The persistence of object file representations

NICHOLAUS S. NOLES, BRIAN J. SCHOLL, and STEPHEN R. MITROFF
Yale University, New Haven, Connecticut

Coherent visual experience of dynamic scenes requires not only that the visual system segment scenes into component objects but that these object representations persist, so that an object can be identified as the same object from an earlier time. Object files (OFs) are visual representations thought to mediate such abilities: OFs lie between lower level sensory processing and higher level recognition, and they track salient objects over time and motion. OFs have traditionally been studied via *object-specific preview benefits* (OSPBs), in which discriminations of an object's features are speeded when an earlier preview of those features occurred on the same object, as opposed to on a different object, beyond general displaywide priming. Despite its popularity, many fundamental aspects of the OF framework remain unexplored. For example, although OFs are thought to be involved primarily in online visual processing, we do not know how long such representations persist; previous studies found OSPBs for up to 1,500msec but did not test for longer durations. We explored this issue using a modified *object reviewing* paradigm and found that robust OSPBs persist for more than five times longer than has previously been tested—for at least 8 sec, and possibly for much longer. Object files may be the "glue" that makes visual experience coherent not just in online moment-by-moment processing, but on the scale of seconds that characterizes our everyday perceptual experiences. These findings also bear on research in infant cognition, where OFs are thought to explain infants' abilities to track and enumerate small sets of objects over longer durations.

The raw input of visual processing is markedly different from our conscious percepts of the world. In particular, we experience the world as structured scenes of discrete objects, each with individual properties, whereas raw retinal images consist of undivided arrays of unbound features. The segmentation processes that are responsible for this critical transformation have been a major focus of research in vision science for decades. This work has demonstrated that some types of segmentation occur relatively early, and that later visual processes (e.g., attention) can then operate over fundamentally *object-based* representations (for a review, see Scholl, 2001). This research on object-based visual processing has covered many specific topics, including the nature of the initial segmentation processes (e.g., Driver, Davis, Russell, Turatto, & Freeman, 2001); the processes by which individual features are bound to specific object representations (e.g., Treisman & Gelade, 1980); the recognition of specific object types (e.g., Biederman, 1987) and views (e.g., Tarr, Williams, Hayward, & Gauthier, 1998); the composition of objects in terms of individual parts (e.g.,

Hoffman & Richards, 1984); the grouping processes by which objects are completed behind occluding surfaces (e.g., Nakayama, He, & Shimojo, 1995); and many others.

Much less research, however, has explored the *persistence* of these object representations in online visual processing. Our experience of the visual world, in addition to being structured, is inherently dynamic: We do not perceive a series of unconnected snapshots of the world, but rather a richly connected sequence of visual events involving persisting objects that retain their individual identities across time and motion. Without such a computation of persisting objecthood, visual experience would be incoherent. Yet, beyond work on low-level motion mechanisms, surprisingly little work has explored the principles that guide the construction and maintenance of representations of portions of the visual field as the *same* objects over time. Of the work that has explored dynamic object persistence, perhaps the most popular paradigm has involved the object file framework.

Object Files

The visual representations that are appealed to by psychologists typically consist of either low-level visual features ("It's red," "It's round") or higher level object types ("It's a ball," "It's a boat"). Our ability to track persisting objects cannot be fully explained in terms of such representations, however, since both lower level features and recognized object types can change within the lifetime of an individual object, even in online visual experience. This is nicely illustrated in the familiar phrase, "It's a bird! It's a plane! Superman!" (Kahneman, Treisman, & Gibbs, 1992). For the perceiver in such a situation, the

visual features and the identity of the object are in flux as the object looms closer, but there is never any doubt that "it" is the same persisting individual. Moreover, we can also easily perceive the opposite situation, wherein two featurally identical snapshots are seen as distinct objects: "Two identical red squares in successive fields may be perceived as distinct objects if the spatial/ temporal gap between them cannot be bridged, but the transformation of frog into prince is seen as a change in a single visual object" (Kahneman et al., 1992, p. 179). These types of situations have led some researchers to posit additional midlevel object representations that are responsible for tracking individual token objects over time in online visual processing.

