyboard ("N" or "B"). On half of the trials, the two bars fell on the two sides of a base pair (Congruent condition), and on the other half of the trials, the two bars were on different base pairs (Incongruent condition). On each trial, a scene was presented until observers pressed one of the two keys, and then after 1 s the next trial began.

Results and discussion

We first analyzed the accuracy of familiarity judgments. The results showed that observers were accurate in discriminating co-occurring (learned) pairs from novel pairs (72 % correct judgments regarding which of the sequentially presented pairs was more familiar), which was significantly above chance, $t(34) = 7.75$, $p < .001$, replicating previous results (e.g., Fiser & Aslin, 2001). Because the frequencies of the individual shapes in the base pairs and the non-base pairs were equal, statistical co-occurrences of the constituent shapes were the only information to differentiate between the two types of pairs. This initial finding suggests that observers learned statistical co-occurrences through incidental exposure, and this learning gave rise to groupings that are accessible and can drive familiarity judgments (possibly through the use of recognition memory).

Next, to determine whether this grouping information can drive object-based attention effects, we compared the mean RTs of correct responses between congruent and incongruent trial types in the object-based attention task. In this and all the following experiments, trials with RTs higher than 5,000 ms or lower than 150 ms were excluded from analysis to reduce the influence of the outliers (less than 1 % of the trials for each experiment). Observers responded significantly faster in the congruent condition (M = 604 ms, SD = 71 ms) than in the incongruent condition (M = 611 ms, SD = 69 ms), $t(34) = 2.15$, $p < .04$, $d = .35$, demonstrating the equivalent of an object-based attentional effect. The accuracy results showed no difference between the congruent condition (96.0 % correct) and the incongruent condition (95.7 %), $t(34) = 1.1$, $p = .27$, $d = .16$, and thus the slower response in the incongruent condition was not due to a speed-accuracy tradeoff. In sum, the results suggest that the co-occurring shapes enjoy an attentional benefit. Because there were no image-based grouping cues in any of our displays, these findings demonstrate that statistical co-occurrences in and of themselves are able to create perceptual units (objects) that serve as the units of attention.

We also examined the congruency effect in the object-based attention task between the first 50 trials and the final 50 trials. There was a congruency effect present in both intervals, with a slightly smaller congruency effect in the final 50 trials (6.1 ms) compared the first 50 trials (10.5 ms). None of these finer-grained analyses were statistically significant because of increased variability when the trial numbers were reduced. The relevant result is that the visual statistical learning from the initial familiarization phase persists across the entire object-based attention task, with little evidence for additional learning while observers perform that task.

Finally, to address our question regarding the relationship between object-based attention and familiarity judgments, we conducted a Pearson correlational analysis between accuracy in the familiarity judgment test and the object-based attention effect. The attention effect was calculated for each subject by subtracting the mean RT of the congruent condition from the mean RT of the incongruent condition. We found no reliable association between them, $r = .11$, $p > .50$, and thus failed to find support for the hypothesis of trading relations between familiarity judgment accuracy and attentional effects. There was no correlation between the familiarity judgment and the object-based effect in either the first 50 trials ($r = 0.01$) or the final 50 trials ($r = –0.15$).

However, because in this experiment the individual shapes were randomly assigned to form base pairs for each participant, the idiosyncrasies of the specific pairings might have influenced familiarity judgments and object-based attention differently and in unsystematic ways, leaving very limited power to detect any relationship between them. Therefore, in Experiment 2, we used a fixed set of base pairs of shapes for all observers, allowing us to maximize the chance of detecting this hypothesized relationship. In addition, Experiment 2 was intended to replicate the primary findings of Experiment 1.

Experiment 2

Method

Subjects Thirty-three University of Iowa undergraduates participated for course credit. All reported normal or corrected-to-normal visual acuity.

Stimuli, task, and procedure Experiment 2 was the same as Experiment 1 except that all observers received the same set of base pairs of shapes, which was randomly assigned.

