VISUAL COGNITION, 2010, 18 (1), 82–109

Psychology Press
Taylor & Francis Group

Are objects required for object-files? Roles of segmentation and spatiotemporal continuity in computing object persistence

Tao Gao and Brian J. Scholl
Yale University, New Haven, CT, USA

Two central tasks of visual processing are (1) to segment undifferentiated retinal images into discrete objects, and (2) to represent those objects as the same persisting individuals over time and motion. Here we explore the interaction of these two types of processing in the context of object files—mid-level visual representations that "stick" to moving objects on the basis of spatiotemporal properties. Object files can be revealed by object-specific preview benefits (OSPBs), wherein a "preview" of information on a moving object speeds the recognition of that information at a later point when it appears again on the same object (compared to when it reappears on a different moving object), beyond display-wide priming. Here we explore the degree of segmentation required to establish object files in the first place. Surprisingly, we find that no explicit segmentation is required until after the previews disappear, when using purely motion-defined objects (consisting of random elements on a random background). Moreover, OSPBs are observed in such displays even after moderate (but not long) delays between the offset of the preview information and the onset of the motion. These effects indicate that object files can be established without initial static segmentation cues, so long as there is spatiotemporal continuity between the previews and the eventual appearance of the objects. We also find that top-down strategies can sometimes mimic OSPBs, but that these strategies can be eliminated by novel manipulations. We discuss how these results alter our understanding of the nature of object files, and also why researchers must take care to distinguish "true OSPBs" from "illusory OSPBs".

Keywords: Object files; Segmentation; Motion; Object persistence; Spatiotemporal continuity.

Please address all correspondence to Tao Gao, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, USA. E-mail: Tao.Gao@yale.edu

For helpful conversation and/or comments on earlier drafts, we thank Erik Cheries, Wen-Chi Chiang, Yisheng Dong, Jon Flombaum, Rob Gordon, John Henderson, Josh New, Steve Mitroff, Nic Noles, Mowei Shen, Rende Shui, Yaoda Xu, and an anonymous reviewer. We also thank Yisheng Dong, Mowei Shen, and Rende Shui for assistance with data collection.

http://www.psypress.com/viscog DOI: 10.1080/13506280802614966

The currency of our visual experience is discrete objects: What we see when we view a dynamic scene are things like people and cars, and not merely patches of colour and shape. To make this type of experience possible, two challenges must be met. First, the continuous visual input must be segmented into discrete visual objects. This challenge has been a major focus of research in vision science for decades, involving topics such as image segmentation (for a review see Driver, Davis, Russell, Turatto, & Freeman, 2001), feature binding (Müller, Elliott, Hermann, & Mecklinger, 2001), object type recognition (Peissig & Tarr, 2007), and object-based attention (Scholl, 2001). Second, we must represent objects as the same persisting individuals over time and motion. This challenge has received rather less attention, but has nevertheless been studied in contexts such as apparent motion (e.g., Dawson, 1991; Kolers, 1964), attentive tracking (e.g., Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999), object substitution masking (e.g., Moore & Lleras, 2005), and the tunnel effect (e.g., Flombaum, Kundey, Santos, & Scholl, 2004; Michotte, Thinès, & Crabbé, 1964/1991).

In this paper, we explore how these two challenges interact with each other. In particular, we report several experiments aimed at determining (1) how segmentation processes form representations that persist over time and motion, and (2) how and when objects' surface features can be bound to those persisting object representations.

OBJECT FILES AND OBJECT REVIEWING

One of the most popular ways of theorizing about object persistence is in terms of "object files". An object file is a mid-level visual representation that "sticks" to a moving object over time on the basis of spatiotemporal properties, and stores (and updates) information about that object's surface properties (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992). In this way, object files help to construct our conscious perception of objects and how they behave—e.g., telling us "which went where" or underlying the perception of object persistence despite featural change or momentary periods of occlusion. Object files are thought to underlie object persistence via three steps: (1) A *correspondence* operation, which uses spatiotemporal information for each visual object to determine whether it is novel or whether it moved from a previous location; (2) a *reviewing* operation, which retrieves previously stored object properties (e.g., colour, shape) of those objects; and (3) an *impletion* operation, which uses both current and reviewed information to construct an evolving conscious percept, perhaps of object motion.

The object-file framework has been used to interpret the results of several experimental paradigms in adult visual cognition research (including multiple-object tracking and the tunnel effect; e.g., Flombaum & Scholl, 2006; vanMarle & Scholl, 2003), and such representations are also thought to mediate the representation and processing of object arrays in infants' "object cognition" (e.g., Carey & Xu, 2001; Cheries, Mitroff, Wynn, & Scholl, in press; Feigenson, Carey, & Hauser, 2002; Scholl & Leslie, 1999). Perhaps the most direct evidence for the operation of object files, however, comes from the *object-reviewing* paradigm that was introduced along with a seminal elaboration of the object-file framework itself (Kahneman et al., 1992). Theoretically, this paradigm involves both the "correspondence" and "reviewing" operations. When the features of objects encountered at different times match the correspondence computed by spatiotemporal factors—in other words, when the features are similar across two encounters that are seen as temporal stages of a single enduring object in the world—then certain responses are facilitated. When the features do not match the computed correspondence, in contrast, responses are inhibited.

This basic object-reviewing paradigm is illustrated in Figure 1. In the initial preview display, a number of distinct objects are presented and then letters briefly appear on some or all of them. (Letters are not critical, of course—and in fact we use different symbols in some of the experiments reported here.) After the letters disappear, the objects begin moving about the display. Once they come to rest, a single target letter appears on one of the objects, and the observer's task is simply to name that letter as quickly as possible. This response is typically slightly faster when the letter matches one of the initially presented letters (a type of display-wide priming). However, observers are faster still to name the target letter when it is the same letter that initially appeared on that same object (in a congruent trial), compared to when the final letter initially appeared on a different object (in an incongruent trial). This is termed an *object-specific preview benefit* (OSPB). This relative response-time advantage for congruent trials over incongruent trials is necessarily object-based since the objects' spatial locations change from the preview to the target displays.

Several more recent object reviewing studies have employed a different method wherein the task is to answer as quickly as possible whether the target letter matches *either* of the initial previews. This forces subjects to attend to the preview information (which otherwise can be completely ignored), and perhaps for that reason it tends to produce more robust OSPBs (Kruschke & Fragassi, 1996; Noles, Scholl, & Mitroff, 2005). As such, this is the form of object reviewing that we use in the current study.

Object reviewing has been used in many studies over the past 15 years to explore two primary questions. First, several studies have explored the types of information that can be stored in object files (and thus elicit OSPBs), and the main lesson of this work is that object files are extremely flexible in this

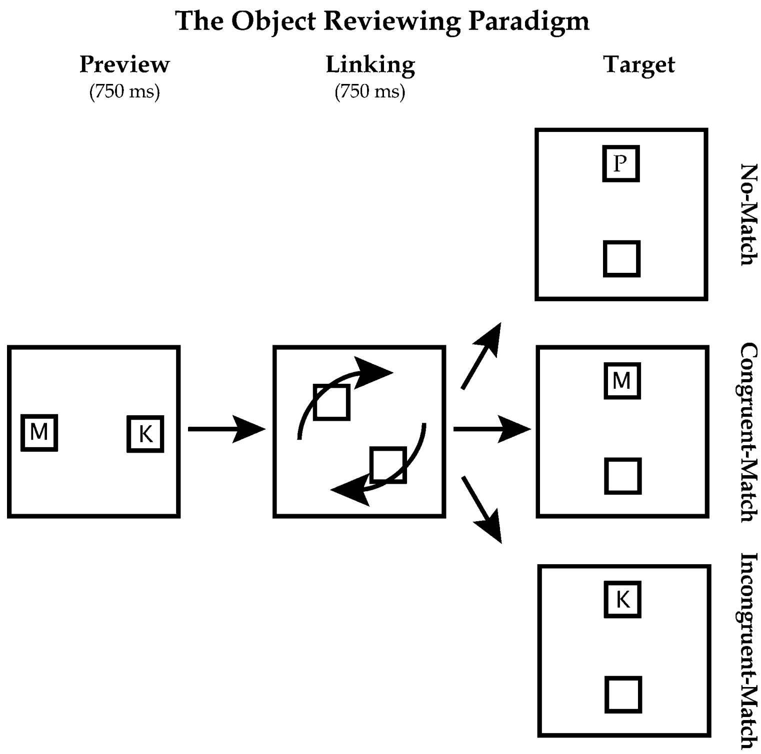

Figure 1. A schematic depiction of the object-reviewing paradigm, as used in the present experiments. Each trial consists of preview, linking, and target displays. The single target can be identical to one of the previews (on match trials) or can be novel to the trial (on no-match trials). On match trials, the target can occur on the same object on which it initially appeared in the preview field (on congruent-match trials) or on the other object (on incongruent-match trials). Observers must press a key to indicate whether or not the target had appeared anywhere in the preview field.