One theory of such representations is the "object file" framework (e.g., Kahneman & Treisman, 1984; Kahneman et al., 1992). An *object file* is a midlevel 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 properties. Object files play a critical role in mediating our perception of persisting individuals, via three primary operations: (1) a *correspondence* operation, which determines for each object whether it is novel or has moved from a previous location; (2) a *reviewing* operation, which retrieves an object's previous characteristics, some of which may no longer be visible; and finally (3) an *impletion* operation, which uses both current and reviewed information to construct a phenomenal percept, perhaps of object motion. Because such operations are critical to computing object persistence, object files are accorded a primary role in vision: They are thought to constitute a "main end product of perceptual processing" (Kahneman et al., 1992, p. 178).

Object Reviewing

Evidence for the operation of such representations comes from the object reviewing paradigm (Kahneman et al., 1992). When the features of two object views at different times match the correspondence computed by spatiotemporal factors—that is, when features are similar across two object views that are seen as temporal stages of a single enduring object in the world—then certain responses are facilitated. In contrast, when the features do not match the computed correspondence, responses are inhibited.

In a typical object reviewing display, a small number of objects (small outlined boxes) are initially presented, and letters are then displayed within them. The letters then disappear and the objects briefly move about the screen. When they halt, a final letter is displayed within 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 displaywide priming effect). However, observers are faster still to name the final letter when it is the *same* letter that initially appeared on that object, as compared with when the final letter initially appeared on a different object—an *object-specific preview benefit* (OSPB; see Figure 1). This effect can thus be used as an index of persisting objecthood: Manipulations that degrade enduring object representations will result in attenuated OSPBs.

A Modified Object Reviewing Paradigm

The magnitudes of the OSPBs that are obtained in standard object reviewing experiments are often extremely small—often on the order of 10 msec. This is due in part to the nature of the underlying effect being measured, which is expected to be relatively small on theoretical grounds: "Matching a previously perceived frog to the prince currently in view may require a few extra milliseconds" (Kahneman et al., 1992, p. 184). However, the original object reviewing paradigm may also minimize such effects more than is necessary. Moreover, anecdotal reports—and our own experience at piloting such experiments—have suggested that the effects are not robust, and that testing many participants may be required in order to reveal statistically significant OSPBs. We think that this has been due to the fact that
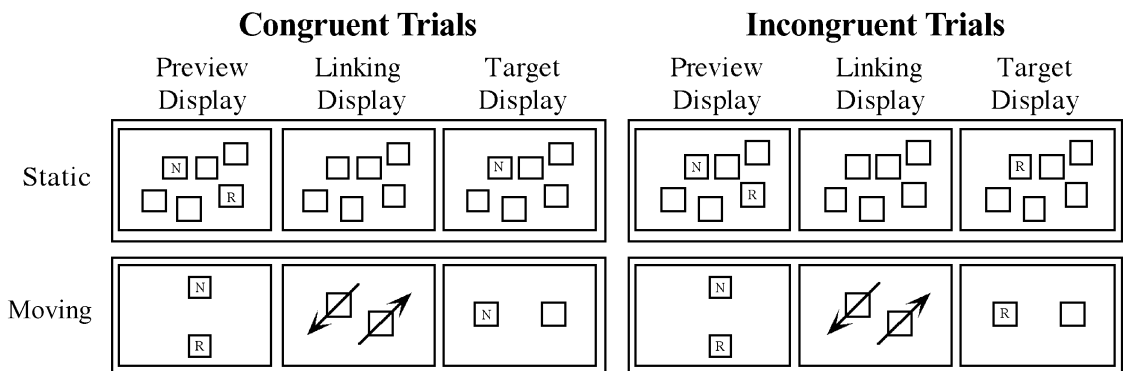