Results and discussion

The basic findings in Experiment 1 were replicated. Observers recognized the co-occurring pairs from the random pairs at a rate significantly above chance (71 % correct), $t(32) = 6.5$, $p < .001$. They also responded to the color bars significantly faster in the congruent condition (M = 615 ms, SD = 87 ms) than in the incongruent condition (M = 622 ms, SD = 89 ms), $t(32) = 2.5$, $p = .02$, $d = .40$, and were marginally significantly more accurate in the congruent condition (95 %) than in the incongruent condition (94 %), $t(32) = 2.0$, $p = .05$, $d = .36$. Therefore, with this fixed set of shape pairings, the results once again suggest that statistical co-occurrences alone are able to create groupings that enjoy an object-based attention benefit and are at the same time accessible for familiarity judgments. As in Experiment 1, the congruency effect was present in the first 50 (12.8 ms) and final 50 (2.5 ms) trials of the object-based attention task, with a smaller effect in the final 50 trials than in the first 50 trials, although this effect was not statistically significant because of the increased variability when looking at a subset of the trials.

In contrast with Experiment 1, a Pearson correlational analysis, shown in Fig. 3, revealed a significant negative association between the object-based attention effect (the RT advantage of the congruent condition compared to the incongruent condition) and familiarity judgment accuracy, $r = –0.66$, $p < .001$. This correlation was present in both the first 50 trials ($r = –0.39$, $p < .05$) and the final 50 trials ($r = –0.36$, $p < .05$). Consistent with the previous findings (Turk-Browne et al. 2009; Vickery & Jiang, 2009), this negative association suggests that different kinds of learning may occur in parallel in a visual statistical learning task, and may be in a trading relation.

Experiment 3

Experiments 1 and 2 demonstrated that statistical co-occurrences alone, even without image-based cues, created groupings that serve as the unit of object-based attention and that are also accessible for familiarity judgments, and that these two effects supported by statistical co-occurrences are negatively correlated. However, two issues remain to be addressed. First, to test the premise that it is the objecthood of the co-occurring shape pairs (rather than something else) that leads to the attentional effect, it is necessary to establish that our attention task is indeed sensitive to object structures when these are explicitly marked by image cues. Second, it remains to be demonstrated that the attention effect, even if truly object-based, really is created by visual statistical learning rather than by some aspect of the test procedure itself such as a response-level bias. If the former, the effect should only arise after prior exposures to statistical co-occurrence, and not without; if the latter, object-based effects should be seen even without prior statistical learning opportunity.

To address these issues we conducted three further experiments. Experiment 3A examined whether our object-based attention task is sensitive to object structure, by using objects defined by Gestalt cues, specifically, common region and closure. Experiments 3B and Experiment 3C examined whether the attentional effect would appear *without* prior statistical learning experience. In Experiments 3B and 3C,

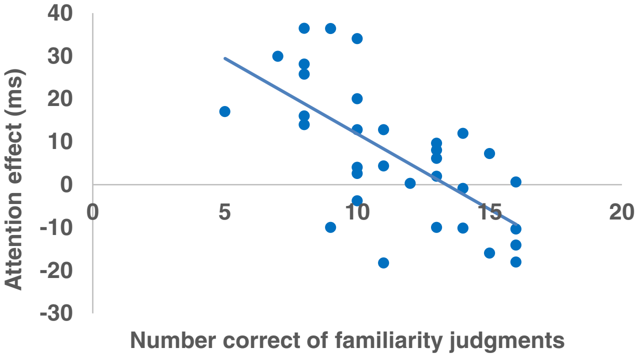

Fig. 3 Relationship between the magnitude of the object-based attention effect, as predicted by familiarity judgment accuracy, in Experiment 2

we removed the familiarization phase of our procedure, thereby eliminating prior exposure to the pairs of shapes.

Experiment 3A

To validate the sensitivity of our attentional task to object structures, Experiment 3A employed object groupings defined by Gestalt cues, while keeping the rest of the attentional task identical to that in the first two experiments. Specifically, we removed one middle line from the original two-by-two grid, resulting in two rectangles (groupings defined by closure) next to each other (see Fig. 4). Previous studies on object-based attention showed that a rectangle is selected by attention as a single unit (e.g., Egly et al. 1994; Vecera, 1994), so, if our task procedure is sensitive to object structures, the color bars should be responded to more efficiently when they are located on the same rectangle compared to when they are on two different rectangles. As the focus was on validating the attentional task as a measure specifically of object-based attention, no statistical learning trials were conducted.