way. For example, object files can store both novel visual features that have never previously been encountered (Mitroff, Scholl, & Noles, 2007), yet also abstract information such as letters independent of their fonts (Henderson, 1994; Henderson & Anes, 1994) or concepts independent of their presentation as words or pictures (Gordon & Irwin, 1996, 2000). Another variant of the object reviewing paradigm even suggests that visual object files can store auditory information (Richardson & Kirkham, 2004). The second primary question that has been addressed in recent object-reviewing experiments concerns the factors that control the maintenance of object files. These studies have demonstrated that object files can persist for at least several seconds (Noles et al., 2005), and that their maintenance is influenced by "core" principles such as cohesion (Mitroff, Scholl, & Wynn, 2004) and perhaps solidity (Mitroff, Scholl, & Wynn, 2005).

THE CURRENT STUDY

Whereas previous studies have focused on the maintenance and contents of object files, here we explore how and when object files are constructed in the first place. In particular, we address a foundational question that has received surprisingly little study: What counts as an "object" in the object file framework? In the limit, of course, many of the same grouping principles that define objects in other situations (e.g., see Feldman, 2007) may influence representations in this paradigm as well. Nevertheless, we may still ask which of these grouping cues are necessary to initially establish object files. Moreover, we can ask how long it takes for explicit segmentation cues to establish object files before features can be stored in them. In this way the current studies explore the interaction between perceptual segmentation and object persistence, two of the major challenges faced by the visual system in the attempt to make sense of the world.

EXPERIMENT 1: REGULAR OBJECTS

Many of the experiments reported in this paper employ objects defined solely by their motion—drawn as random-dot squares on a random-dot background. In order to compare the performance in such conditions with a more typical baseline, however, our first experiment simply measures the OSPBs that arise when using objects with explicit contours (visible in each static frame of motion), drawn on a random-dot background (see Figure 2a).

Method

Participants. Sixteen Yale University undergraduates participated in a single 15-minute session for course credit or payment. The displays were presented on a Macintosh iMac computer using custom software written with the VisionShell graphics libraries (Comtois, 2007). Observers sat without head restraint approximately 50 cm from the monitor. (All measurements here are computed based on this viewing distance.)

Stimuli. The visible background subtended 32°, with each pixel set randomly to black or white. The two objects on each trial were squares subtending 1.71°. The squares were drawn as black outlines with a stroke of 0.2° (such that they were visible on all static frames of motion). Each square's interior was composed of a separately computed random noise pattern that did not change while the objects moved. (In later experiments, this manipulation gave rise to visible motion-defined objects even without the black outlines.) These random-dot interiors were identical for all objects in a trial, but were randomly computed between trials. The squares initially

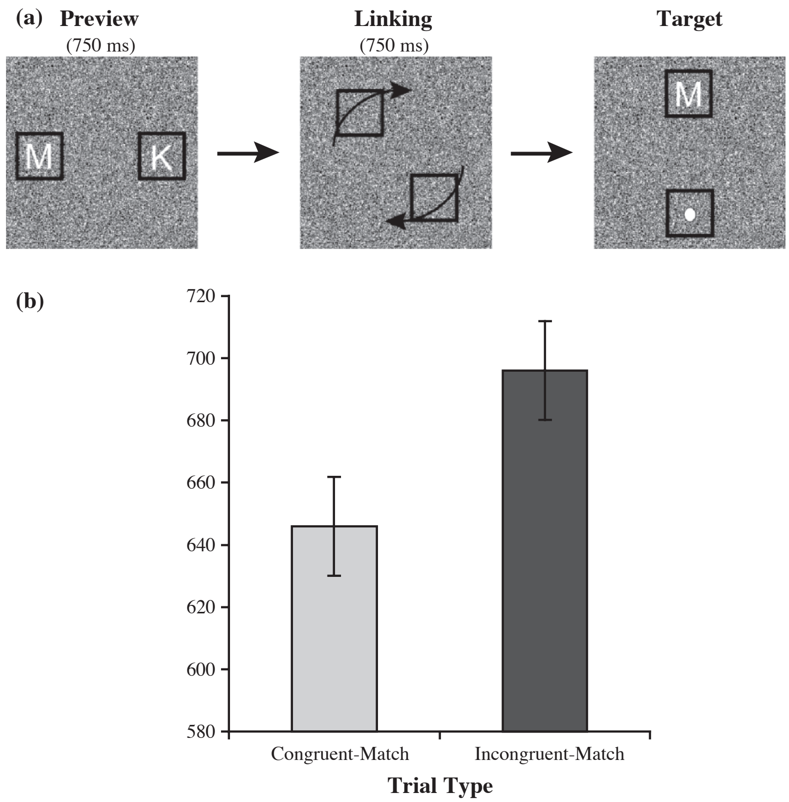

Figure 2. (a) A schematic depiction of a clockwise-motion congruent-match trial from Experiment 1. The objects were initially segmented by the closed black contours, which remained visible throughout the objects' motions. The other location where the target could appear (on other types of trials) is indicated here by a small dot (which was not present in the actual displays). (b) The significant OSPB obtained from Experiment 1. (Error bars depict 95% confidence intervals.)

appeared horizontally aligned in the vertical centre of the display with their centres respectively 5.13° to the right and left of the display's centre. The letters presented on the objects each subtended 0.91°, drawn in a red monospaced font in the centres of the squares, with the particular letters on a given trial drawn without replacement from the set K, M, P, S, T, and V.

Procedure and design. The initial display containing the two squares and the background appeared for 500 ms, after which a letter appeared in each square (in the first panel of Figure 2a). After 750 ms, the two preview letters disappeared, and the squares began their uniform circular motion (either clockwise or counterclockwise, determined randomly on each trial) around the centre of the display (as in the middle panel in Figure 2a). Their motion stopped after 750 ms when they were vertically aligned with their centres 5.13° above and below the centre of the display, respectively.

Immediately after the motion ended, a single target letter appeared in a randomly chosen final square and remained until the observer responded (in the final panel of Figure 2a). We used an adaptation of the object-reviewing paradigm (Kruschke & Fragassi, 1996; Noles et al., 2005): Observers made a speeded response, pressing one key to indicate that the target letter was the same as *either* of the preview letters or another key to indicate that it did not appear in the preview display. Fifty per cent of trials were *no-match trials*, in which the target letter (drawn from the same set as the preview letters) did not appear in either of the original squares. Of the remaining match trials, 50% were *congruent-match trials* (in which the target letter was the same as the preview letter that initially appeared on that square), and 50% were *incongruent-match trials* (in which the target letter was the same as the preview letter that initially appeared on the other square). After 32 practice trials, 160 test trials were presented in a different random order for each observer.

Results and discussion

Overall accuracy was high (96.84%) and did not differ between the congruent-match and incongruent-match conditions (96.25% vs. 95.62%), $t(15) = 0.745$, $p = .468$, $\eta^2 = .036$. All analyses were limited to trials with accurate responses. Trials on which response times fell outside 3 standard deviations from the observer's mean were eliminated (1.6% of the trials). Response times indicated a displaywide priming effect (see Table 1): As in many studies of object reviewing, responses on no-match trials (721 ms) were significantly slower than responses on either congruent-match trials (646 ms), $t(15) = 4.884$, $p < .001$, $\eta^2 = .614$, or incongruent-match trials (696 ms), $t(15) = 2.497$, $p = .025$, $\eta^2 = .294$.[1]

The comparison of interest was the OSPB: faster responses to congruent-match than incongruent-match trials indicated the maintenance of object-specific information, above and beyond displaywide priming. A significant OSPB of 50 ms was observed in this experiment, $t(15) = 4.741$, $p < .001$, $\eta^2 = .600$, as noted in Table 1 and Figure 2b. This result effectively replicates

[1] Such display-wide priming effects are found inconsistently in object-reviewing experiments—sometimes appearing large and robust (e.g., Gordon & Irwin, 1996; Mitroff et al., 2004, 2005) and sometimes being small or nonexistent (Kahneman et al., 1992; Mitroff et al., 2007). These differences appear to depend on the tasks, timings, and stimuli used in particular experiments—and in fact we found such effects only inconsistently even in the experiments reported in this paper. Because no theoretical implications follow from such effects, however, we do not report them in the main text for the remainder of the experiments, though all of the relevant data are presented in Table 1.