Figure 1. Sample displays used in the object reviewing paradigm (Kahneman et al., 1992). In the static displays, the target letter is seen as belonging to one of the objects shown in the preview, because it appears on the same object in the same location. Objecthood and location are unconfounded in the moving displays. In each case, congruent information facilitates target naming for the persisting object relative to incongruent information.

the standard paradigm, wherein the task is simply to name the final letter, allows observers to completely ignore the initially presented letters—a situation that may become increasingly likely as the experiment progresses. Accordingly, we have adopted a modification of the object reviewing paradigm (based on Kruschke & Fragassi, 1996) wherein participants simply press a key on each trial to indicate whether the final letter matched *any* of the initially presented letters. (In other words, participants make a forced choice match/no-match response on each trial.) This response, unlike voice-key naming, requires participants to attend to the initially presented letters, but it still allows us to measure OSPBs (since the participants make the same "match" response to both congruent "same-object" matches and incongruent "different-object" matches). This modification results in more robust effects and in OSPBs that are typically about three times as large as those obtained with the naming response (when computed in the same manner as in Kahneman et al., 1992).[1]

The Present Project: How Long Do Object Files Persist?

The object reviewing paradigm was used in the original study by Kahneman et al. (1992) to investigate many fundamental questions about object files, including capacity limits on the number of object files that can be simultaneously active; the ability of object files to persist across various types of motion (including apparent motion); the importance of the duration for which the initial letters are present; the inability of object files to store multiple temporal stages of a single object; and several other aspects of the object reviewing process. Since this seminal study, the object file framework has often been appealed to by researchers studying both adult vision and cognitive development (for reviews, see Carey & Xu, 2001, and Scholl, 2001), and several additional studies have employed the object reviewing paradigm (e.g., Gordon & Irwin, 1996, 2000; Henderson, 1994; Henderson & Anes, 1994; Pollatsek, Rayner, & Henderson, 1990). However, these additional studies have typically focused only on the question of what types of information (e.g., visual details, abstract identity) can be stored *in* object files. Although this question is a critical part of determining the underlying nature of object files, very few object reviewing experiments have directly explored the principles that cause object files to be constructed, destroyed, or updated. These issues are critical, since they essentially define what "counts" as a persisting object in the object file framework.

In this study, we explore what is perhaps the most basic factor underlying the destruction of object files: their decay over time. In operationalized terms, we can examine the temporal delay between the offset of the initially presented letters and the onset of the final letter (after the motion period), asking the question: How long can this temporal delay last while still producing OSPBs to persist? In short, how long can object files persist?

Kahneman et al.'s (1992) original study did vary this factor, but only across a very small range of delays, from 300–700 msec. Significant OSPBs have been obtained across this range, and many readers have thus assumed that object files operate only on a scale of milliseconds (e.g., Scholl & Leslie, 1999). However, even Kahneman and colleagues were careful to remain agnostic about the upper limit in this context: "When the previous information simply disappears, the preview benefit survives apparently intact for at least 600–700 ms, and *perhaps much longer*" (Kahneman et al., 1992, p. 208, emphasis added).

The purpose of the present study is to determine the timescale over which object file representations function. In the first place, we aim to test a longer range of temporal delays than those tested in the original studies of object files. More generally, the modified object reviewing paradigm described above may provide a more sensitive test of OSPBs at longer durations, since Kahneman and Treisman (1984) did observe an attenuation of OSPBs with increasing delays.

Determining the duration over which object files can persist is important to the study of midlevel visual processing for at least three reasons. First, this project fills a gap in our understanding of the object file framework itself, insofar as the duration of object files constitutes a relatively unexplored factor involved in the destruction and updating of such representations. Second, this project has the potential to broaden the scope of activities in which we think object files may be involved. Although the object file framework has often been thought to be involved in visual processing only on the order of milliseconds, the timescale of our conscious visual experience is often characterized on the order of several seconds. Indeed, we are often unaware of the millisecond-by-millisecond details of visual processing, and many day-to-day visual activities—from driving, to writing a paper, to playing baseball—seem to unfold over (at least) many seconds. As such, finding that OSPBs could persist for several seconds might forge a stronger tie between object files and common visual tasks.