Method

Subjects Twenty-three University of Iowa undergraduates participated for course credits. All reported normal or corrected-to-normal visual acuity.

Stimuli, task, and procedure Only the attentional task was administered. This was essentially the same as the one in Experiments 1 and 2 except for the following changes. First, the grid in this task was made by removing one line (either the horizontal or vertical one) in the middle of the original two-by-two grid, thus creating a display made up of two contiguous rectangles (see Fig. 4). Second, there were no shapes presented inside the grid.

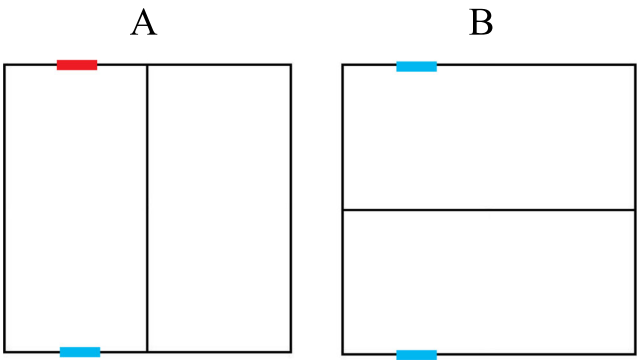

Fig. 4 Example scenes used in the object-based attention test in Experiment 3A. **a** The two *bars* are on one *rectangle*, **b** they are on two *rectangles*

Results and discussion

The results showed that observers responded to the color bars faster when they appeared on one rectangle (M = 568 ms, SD = 78 ms) than when they appeared on two rectangles (M = 574 ms, SD = 83 ms), $t(22) = 2.4$, $p < .05$, $d = .36$. In addition, there was no reliable difference in accuracies between these two conditions (95.6 vs. 95.0 %), $t(22) = 1.2$, $p > .20$, $d = .25$. These results suggest that our attention test procedure is sensitive to object structures, and validates it as a test of *object-based* attention. This justifies our inference in Experiment 1 and Experiment 2 that co-occurring pairs demonstrated an object-based attention effect when stimuli on co-occurring pairs were responded to more efficiently.

Experiment 3B

This experiment included only an object-based attention task, which was identical to the one in Experiment 1 (shape pairings were randomly assigned for each participant). The goal was to examine whether an object-based attention effect would emerge *without* the prior statistical learning experience that was provided in Experiment 1, and thus to examine the possibility that the attentional effect in Experiment 1 might simply have been a response-level effect at test.

Methods

Subjects Twenty-one undergraduates at University of Iowa participated for course credits. All reported normal or corrected-to-normal visual acuity.

Stimuli, task, and procedure An object-based attention task identical to the one used in Experiment 1 was presented to observers.

Results and discussion

The color bars were responded to equally fast in the congruent condition (M = 583 ms, SD = 87 ms) and in the incongruent condition (M = 580 ms, SD = 92 ms), $t(20) = .83$, $p > .40$, $d = .10$. As for accuracies, there was also no reliable difference between these two conditions (96.4 vs. 96.2 %), $t(20) = .35$, $p > .70$, $d = .11$. These results indicate that an object-based attention effect did not arise during the attention test itself, thus ruling out the possibility that the effect observed in Experiment 1 was due to any response-level bias at test or other aspect of the attentional task procedure itself. This in turn suggests that the object-based attention effect in

Experiment 1 was in fact due to the prior statistical learning experience.

Experiment 3C

This experiment consisted of the object-based attention task from Experiment 2, which employed a fixed set of shape pairings for all observers. The goal was to examine whether an object-based attention effect would emerge without the prior statistical learning experience that was provided in Experiment 2.

Methods

Subjects Fifteen undergraduates at University of Iowa completed this experiment for course credits. All reported normal or corrected-to-normal vision.

Stimuli, task, and procedure An object-based attention task identical to the one used in Experiment 2 was presented to observers,

Results and discussion

Here again there was no reliable difference between the response times in the congruent (M = 619 ms, SD = 98 ms) and those in the incongruent condition (M = 618 ms, SD = 94 ms), $t(14) = .04$, $p > .90$, $d < .1$, or between the accuracies in the congruent and the incongruent conditions (96.4 vs. 96.2 %), $t(20) = .35$, $p > .70$, $d = .11$. These results suggest that the object-based attention effect revealed in Experiment 2 was not driven by the attention test itself, but rather by the prior statistical learning experiences.