TABLE 1
Mean response times (in ms) from the no-match, congruent-match, and incongruent-match conditions for each experiment

Experiment	*No match*	*Congruent match*	*Incongruent match*
1. Regular objects	720.96	646.34	696.25
2. Motion-defined objects	693.70	644.33	674.86
3. Global rotation	684.09	637.93	660.75
4. Global rotation, uncertain positions	775.75	733.25	743.78
5. Motion-defined, uncertain positions	892.05	825.72	879.17
6a. Temporal delay, 100 ms	875.93	792.72	837.49
6b. Temporal delay, 500 ms	878.86	813.27	825.49

the standard OSPB with our modified paradigm, shows that visually complex displays filled with random noise do not greatly inhibit such effects, and serves as a baseline for the following experiments.

EXPERIMENT 2: MOTION-DEFINED OBJECTS

During the motion phase of Experiment 1, the objects were defined in an especially simple and intuitive way—by small closed shapes, which were visible from the very beginning of the trial. Which aspects of those shapes are necessary in order to maintain object files through the motion? In particular, what types of segmentation cues are required to set up the object files in the first place, before the previews appear and the motion begins?

Here we begin our exploration of this question by eliminating all initial static cues to objecthood beyond the preview "features" themselves, defining the objects simply by their motion, as random patterns moving on a random background. As a result, the objects in this experiment were not visible until the motion began—after the offset of the initial previews. This condition is depicted in Figure 3a. However, since the objects were invisible in any given static frame, we have adopted the convention here of drawing visible motion-defined objects in the figures with dashed-line borders, to make them visible. In the animations themselves, of course, these borders were not drawn (for examples of the actual animations used in these studies, see http://www.yale.edu/perception/OF-Segmentation/).

Method

This experiment was identical to Experiment 1, except as noted here. Sixteen observers participated in this experiment; six had also participated in Experiment 1. The objects in this experiment were identical to those in

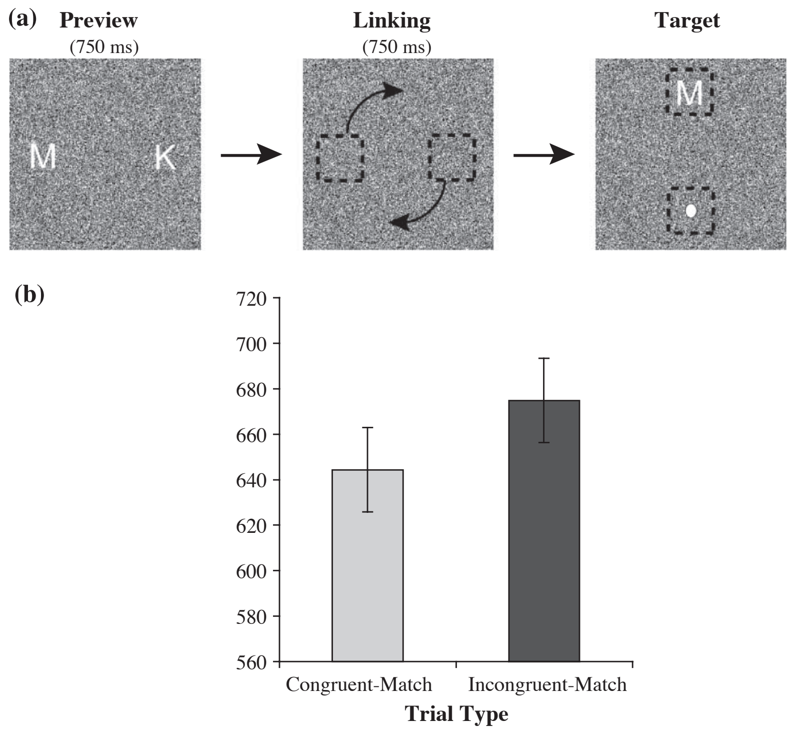

Figure 3. (a) A schematic depiction of a clockwise-motion congruent-match trial from Experiment 2. The objects were defined solely by their motion (as random visual noise patches moving on a random noise background), and were thus invisible during the preview display. Here the objects are indicated by dashed lines, which did not appear in the displays themselves. The other location where the target could appear (on other types of trials) is indicated here by a small dot (which was not present in the actual displays). (b) The significant OSPB obtained from Experiment 2. (Error bars depict 95% confidence intervals.)

Experiment 1, except that the black outlines were not drawn: each object was defined only by its invariant random-noise interior. Because these random-noise shapes were again presented on a random-noise background, they were only visible when the objects moved (as in Figure 3a).

Results

Overall accuracy was high (96.06%) and did not differ between congruent-match and incongruent-match conditions (95.78% vs. 94.06%), $t(15) = 1.842$, $p = .085$, $\eta^2 = .184$. Trials on which response times fell outside 3 standard deviations from the observer's mean were eliminated (1.26% of the trials). A significant OSPB was observed in this experiment, $t(15) = 2.682$, $p = .017$,

$\eta^2 = .324$, as noted in Table 1 and Figure 3b, and the magnitudes of the OSPBs observed in Experiments 1 and 2 did not differ, $t(30) = 1.250$, $p = .221$, $\eta^2 = .050$.

Discussion

This result suggests a striking answer to the question of what segmentation cues are required to define the "objects" of object files in the first place. Whereas some other projects have explored effects of individual surface features such as closure (Mitroff, Arita, & Fleck, 2009), the results of this experiment indicate that *no* initial static surface features are required during the motion for OSPBs: As long as the objects can be perceived (here defined by their coherent motion alone), they can give rise to distinct mid-level visual representations that persist over time and that store information from the preview displays. Perhaps more importantly, note that the objects only became visible in this experiment (due to their motion) *after* the previews had already been removed from the display. This indicates either (1) that the object files were initially formed from the preview stimuli themselves, and/or (2) that the object files that are generated once the motion begins can contain information that was presented (and then disappeared) *before* the object files themselves existed. We discuss these possibilities at greater length in the General Discussion.

EXPERIMENT 3: GLOBAL ROTATION

Previous experiments have explored whether certain cues such as perceptual closure are required in order to set up and maintain object files, as measured by the object reviewing paradigm (e.g., Mitroff et al., 2009). In Experiment 2, however, we removed *all* such cues from the initial displays, and the objects only existed during their motion by virtue of their motion-defined boundaries. Thus Experiment 2 removed all static surface features that could have defined the objects in the first place. In Experiment 3, we attempt to push this kind of effect even further: we still used random-dot patterns to define the objects and the background, but now we removed *all* cues to separate objects during the motion: The background rotated along with the objects, rendering them invisible at all times. Thus observers simply perceived a single global random-dot disc that rotated in place (see Figure 4a).

Whereas there were perfectly visible discrete "objects" in Experiment 2 (after the motion began), there are no discrete objects *per se* in this experiment. Nevertheless, we can still test for OSPBs in exactly the same manner as before, considering the two regions of the global disc (where the previews appeared) as the "objects" for the purpose of the analysis. If

OSPBs are still observed, this would pose a challenge to the conception of object files as truly "object"-based.

Method

This experiment was identical to Experiment 2, except as noted here. Sixteen new observers participated in this experiment. After the offset of the preview letters, the entire background—now drawn as a single disc subtending much of the display (15.16° in diameter)—began to rotate around the centre of the screen (either clockwise or counterclockwise, determined randomly on each trial; see Figure 4a). As a result, the two squares from Experiment 2 remained static relative to the rotating background at all times, and were thus invisible. Indeed, during the motion phase of this experiment, every single static frame looked identical—though the resulting animation was perceived in terms of a rotating disc. The congruent-match and incongruent-match conditions were still defined, however, now in terms of the relative position of the target letter to the invisible objects (i.e., to those local regions of the rotating disc).

Results and discussion

Overall accuracy was high (96.91%) and did not differ between congruent-match and incongruent-match conditions (96.56% vs. 96.41%), $t(15) = 0.151$, $p = .882$, $\eta^2 = .002$. Trials on which response times fell outside 3 standard deviations from the observer's mean were eliminated (1.29% of the trials). A significant OSPB was observed in this experiment, $t(15) = 2.897$, $p = .011$, $\eta^2 = .359$, as noted in Table 1 and Figure 4b. The magnitude of this OSPB was significantly smaller than that in Experiment 1, $t(30) = 2.060$, $p = .048$, $\eta^2 = .124$, but did not differ from that in Experiment 2, $t(30) = 0.557$, $p = .582$, $\eta^2 = .010$.