A final related reason for exploring the persistence of object file representations relates to the possible explanatory value of the object file framework in the study of infant cognition. Using looking-time measures to study the infant's "object concept," developmental psychologists have demonstrated that infants even a few months old have a substantial amount of "initial knowledge" about objects, in domains such as physics and arithmetic (for reviews, see Baillargeon, 2002; Spelke, Phillips, & Woodward, 1995; Wynn, 1998). This research has shown, for example, that infants have some appreciation of the fact that objects must trace spatiotemporally continuous paths through space (Spelke et al., 1995); that seeing two feature clusters in different locations at the same time implies the existence of at least two objects (e.g., Xu & Carey, 1996); and that one plus one yields two, and other simple arithmetical facts (Feigen-

son, Carey, & Hauser, 2002; Wynn, 1992). A critical question in this literature is the nature of this "initial knowledge," and as will be further discussed in the General Discussion section, some researchers have appealed to the object file framework in order to explain these abilities (e.g., Carey & Xu, 2001; Scholl & Leslie, 1999). In this context, however, it is often stressed that the relevant infancy experiments require object representations to persist for at least several seconds, whereas it has not been previously demonstrated that object files can persist for longer than 1,500 msec (which was shown by Gordon & Irwin, 2000). Thus, more focused studies of the persistence of object files may also be important to our understanding of the roles they may play in other areas of cognitive science.

In Experiments 1 and 2, we first replicated the essential findings of the object file framework in both static and dynamic displays using our modified object reviewing paradigm. In Experiments 3, 4, and 5, we then tested for the presence of OSPBs in dynamic displays with temporal gaps of up to 8 sec—more than five times the previously tested values—using several types of object motion. On the basis of a hypothesized role for object files in everyday visual experience—and in the underlying performance in some infant cognition experiments—we predicted that object files should persist for at least 5 sec.

GENERAL METHOD

Participants

All observers were Yale University students and received either course credit or a small monetary payment for their participation in a single testing session. No observer participated in more than one experiment. All had normal or corrected-to-normal vision.

Apparatus and Materials

The displays were presented on a Macintosh iMac computer using custom software written using the Vision Shell graphics libraries (Comtois, 2003). The observers sat comfortably without head restraint approximately 50 cm from the monitor. (All measurements below are computed based on this viewing distance.)

Each trial began with two squares (each subtending 2.32°) presented as outlines (either white on black or black on gray, depending on the experiment). The preview and target letters were drawn in a monospaced font subtending 1.16° (again, either white on black or black on gray) and centrally located in the square frames. The preview letters were drawn without replacement from the set {K, M, P, S, T, V}, and the target letter either was one of the two preview letters from the given trial (on congruent and incongruent match trials) or was drawn from the remaining letters of the set (on no-match trials).

Procedure

The observer pressed a key to start each trial, causing the two square frames to appear 500 msec later. After 500 msec, a different preview letter appeared for 1,000 msec in each of the two frames. After an ensuing "linking display"—which was different for each experiment, and which often involved the frames moving to new locations—a single target letter appeared in one of the two square frames (equally often in each). The observers were instructed to press one key if the target letter was the same as *either* of the two preview letters (a match trial) or to press another key if the target letter was novel (a no-match trial). Half of the trials were match tri-

als, and half were no-match trials. On match trials, the target letter was equally likely to match the preview letter from either of the two square frames. Regardless of condition, the target letter was equally likely to appear in either frame. Detailed verbal and written instructions were provided to each observer. The observers were instructed to make the match/no-match responses as quickly as possible while remaining accurate. Every 50 trials, a dialogue box appeared to inform the observers of their progress and to allow them to take a break if they wished.