General discussion

The present experiments demonstrated that task-free viewing of simple visual scenes resulted in learned perceptual co-occurrences that produce object-based attentional effects. Because the co-occurring shapes were not marked by any image-based grouping cues, our findings provide the first evidence that statistical co-occurrences alone are capable of creating perceptual units that serve as the unit of attention. It should be noted that Experiment 3 in Turk-Browne et al. (2005) also provided evidence bearing on this conclusion, although it aimed to answer a different question. In that experiment, observers were first exposed to shapes one at a time in a sequence, in which co-occurring triplets were embedded without perceptual cues that might bind the co-occurring elements. The results showed that in a later implicit test, observers were faster to detect a shape target if it was the second or the third element in an attended co-occurring triplet. However, without validation of that task as object-based, the effect could be also explained in other ways—for example, as anticipation of the next element based on a prior one, even if triplets were not treated as objects. In contrast, the present work addressed the question of objecthood more directly via an object-based attention task, and employed task validation, replication, and controls that together provide strong evidence that statistical co-occurrences alone can create perceptual groupings that serve as the unit of object-based attention.

Our results are important for understanding how familiarity of visual co-occurrences influences object-based attention. Li and Logan (2008) reported that pairs of Chinese characters that formed words could guide attention: when cued to one character of a word, attention spread or shifted to the semantically related character before spreading or shifting to a semantically unrelated character. Although such findings could result from higher-level semantic or lexical information guiding attention, our findings indicate that visual co-occurrences, devoid of any semantic or lexical information, can direct attention in an object-based manner. Importantly, the magnitude of object-based effects is similar between our results (~7 ms difference between the congruent and incongruent conditions) and those reported by Li and Logan (~8 ms difference), although we should note that the stimuli and procedures were very different across these studies, making direct comparisons difficult.

A more general implication of the current findings centers on the Gestalt cues that define perceptual groups and guide attention in typical object-based attention tasks. Historically, many Gestalt psychologists suggested that these perceptual organization processes might be innate or hard wired into the visual system (e.g., Koffka, 1935; Zuckerman & Rock, 1957). However, our experimental results are consistent with the alternative possibility that Gestalt cues are learned through visual experience and possibly reflect the co-occurrences of image elements under different conditions (see Vecera & Palmer, 2006). For example, two image elements or regions that are the same color or luminance (Gestalt similarity) may be more likely to co-occur on the same larger region or object than on different regions or objects. As an object or region moved or otherwise changed, those elements would then be more likely to undergo a similar change (e.g., both would be translated by movement to a similar extent), further reinforcing their pairing or grouping by similarity. Such a view was articulated as early as the 1950s by Brunswick and his colleagues (Brunswick 1956; Brunswick & Kamiya, 1953), who sampled a number of pairs of visual elements in real life images and found that pairs that were from the same object possessed more of the Gestalt perceptual properties (e.g., closeness or similarity) than those pairs from different objects. Based on these findings, Brunswick and his colleagues proposed that the Gestalt cues could be learned via generalized probability learning rather than being innate. Interestingly, our study directly manipulated statistical co-occurrences of visual elements while holding Gestalt cues constant, and demonstrated that statistical co-occurrences alone were able to create groupings that are treated as single objects by attention. Our finding and those of Brunswick and his colleagues, taken together, point to a strong suggestion that there might be nothing special about the Gestalt cues and that they could have been learned through statistical learning. Such an account would fold Gestalt grouping cues into 'empirical' approaches to perception, in which environmental regularities are the driving force behind different visual processes (e.g., Geisler, 2008; Geisler & Perry, 2009; Purves & Lotto, 2011).