Whereas the results of Experiment 2 can be accommodated within an object-based framework—indicating that objects can be defined by their motion alone—the results of this experiment cannot. This is true for the simple fact that there were *no discrete objects, in any sense*, present in these displays: there was only a single rotating disc, on which previews appeared. As such, these results seem to challenge the core of the object file framework. If OSPBs are taken to indicate the presence of object-files (which is universally assumed in this literature, by the very logic of the object-reviewing paradigm), then these results seem to indicate that objects are not required for object files. But, since the whole purpose of the object reviewing paradigm is to reveal and explore the construction and maintenance of *object-specific* information in mid-level visual cognition, these results call the

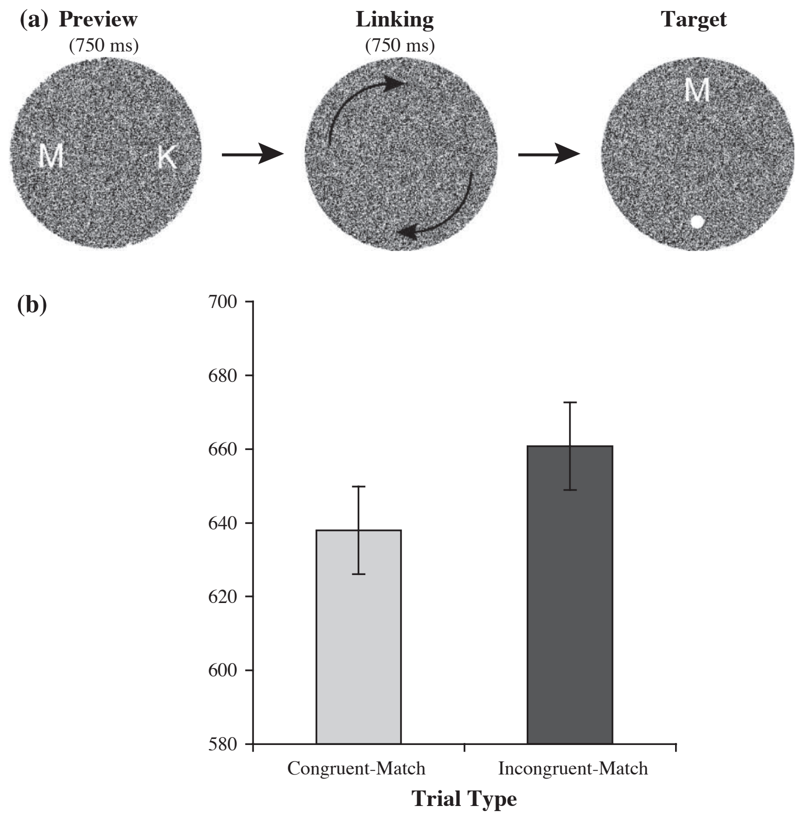

Figure 4. (a) A schematic depiction of a clockwise-motion congruent-match trial from Experiment 3. The entire background disc rotated along with the objects, rendering them invisible throughout the entire trial; observers simply saw a global rotating disc. The other location where the target could appear (on other types of trials) is indicated here by a small dot (which was not present in the actual displays). (b) The significant OSPB obtained from Experiment 3 (later reinterpreted as an ''illusory OSPB''). (Error bars depict 95% confidence intervals.)

entire logic of the paradigm into question. Because of the gravity of this inference, however, we probed this effect more closely in Experiment 4, in an attempt to test an alternate explanation that would explain these results in a way which maintained the ''object-specific'' nature of object files.

EXPERIMENT 4: GLOBAL ROTATION WITH UNCERTAIN TARGET POSITIONS

If object files truly require objects, then why did the object-less displays in Experiment 3 yield robust OSPBs? One possibility is that this effect was driven not by the automatic object-specific integration of information in mid-level visual processing, but rather by a higher-level strategy that combined (1) prioritized encoding of one of the two previews with (2) strategic scanning along with the rotary motion.

Consider what it is like to experience a trial of the previous experiment: Two previews appear and then disappear, and then they must be compared with a final target. Assume first that in such situations observers must scan the two previews to encode them, by attending them or even overtly looking at them in sequence during the 750 ms presentation interval. This means that the eyes (or perhaps just the locus of attention and encoding) will end up at one of the two preview locations at the end of this period, and that the representation of that preview item may be especially salient, having just been encoded.

Even without any objects, however, observers know (all too well, after so many trials!) that the target never appears in the same location where either of the previews appeared: Instead, the target reliably appears in an orthogonal position (at the top or bottom of the display, whereas the previews appear along the left or right of the display). Thus, even without any segmentation cues, observers must move their eyes and/or their attention from the location of the most recently encoded preview item to one of the two possible target locations. But which one? Observers could simply wait at the location of the most recently encoded preview item until the target actually appears, and then move their eyes and/or their attention to this new location. But this is not a natural strategy, given the salience of the intervening rotary motion—and it may be especially natural to simply move their eyes and/or attention along with that motion, which will take them directly to one of the two possible target locations. But now consider what will happen if the target appears in that location: Observers will still have the most recently encoded preview item prioritized in memory, and may thus be especially fast to respond when it matches the target. Put more directly: Observers may simply be slightly faster when the same item appears in quick succession right where they are looking (or attending)—something that may occur on a considerable number of trials in this situation, and will yield a pattern of results that looks exactly like an OSPB.

In this experiment, we tested this possibility in the most direct way possible: We simply replicated Experiment 3, but we doubled the number of trials, and allowed the target to appear on the same location as one of the two previews on half of the trials (see Figure 5a). These trials themselves cannot be analysed in terms of OSPBs, of course, since objects and locations are perfectly confounded here, and so they are set aside from the primary analysis. However, if the significant result observed in Experiment 3 was strategic in nature, then this manipulation should eliminate it even on the remaining trials where the target appears in a new location: Now there should be no incentive to move attention away from the location of most recently encoded preview item, since the target is just as likely to appear in that same location as anywhere else. And, since there are no discrete objects visible at any time, this display should not generate an OSPB driven by automatic object-specific encoding.

Method

This experiment was identical to Experiment 3, except as noted here. Sixteen new observers participated in this experiment (which took roughly twice as long as Experiment 3, given that there were twice as many trials). To minimize possible effects of overt rehearsal of preview and target letters, and thus

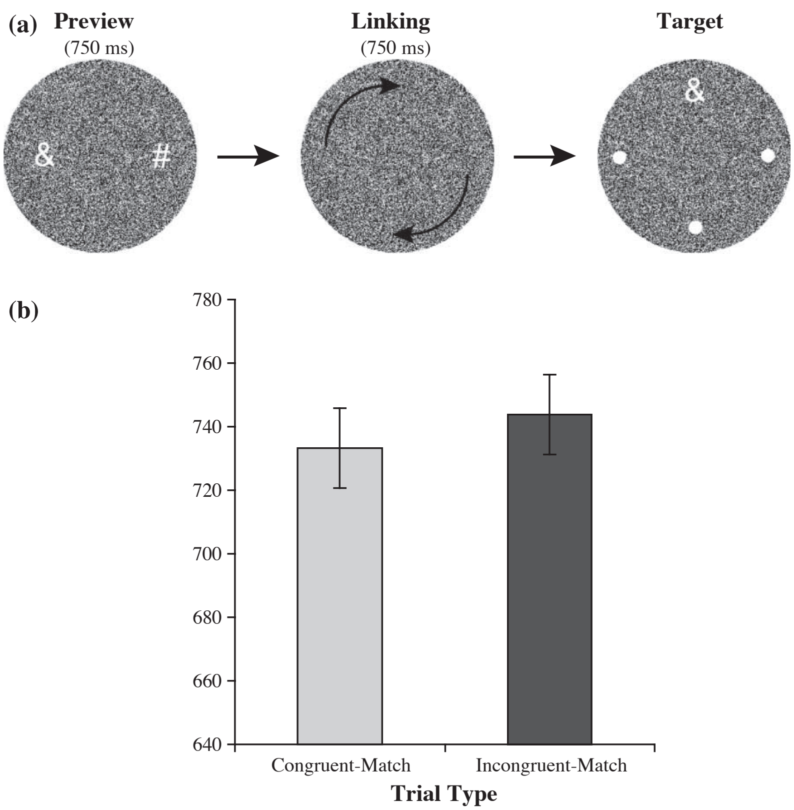

Figure 5. (a) A schematic depiction of a clockwise-motion congruent-match trial from Experiment 4. This experiment was identical to Experiment 3 (with a global rotating disc), except that there were now more locations where the target could appear (on other types of trials)—indicated here by small dots (which was not present in the actual displays). (b) The null effect observed in Experiment 4. (Error bars depict 95% confidence intervals.)

maximize our chances of observing OSPBs, this and all subsequent experiments employed symbols rather than letters: The two previews were randomly selected from a pool of six possible choices ($, &, +, #, =, %). The only other change in the procedure was that the number of trials was doubled, and the target on the new 160 trials appeared at one of the two preview locations (randomly chosen). The resulting 320 test trials were presented in a different random order for each observer.