EXPERIMENT 1
Static Object Reviewing

Our modifications of the object reviewing paradigm are based on previous work that explored the role of object files in causal perception (Kruschke & Fragassi, 1996), but they have not been tested using simpler dynamic displays with noninteracting objects. Accordingly, we first sought to replicate the standard OSPB effects using the types of stimuli employed in the original object reviewing experiments of Kahneman and colleagues (Kahneman et al., 1992). This test would serve to demonstrate the sensitivity of the match/no-match task in such displays, so that any null object-specific effects in later experiments could be attributed to the particular duration manipulations used rather than to the paradigm itself. Moreover, this test would allow us to compare the resulting magnitudes of the OSPBs with those obtained with the more standard naming task. In this experiment, we tested the simplest such situation, the one involving static displays.

Method

Nine observers participated in individual 20-min testing sessions. All stimuli were drawn in black against a gray background. The two square frames were positioned 2.9° above and below the center of the display (all distances are measured from the squares' midpoints), so that their nearest edges were 3.48° apart. After the preview letters disappeared, the linking display consisted simply of the (again empty) frames for 1,000 msec, with no motion. The observers completed 20 practice trials, the results of which were not recorded, followed by 240 trials, presented in a separate random order for each observer.

Results and Discussion

On average, the observers were 97.08% accurate ($SD = 1.49\%$), and all analyses were conducted on only those trials with correct responses. The observers' overall mean response time was 591.64 msec. As in all object reviewing experiments, OSPBs were obtained by calculating the difference in response times (RTs) between incongruent and congruent match trials. OSPBs imply the maintenance of object-specific information (i.e., of information bound to one of the boxes in these displays) beyond simple displaywide priming. The observers' mean RTs were 566.91 msec on congruent match trials and 608.07 msec on incongruent match trials, yielding an OSPB of 41.16 msec [see Figure 2A; $t(8) = 5.94, p < .01$].

The significant OSPB obtained here indicates the maintenance of object-specific information and demonstrates that the modified object reviewing paradigm is
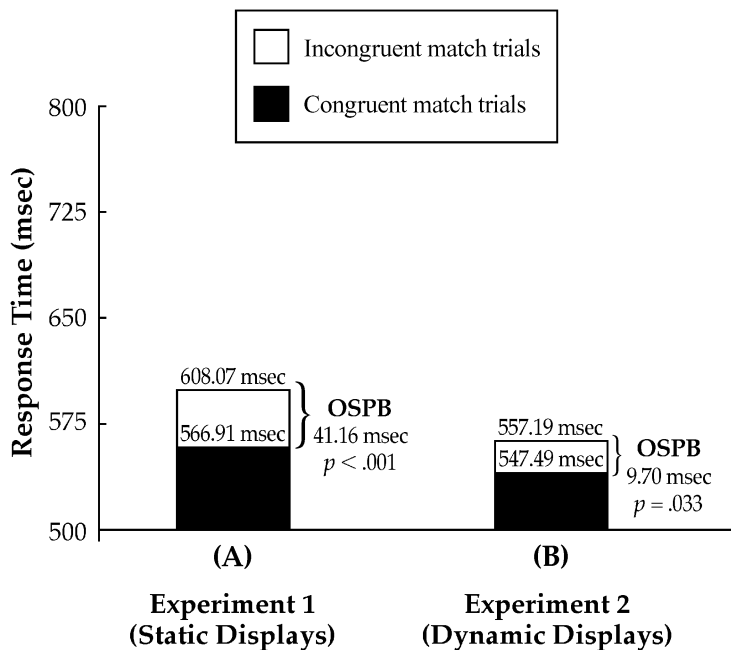
Figure 2. Mean response times for incongruent and congruent match trials in (A) the static displays of Experiment 1 and (B) the dynamic displays of Experiment 2.

sensitive enough to pick up effects in simple static displays of roughly the same magnitude as those found in the original object reviewing experiments of Kahneman et al. (1992).