Finally, we obtained a negative association between familiarity judgment performance and the magnitude of the object-based attention effect, suggesting that different kinds of learning may occur in parallel from visual statistical learning. One recent fMRI study on visual statistical learning showed that medial temporal brain regions, associated with declarative memory, and the striatum, associated with procedural memory, were both sensitive to statistical structures (Turk-Browne et al., 2009). In addition, ample evidence has indicated that declarative memory and procedural memory are simultaneously involved in various learning tasks and some evidence has suggested these two systems compete with each other (e.g., Poldrak, Clark, Pare-Blagoev, Shohamy, Moyano, Myers, & Gluck 2001; Wagner, Maril, & Schacter, 2000; for a review, see Poldrack & Packard, 2003). Taken together, we can speculate that procedural memory and declarative memory may be involved in visual statistical learning in parallel, and that the former may underlie the attentional benefit of the co-occurring visual elements and the latter may underlie the recognition of these structures. However, as noted by Vickery & Jiang (2009), uncovering the nature of the kinds of learning involved in visual statistical learning, including the learning that supports object-based attention effect, calls for considerably more study, especially investigations that use multiple measures of learning and combine multiple behavioral, neuropsychological and neuroimaging methodologies.

References

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science, 15*(7), 460–466. doi:10.1111/j.0956-7976.2004.00702.x

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Brunswick, E., & Kamiya, I. (1953). Ecological cue-validity of "proximity" and of other Gestalt factors. *The American Journal of Psychology, 66,* 20–32.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics, 74*(5), 784–802.

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

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499–527.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458–467.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology, 59,* 167–192. doi:10.1146/annurev.psych.58.110405.085632

Geisler, W. S., & Perry, J. S. (2009). Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience, 26*(1), 109–121. doi:10.1017/S0952523808080875

Hollingworth, A., Maxcey-Richard, A. M., & Vecera, S. P. (2012). The spatial distribution of attention within and across objects. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 135–151. doi:10.1037/a0024463

Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt Brace.

Li, X. S., & Logan, G. D. (2008). Object-based attention in Chinese readers of Chinese words: Beyond Gestalt principles. *Psychonomic Bulletin & Review, 15,* 945–949.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology.* Cambridge, MA: MIT press.

Palmer, S. E. (2002). Perceptual organization in vision. In H. Pashler (ed.), *Stevens' Handbook of Experimental Psychology*, Sensation and Perception. New York: John Wiley and Sons.

Poldrack, R. A., Clark, J., Pare-Blagoev, J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414,* 546–550.

Poldrack, R. A., & Packard, M. G. (2003). Competition between memory systems: Converging evidence from animal and human studies. *Neuropsychologia, 41,* 245–251.

Purves, D., & Lotto, R. B. (2011). *Why we see what we do redux: A wholly empirical theory of vision*. Sunderland: Sinauer Associates.

Richard, A. M., Lee, H.-K., & Vecera, S. P. (2008). Attentional spreading in object-based attention. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 842–853. doi:10.1037/0096-1523.34.4.842

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn_a_00578

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology. General, 134*(4), 552–564. doi:10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21,* 1934–1945.

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

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General, 123*(2), 146–160. doi:10.1037/0096-3445.123.2.146

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

Vecera, S. P., & Palmer, S. E. (2006). Grounding the figure: Surface attachment influences figure-ground organization. *Psychonomic Bulletin & Review, 13*(4), 563–569.

Vickery, T. J., & Jiang, Y. V. (2009). Associative grouping: Perceptual grouping of shapes by association. *Attention, Perception, & Psychophysics, 71*(4), 896–909.

Wagner, A. D., Maril, A., & Schacter, D. L. (2000). Interactions between forms of memory: When priming hinders new episodic learning. *Journal of Cognitive Neuroscience, 12,* 52–60.

Wertheimer, M. (1923/2000). Laws of organization in perceptual forms, in S. Yantis (eds), *Visual perception: Essential readings*. Philadelphia: Psychology Press. pp. 216–224. (Original work published 1923.).

Zemel, R. S., Behrmann, M., Mozer, M. C., & Bavelier, D. (2002). Experience-dependent perceptual grouping and object-based attention. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 202–217.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science, 24*(5), 667–677. doi:10.1177/0956797612460407

Zuckerman, C. B., & Rock, I. (1957). A reappraisal of the roles of past experience and innate organizing processes in visual perception. *Psychological Bulletin, 54*(4), 269–296.