Results and discussion

The additional 160 trials that were added to this experiment cannot be analysed in terms of OSPBs, since spatial location was perfectly confounded with persisting objecthood.[2] The remaining 160 trials, however, were analysed as before. Overall accuracy was high (95.55%) and did not differ between congruent-match and incongruent-match conditions (94.69% vs. 94.84%), $t(15) = 0.131$, $p = .898$, $\eta^2 = .001$. Trials on which response times fell outside 3 standard deviations from the observer's mean were eliminated (1.47% of the trials). No significant OSPB was observed, as noted in Table 1 and Figure 5b (733 vs. 744 ms for congruent-match and incongruent-match trials, respectively), $t(15) = 1.266$, $p = .225$, $\eta^2 = .097$. (This null effect cannot be properly compared with the OSPB from Experiment 3, however, given that observers experienced twice the number of overall trials, and that we began using symbols instead of letters for the previews and targets. However, note that we are able to conduct this type of comparison with a better-matched OSPB in one of the following experiments.)

These results are consistent with the "strategic" explanation of the results of Experiment 3 that was suggested above. Simply adding in two new possible target locations in this experiment eliminated the OSPB. There is no reason to expect that this manipulation would affect the computation of

[2] Of course, we can analyse these "spatial priming" trials in a similar way to see whether there is a *location*-specific preview benefit. No theoretical implications follow for the issues being explored in this paper in either case, but we did run this analysis for the sake of completeness. Here, no-match trials are defined as before; congruent-match trials are those in which the target matched the preview that appeared in that same *location*; and incongruent-match trials are those in which the target matched the preview that appeared in the other preview location. Overall accuracy in these spatial priming trials was high (94.92%), though accuracy was marginally higher on congruent-match versus incongruent-match trials (95.31% vs. 93.28%), $t(15) = 1.932$, $p = .072$, $\eta^2 = .199$. Trials on which response times fell outside 3 standard deviations from the observer's mean were eliminated (1.13% of the trials), along with incorrect trials. No significant spatial priming effect was observed by comparing the RTs from the congruent-match versus incongruent-match condition (719 vs. 734 ms, respectively), $t(15) = 1.738$, $p = .103$, $\eta^2 = .168$. The spatial-priming results of Experiments 5 and 6 were also analysed in the same manner, and all of these results are reported in Table 2.

TABLE 2
Mean response times (in ms) and accuracy (in parentheses) from the no-match, congruent-match, and incongruent-match conditions for spatial priming trials in Experiments 4–6

Experiment	*No match*	*Congruent match*	*Incongruent match*
4. Global rotation, uncertain positions	768.17 (95.55%)	718.96 (95.31%)	734.10 (93.28%)
5. Motion-defined, uncertain positions	875.77 (95.31%)	865.10 (95.63%)	873.60 (96.25%)
6a. Temporal delay, 100 ms	868.09 (94.95%)	803.31 (95.31%)	799.34 (93.33%)
6b. Temporal delay, 500 ms	867.09 (94.95%)	803.08 (94.79%)	798.03 (93.85%)

object persistence in mid-level visual cognition, but we had a compelling reason to think that this would affect observers' strategies. Thus, we tentatively conclude that the significant effect observed in Experiment 3 might reflect only an "illusory OSPB"—and, as such, that those results may not in fact impugn the object-specific nature of object files.

As discussed later, however, this possibility calls into question some other results that have been obtained with the object reviewing technique, and suggests that researchers who employ this tool should always attempt—in the way used here or some other way—to separate out *true* OSPBs (i.e., those that are fuelled by automatic mid-level visual processing) from *illusory* OSPBs (i.e., those that may be fuelled only by observers' circumstantial strategies).

EXPERIMENT 5: MOTION-DEFINED OBJECTS WITH UNCERTAIN TARGET POSITIONS

The results of the previous experiment show that when strategic effects are well controlled, OSPBs are not observed with displays containing only a single global rotating object—thus supporting the object-specific nature of the "object file" framework. However, this experiment also raised an even more dire concern: What if *all* OSPBs reflected the kinds of strategic processing discussed in the previous experiment, rather than more automatic aspects of mid-level visual processing? In order to explain away the results of Experiment 3 in terms of strategic effects ("illusory OSPBs"), while retaining the results of Experiment 2 as "true OSPBs", we must assume that the latter reflected nonstrategic processing. Here we test this directly, in the same manner as in Experiment 4: We retain the novel manipulation wherein the target can appear in either the original or final positions of either object, but now we test this manipulation with discrete motion-defined objects (as in Experiment 2). This manipulation (see Figure 6a) destroyed the apparent OSPB with the rotating disc, but we predict no such effect here: When objects are present, object-specific processes should be invoked, which do not depend on observers' strategies.

Method

This experiment was identical to Experiment 4, except that the rotating background disc from Experiment 3 was replaced with the static random-dot background from Experiment 2 (such that the discrete objects were again visible due to their motion). Twelve new observers participated in this experiment.

Results and discussion

Our primary analysis was again conducted only over those 160 trials wherein the target was not presented at one of the original preview locations. Overall accuracy was high (94.27%) and did not differ between congruent-match and incongruent-match conditions (93.54% vs. 92.92%), $t(11) = 0.376$, $p = .714$, $\eta^2 = .013$. Trials on which response times fell outside 3 standard deviations from the observer's mean were eliminated (1.28% of the trials) along with incorrect trials. A significant OSPB was observed, as noted in Table 1 and Figure 6b (826 ms vs. 879 ms, for congruent-match and incongruent-match trials, respectively), $t(11) = 2.698$, $p = .021$, $\eta^2 = .398$. The magnitude of this OSPB was considerably larger than the nonsignificant effect observed in Experiment 4, $t(26) = 2.191$, $p = .038$, $\eta^2 = .156$. (Note also that this OSPB was even larger than that from Experiment 2—perhaps because the use of symbols rather than letters inhibits types of overt rehearsal that can add noise to the underlying OSPBs.)

These results, when combined with those of the previous experiment, demonstrate that subtle manipulations of where targets may appear can dramatically affect the resulting (illusory) OSPBs in objectless displays, but do not affect OSPBs that arise from segmented displays with discrete objects. We thus conclude that the OSPBs observed in segmented displays reflect true object-specific processing, but that the apparent OSPB in the objectless display of Experiment 3 reflects only higher level strategic effects.

EXPERIMENT 6: INITIAL OBJECT SEGMENTATION VIA THE PREVIEWS THEMSELVES?

One striking result of this study so far is that of Experiments 2 and 5, wherein OSPBs are observed even when the objects themselves do not appear until the offset of the previews. This raises the possibility that the initial presentation of the preview letters themselves establish the object files, which are then later bound to the moving objects which subsequently appear. After all, although the previews have typically been treated in this

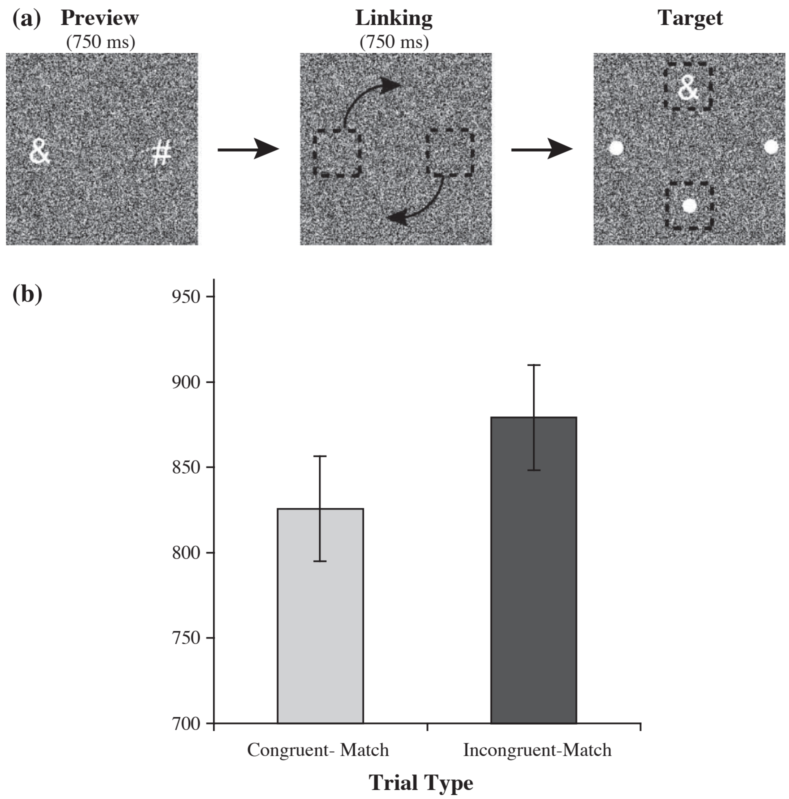

Figure 6. (a) A schematic depiction of a clockwise-motion congruent-match trial from Experiment 5. This experiment was identical to Experiment 2 (with motion-defined objects), except that there were now more locations where the target could appear (on other types of trials)—indicated here by small dots (which was not present in the actual displays). (b) The significant OSPB obtained from Experiment 4. (Error bars depict 95% confidence intervals.)

paradigm as the "features" that are stored *in* object files, they are also local shapes (and thus objects) on their own.

Note, however, there was no temporal gap between these two events in the previous experiments: the objects appeared (via their motion) immediately as the previews offset. Thus it is possible that the preview information was still decaying on the display itself, and was thus visible as the motion began—thus directly facilitating spatiotemporal integration between the previews and the moving objects. In this experiment, however, we introduced delays between the offset of the previews and the onset of the motion (and thus the appearance of the "objects"; see Figure 7a). If OSPBs should persist despite these brief pauses, this would lend support to the idea that the previews themselves can establish object files, before the appearance of the moving boxes, which are then integrated into the same evolving representations.

Method

This experiment was identical to Experiment 5, except as noted here. Twenty-four Zhejiang University undergraduates participated in a single 50-minute session for payment. The displays were presented on a PC computer using custom software written with the Psychotoolbox graphics libraries and Matlab (Brainard, 1997; Pelli, 1997). The number of trials was doubled. On 50% of the trials, the delay between the offset of the preview and the onset of the motion was 100 ms; on the remaining trials, this preview-motion ISI was 500 ms. After 32 practice trials (the result of which were not recorded), the 640 test trials were presented in a different random order for each observer.

Results

Our primary analysis was conducted only over those 320 trials wherein the target was not presented at one of the original preview locations. Overall accuracy was high and did not differ between congruent-match and incongruent-match conditions for either the 100-ms preview-motion ISI ($M = 94.92\%$; 94.69% vs. 94.27%), $t(23) = 0.397$, $p = .695$, $\eta^2 = .007$, or the 500-ms preview-motion ISI ($M = 94.18\%$; 93.23% vs. 93.33%), $t(23) = 0.096$, $p = .924$, $\eta^2 = .001$. Trials on which response times fell outside 3 standard deviations from the observer's mean were eliminated (1.36% of the trials for the 100-ms ISI, 1.35% for the 500% ISI) along with incorrect trials. A significant OSPB was observed, as noted in Table 1 and Figure 7b, for the 100-ms ISI trials (793 ms vs. 837 ms, for congruent-match and incongruent-match trials, respectively), $t(23) = 6.013$, $p < .001$, $\eta^2 = .611$, but not for the 500-ms ISI trials (813 ms vs. 825 ms), $t(23) = 1.073$, $p = .294$, $\eta^2 = .048$—and these two magnitudes were significantly different, $t(23) = 2.267$, $p = .033$, $\eta^2 = .183$. Moreover, the OSPB magnitude in the 100-ms ISI condition did not differ from that obtained with the 0-ms ISI using these same methods in Experiment 5, $t(34) = 0.498$, $p = .621$, $\eta^2 = .007$.

Discussion

The significant OSPB observed in the 100-ms preview-motion ISI condition, unlike the 0-ms ISI tested in Experiments 2 and 5, cannot be explained by appeal to any type of visual persistence of the preview information.

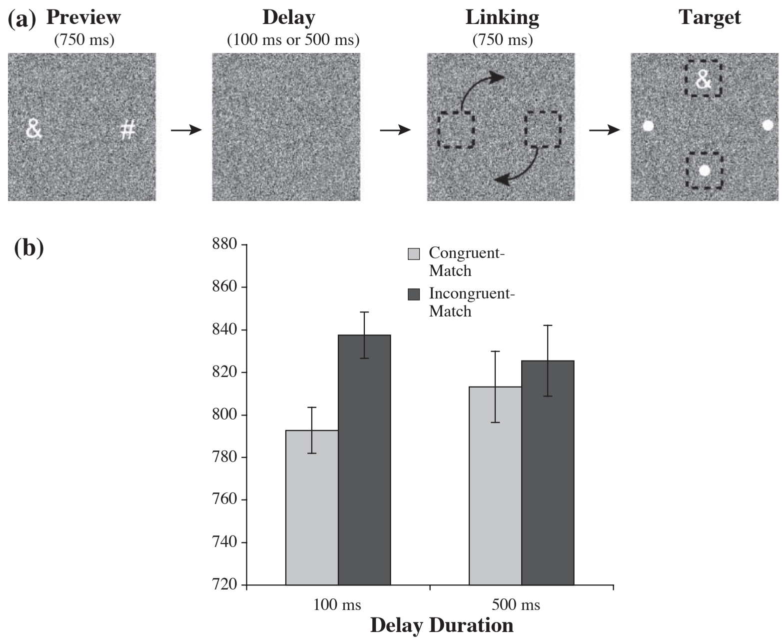

Figure 7. (a) A schematic depiction of a clockwise-motion congruent-match trial from Experiment 6. This experiment was identical to Experiment 5 (with motion-defined objects), except that there was now a pause of either 100 ms or 500 ms between the offset of the previews and the onset of the motion. As in Experiments 4 and 5, the target could appear in multiple potential locations (on other types of trials)—indicated here by small dots (which was not present in the actual displays). (b) The results of Experiment 6, consisting of the significant OSPB observed in the 100 ms delay condition, and the null effect observed in the 500 ms delay condition. (Error bars depict 95% confidence intervals.)

Rather, this result indicates that the preview letters themselves were able to establish the object files, and that the moving boxes that later appeared in those same locations were then treated as the subsequent stages of those same (preview-)objects. Such spatiotemporal integration is perhaps not surprising given that the boxes enjoyed excellent spatiotemporal continuity with the previews, appearing as they did moments later and in the same location. Spatiotemporal integration does have a temporal component, however, and accordingly the null effect observed with the 500-ms ISI suggests that the boxes and the previews will only be integrated in this way when there is not a sizeable temporal gap to separate them. Note also that the robust 100-ms ISI effect was observed with the same degree of statistical power (and, indeed, in the same observers, with intermixed trials) as for the 500-ms ISI, effectively ruling out any strategic explanation of these results—since it is difficult to imagine an overt rehearsal strategy that could not accommodate a 500 ms delay.[3]

GENERAL DISCUSSION

The experiments reported here explored the interaction of segmentation and persistence in the object file framework, using the object reviewing paradigm, with three primary manipulations: (1) What sorts of objects were displayed, (2) when they were displayed, and (3) where the targets could appear. Corresponding to these manipulations, we observed three primary results. First, no static surface features at all beyond the previews themselves are required to establish object files and to generate OSPBs. Second, the letters and symbols that have been treated as "surface features" in this paradigm can be imported into object files (yielding robust OSPBs) even when those previews disappear 100 ms (but not 500 ms) before the moving objects themselves first appear—thus indicating that the previews alone are able to establish object files, to which the moving boxes are then bound via spatiotemporal continuity. Third, we discovered that the potential number and locations of targets can dramatically influence the resulting OSPBs in some contexts but not others—a result we interpret in terms of a new distinction between automatic "true" OSPBs and strategic "illusory" OSPBs.

In the remainder of this paper, we discuss the implications of each of these results in turn.

Are objects required for object files?

What counts as an object? This question may not have a single answer in the context of real objects in the world (Marr, 1982), but it may have a

[3] By the same token, note that this contrast effectively rules out another potential alternate interpretation of OSPBs. Earlier, we noted that a combination of discrete moving objects plus highly predictable probe locations could effectively produce "illusory OSPBs". One might also wonder, though, whether such effects could arise due to strategic tracking of discrete moving objects even when the probe locations are harder to predict (as in Experiments 5 and 6). This possibility has not been previously considered in the context of object reviewing, but similar ideas have been tested in the context of transsaccadic integration (Gajewski & Henderson, 2005; Gordon, Vollmer, & Frankl, 2008). However, this possibility is difficult to reconcile with the present experiment, since there is no reason why the ultimately small difference between 100 ms and 500 ms should make a difference, if OSPBs are simply driven by strategic oculomotor tracking of the objects: Such tracking should be just as natural (or not) in either case, yet only one of them yields an OSPB. More generally, this possibility would predict that "OSPBs" would always arise whenever there are readily "trackable" displays in which two discrete objects each move to a new location—but in fact only a subset of such studies have revealed reliable OSPBs, as in the many previous object reviewing studies cited in the introduction.

well-defined answer (or answers) in the context of visual representations. For example, this issue has been the focus of a considerable amount of research in the study of object-based attention. Seminal work by many laboratories (e.g., Duncan, 1984; Egly, Driver, & Rafal, 1994) indicated that attention does not operate in the manner of a spatial spotlight, but that it is influenced by objects' boundaries. In particular, these studies and many others have demonstrated *same-object advantages*: Across several methodological variants, attending to two features of a single object is easier and faster than attending to two features of two different objects (for a review see Scholl, 2001). Thus attention is (in some cases necessarily) "object based". This conclusion immediately raises the question of what can count as an "object" of attention. Accordingly, many additional studies have now explored the features that are necessary and sufficient for same-object advantages—including connectedness (Scholl, Pylyshyn, & Feldman, 2001; Watson & Kramer, 1999), closure (Avrahami, 1999; Marino & Scholl, 2005), symmetry (Saiki, 2000), curvature (Barenholtz & Feldman, 2003; Ben-Shahar, Scholl, & Zucker, 2007), and several other features (see Feldman, 2007). Perhaps the central lesson of this work is just that such features matter: Even though the resulting definition of "objecthood" may not always match up with our intuitions (see Ben-Shahar et al., 2007), it is clear that attention—and same-object advantages—are acutely dependent on the individual segmentation and grouping cues that collectively parse the visual world into discrete units.

Studies of "object"-specific preview benefits have similarly motivated the question of what counts as an object in this context. One might assume that the answers would be identical to those discussed earlier: The "objects" that fuel same-object advantages might be the same "objects" that give rise to OSPBs. However, the few studies of such questions using object reviewing have suggested that this is not so. Indeed, whereas the central lesson of work on object-based attention is that various segmentation and grouping cues matter a great deal, the central lesson of work on object-reviewing is that such cues often barely matter at all.

Whereas most traditional object-reviewing studies limited their "objects" to intuitive rectangles and discs, two recent studies have explored a wider array of possible cues. In one of these studies (Gao, Shen, & Dong, 2008), robust OSPBs were obtained regardless of whether the objects enjoyed properties such as closure, or connectedness—arising even from sparse dot clusters with no grouping cues at all. Moreover, this study demonstrated that the very same stimuli which dramatically affect same-object advantages (e.g., the "box and line" used by Duncan, 1984) made no difference whatsoever in object reviewing. Another recent study focused on the role of closure as a cue to define objects, and similarly found that under some circumstances (especially when using blocked conditions), OSPBs can arise without any closed contours, and even when the critical surface features (i.e., the letters, as in most studies) occurred outside of closed contours (Mitroff et al., 2009).

The present results similarly support the conclusion that object reviewing is largely insensitive to the cues that define objects in other contexts. Indeed, in one way the current studies extend such manipulations to their limit: we observed that no static cues at all beyond the "features" that constitute the previews themselves are required to establish object files, since OSPBs were reliably observed (and were not even any weaker) even when the objects were composed of pure random noise, moving on a background of random noise. These types of motion-defined objects thus demonstrate that the objects of OSPBs can be constructed "on the fly" without initial segmentation cues beyond the appearance of the previews themselves, so long as the those previews are spatiotemporally continuous with the moving motion-defined objects that subsequently appear.

At the same time, the current results place an important new limit on the question of what can count as an object of object reviewing, and demonstrate that there is still an important sense in which this processing is truly "object" based. It would be consistent with all previous studies to suggest a deflationary answer: Maybe OSPBs would arise without any cues at all to discrete objects at any time in the display. We tested this possibility in Experiments 3 and 4, and discovered that "true" OSPBs (as discussed in more detail later) do require discrete individuals during the motion phase: motion-defined squares (as in Figure 6) yield robust OSPBs, but the same positions on a globally rotating disc (as in Figure 5) do not. Thus it appears that mid-level visual processing does involve object-specific tracking, but it does not involve "region"-specific tracking in unsegmented displays.

The role of spatiotemporal continuity in maintaining object files

In general, two very different types of information may be used to identify objects as the same individuals over time. First, we can simply note what the objects look like: A glimpsed object that is flat and blue is unlikely to be the same individual as a later glimpsed object that is tall and red. Second, we can identify objects as the same based on how and where they move through the environment, employing a principle of spatiotemporal continuity: There must be a continuous spatiotemporal path between two objects in order for them to be treated as the same individual. (This is true even in neural processing, wherein ventral cortex requires spatiotemporal continuity in order to treat subsequent faces as the same, as revealed by repetition attenuation; see Yi et al., 2008.) In practice, however, these two types of information are not equally weighted: Across many different paradigms, spatiotemporal information is found to trump surface features, such that two shapes may be reflexively interpreted as subsequent stages of the same persisting individual even when they look entirely different (for a review see Flombaum, Scholl, & Santos, in press). This is true, for example, in apparent motion (e.g., Burt & Sperling, 1981; Navon, 1976), multiple object tracking (Scholl, Pylyshyn, & Franconeri, 1999), and the tunnel effect (Flombaum et al., 2004).

The present study suggests that the same may hold for object files in the context of the object reviewing paradigm. The fact that we observed robust OSPBs even when the previews disappeared before the onset of the moving boxes (especially in the 100-ms ISI condition of Experiment 6) suggests that the previews themselves must have been sufficient to establish the object files. The spatiotemporal continuity between those symbols and the motion-defined boxes was then apparently sufficient to bind these entities into the same enduring object representations, despite the vast featural differences between the red contours of the preview symbols and the colourless motion-defined squares. This is consistent with the view that surface features may be stored *in* object files, but only spatiotemporal information suffices to direct their maintenance over time (cf. Mitroff & Alvarez, 2007). Previous studies have explored the constraints on spatiotemporal integration, focused on principles of cohesion (Mitroff et al., 2004) and solidity (Mitroff et al., 2005). The present results focus instead on the temporal component of spatiotemporal *continuity*, suggesting that the object files established by the preview letters will decay at some point between 100 and 500 ms, as observed in Experiment 6. This is consistent with previous studies, especially of apparent motion, which is observed (even between featurally different shapes) with 100 ms but not 500 ms delays between two flashes.

This discussion also highlights one of the reasons to expect that the "objects" of object reviewing and object-based attention would differ, in that they prioritize different types of features. Both processes are likely to operate over the same visual object representations in some sense, of course: after all, both of these processes prioritize discrete objects, and it seems unlikely that the initial segmentation processes that serve to individuate these objects would be different in each case. However, at the same time, it seems natural to expect that processes related to attention versus object-files would *do different things with those object representations*, since they are thought to serve rather different purposes. Attention prioritizes objects for further processing, and some of this processing intrinsically involves visual features—for example binding them into particular arrangements within individual objects. The object-file system, in contrast, serves a different purpose: It is primarily concerned with maintaining object *persistence*, and in such cases spatiotemporal cues are dominant. (For discussion of just why spatiotemporal cues trump surface features during computations of persistence across time, see Flombaum et al., in press.)

Improving the object reviewing paradigm

Beyond its theoretical implications, this study also has methodological implications for how we should test for OSPBs. In all previous studies of object reviewing of which we are aware, the target could only ever appear in one of two locations, corresponding to the two final locations of the moving objects. The results of Experiments 4 and 5, however, suggest that this common feature of the object reviewing paradigm in practice is problematic. When targets could appear in only these two typical positions, we found that OSPBs were observed even without any cues to discrete objects at any point in the trial. However, when we allowed targets to appear at four possible locations—the two final locations of the moving objects (as usual), but also their initial locations where the previews were presented—then OSPBs remain robust for displays with moving objects, but were eliminated in the "objectless" displays.

There is no reason to expect that this manipulation of target position across trials would affect the computation of object persistence via automatic visual processing. Indeed, we know of no such mechanism by which such processing *could* be affected in this manner. However, such manipulations could very well affect observers' higher level strategies for keeping track of the preview information—for example in the ways explored at length in the introduction to Experiment 4. We can interpret these results in terms of the need to distinguish true OSPBs that arise from actual mid-level visual processing from similar effects—illusory OSPBs—that may mimic this same pattern due only to strategic scanning and rehearsal. After recognizing the possibility of illusory OSPBs, it of course becomes extremely important to evaluate which OSPBs are "true".

We have proposed that in most object reviewing designs, this can be done—i.e., strategic effects can be thwarted—simply by allowing targets to appear at the initial preview locations. (This effectively serves to remove any incentive for strategic scanning away from these initial locations.) In our studies, this manipulation made a dramatic difference indeed, leading us to conclude that OSPBs were indeed object based rather than merely reflecting some sort of "region-based" processing that required no segmentation and individuation. We thus commend this manipulation to others, as what we hope will become an essential part of object-reviewing experiments—and of course we must also now test whether previously reported OSPBs are true or merely illusory in this way.

CONCLUSIONS

This paper began with an attempt to relate two of the primary challenges of object perception: (1) Segmenting undifferentiated retinal images into discrete objects, and then (2) representing those objects as the same persisting individuals over time and motion. In some ways the results of the studies reported here support a connection between these two challenges—for example, by demonstrating that object files really do require discrete objects. In other ways, however, our results emphasize the distinction between the representations that help meet these two challenges—for example by demonstrating that object files can be maintained despite radical featural differences between the initial preview information that establishes the object files in these displays and the motion-defined boxes that are later bound to them.

REFERENCES

Avrahami, J. (1999). Objects of attention, objects of perception. *Perception and Psychophysics*, *61*, 1604–1612.

Barenholtz, E., & Feldman, J. (2003). Perceptual comparisons within and between object parts: Evidence for a single-object superiority effect. *Vision Research*, *43*, 1655–1666.

Ben-Shahar, O., Scholl, B. J., & Zucker, S. W. (2007). Attention, segregation, and textons: Bridging the gap between object-based attention and texton-based segregation. *Vision Research*, *47*, 845–860.

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

Burt, P., & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, *88*, 171–195.

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object-files and object tracking. *Cognition*, *80*, 179–213.

Cheries, E. R., Mitroff, S. R., Wynn, K., & Scholl, B. J. (in press). Do the same principles constrain persisting object representations in infant cognition and adult perception? The cases of continuity and cohesion. In B. Hood & L. Santos (Eds.), *The origins of object knowledge*. Oxford, UK: Oxford University Press.

Comtois, R. (2007). VisionShell PPC [Software]. Watertown, MA: Author.

Dawson, M. R. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review*, *98*, 569–603.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention, and phenomenal visual objects. *Cognition*, *80*, 61–95.

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

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

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files vs. analog magnitudes. *Psychological Science*, *13*, 150–156.

Feldman, J. (2007). Formation of visual "objects" in the early computation of spatial relations. *Perception and Psychophysics*, *69*, 816–827.

Flombaum, J. I., Kundey, S., Santos, L. R., & Scholl, B. J. (2004). Dynamic object individuation in rhesus macaques: A study of the tunnel effect. *Psychological Science*, *15*, 795–800.

Flombaum, J. I., & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 840–853.

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (in press). Spatiotemporal priority as a fundamental principle of object persistence. In B. Hood & L. Santos (Eds.), *The origins of object knowledge*. Oxford, UK: Oxford University Press.

Gajewski, D. A., & Henderson, J. M. (2005). The role of saccade targeting in the transsaccadic integration of object types and tokens. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 820–830.

Gao, T., Shen, M., & Dong, Y. (2008). What counts as an object (file)? Grouping cues and object persistence. *Manuscript in preparation*.

Gordon, R. D., & Irwin, D. E. (1996). What's in an object file? Evidence from priming studies. *Perception and Psychophysics*, *58*, 1260–1277.

Gordon, R. D., & Irwin, D. E. (2000). The role of physical and conceptual properties in preserving object continuity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 136–150.

Gordon, R. D., Vollmer, S. D., & Frankl, M. L. (2008). Object continuity and the transsaccadic representation of form. *Perception and Psychophysics*, *70*, 667–679.

Henderson, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General*, *123*, 410–426.

Henderson, J. M., & Anes, M. D. (1994). Roles of object file review and type priming in visual identification within and across eye fixations. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 826–839.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). Orlando, FL: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.

Kolers, P. A. (1964). The illusion of movement. *The Scientific American*, *211*, 98–106.

Kruschke, J. K., & Fragassi, M. M. (1996). The perception of causality: Feature binding in interacting objects. In G. W. Cottrell (Ed.), *Proceedings of the 18th annual conference of the Cognitive Science Society* (pp. 441–446). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Marino, A. C., & Scholl, B. J. (2005). The role of closure in defining the "objects" of object-based attention. *Perception and Psychophysics*, *67*, 1140–1149.

Marr, D. (1982). *Vision*. New York: W. H. Freeman.

Michotte, A., Thinès, G., & Crabbé, G. (1964). Les complements amodaux des structures perceptives. In *Studia Psychologica*. Louvain: Publications Universitaires. Reprinted and translated 1991 as: A. Michotte, G., Thinès, & G. Crabbé. Amodal completion of perceptual structures. In G. Thines, A. Costall, & G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception* (pp. 140–167). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, underlie object persistence. *Psychonomic Bulletin and Review*, *14*, 1199–1204.

Mitroff, S. R., Arita, J. T., & Fleck, M. S. (2009). Staying in bounds: Contextual constraints on object file coherence. *Visual Cognition, 17*, 195–211.

Mitroff, S. R., Scholl, B. J., & Noles, N. (2007). Object files can be purely episodic. *Perception*, *36*, 1730–1735.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2004). Divide and conquer: How object files adapt when a persisting object splits into two. *Psychological Science*, *15*, 420–425.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition*, *96*, 67–92.
Moore, C. M., & Lleras, A. (2005). On the role of object representations in substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1171–1180.
Müller, H., Elliott, M., Herrmann, C., & Mecklinger, A. (2001). Neural binding of space and time: An introduction. *Visual Cognition*, *8*, 273–285.
Navon, D. (1976). Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 130–138.
Noles, N., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object-file representations. *Perception and Psychophysics*, *67*, 324–334.
Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, *58*, 75–96.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.
Richardson, D. C., & Kirkham, N. Z. (2004). Multi-modal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, *133*, 46–62.
Saiki, J. (2000). Occlusion, symmetry, and object-based attention: Comment on Behrmann, Zemel, & Mozer (1998). *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 424–433.
Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*, 1–46.
Scholl, B. J., & Leslie, A. M. (1999). Explaining the infant's object concept: Beyond the perception/cognition dichotomy. In E. Lepore & Z. Pylyshyn (Eds.), *What is cognitive science?* (pp. 26–73). Oxford, UK: Blackwell.
Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290.
Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multi-element tracking. *Cognition*, *80*, 159–177.
Scholl, B. J., Pylyshyn, Z. W., & Franconeri, S. (1999). When are spatiotemporal and featural properties encoded as a result of attentional allocation? [Abstract]. *Investigative Ophthalmology and Visual Science*, *40*, S797.
VanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects vs. substances. *Psychological Science*, *14*, 498–504.
Watson, S., & Kramer, A. (1999). Object-based visual selective attention and perceptual organization. *Perception and Psychophysics*, *61*, 31–49.
Yi, D.-J., Turk-Browne, N. B., Flombaum, J. I., Kim, M., Scholl, B. J., & Chun, M. M. (2008). Spatiotemporal object continuity in human ventral visual cortex. *Proceedings of the National Academy of Sciences*, *105*, 8840–8845.

Manuscript received May 2008
Manuscript accepted October 2008
First published online February 2009