EXPERIMENT 2
Dynamic Object Reviewing

Static displays have traditionally yielded the most robust OSPBs, and as such served as a good first test of the modified object reviewing paradigm in Experiment 1. However, static displays can never unconfound objects and locations: Perhaps the results of Experiment 1 did not reflect any sort of object-specific processing at all, but simply preview benefits tied to the spatial position of the relevant preview letter (or some combination of the two). Following Kahneman et al. (1992), we verified in this experiment that OSPBs truly reflect object-specific processing by having the objects move during the linking display (as in the bottom frame of Figure 1).

Method
This experiment was identical to Experiment 1, except as noted here. Sixteen observers participated in individual 20-min testing sessions. All stimuli were drawn in black against a gray background. The two square frames were initially horizontally aligned, with their centers 2.90° to the left and right of the center of the display. During the linking display (still lasting 1,000 msec after the preview letters disappeared), the frames moved diagonally in opposite directions (either clockwise or counterclockwise, determined randomly on each trial) until they were vertically aligned, with their

centers 2.90° above and below the center of the display, respectively. (See the lower panel of Figure 1 for a schematic depiction of this motion.) Each frame traveled at a constant speed of 4.10°/sec.

Results and Discussion
On average, the observers were 96.85% accurate ($SD = 2.76\%$), and all analyses were conducted on only those trials with correct responses. The observers' overall mean response time was 549.31 msec. The mean RT on congruent match trials was 547.49 msec, and on incongruent match trials it was 557.19 msec, yielding an OSPB of 9.70 msec [see Figure 2B; $t(15) = 2.35$, $p = .03$].

Because the objects moved in this experiment—so that the target letter was in a novel location, equidistant from each original preview letter—the resulting OSPB must reflect the maintenance of object-specific information. We are thus confident in the sensitivity of the modified object reviewing paradigm, and in the remaining experiments we used it to explore the persistence of object representations across a variety of temporal delays, using several sorts of motion.

EXPERIMENT 3
Persistence Through Translating Motion

The first two experiments, although they were conducted primarily to test the sensitivity of the modified object reviewing paradigm, have already demonstrated that OSPBs last 40% longer than the durations previously tested by Kahneman et al. (1992). Emboldened by this result, we now tested the effect for durations up to four times

this value. The linking-display motions in this experiment consisted of the simple parallel translation of a pair of vertically aligned boxes across the display, from left to right (see Figure 3A). The objects moved at a constant speed, stopping after 500, 1,000, 2,000, or 4,000 msec.

If object files are relatively transient in nature, and if previous experiments span the range of durations across which they can persist, then we would expect to see significant OSPBs for only the first two of these durations. However, if object files persist for much longer than the results of previous experiments have suggested—if they can operate on a scale of several seconds rather than only milliseconds—then we would expect to see significant OSPBs for the longer durations as well. Moreover, comparing the resulting OSPB magnitudes would give us some hint of the temporal flexibility of object files: Linearly decreasing OSPBs would suggest the upper bound for which object files can persist, whereas roughly constant OSPBs would suggest that even the values tested here do not approach the maximum.

Method

Twenty-three observers participated in individual 45-min testing sessions. All stimuli were drawn in white against a black background. The two square frames were initially vertically aligned, with their centers 2.32° above and below the center of the display. Unlike the previous experiments, however, the pair of frames was now presented toward the left extreme of the display instead of in the center—such that the centers of the frames were 11.6° to the left of the center of the display. During the linking display (after the preview letters had disappeared), the pair of frames translated to the right at 5.80°/sec for either 500, 1,000, 2,000, or 4,000 msec (thus moving 2.9°, 5.8°, 11.6°,

or 23.2°, respectively). In this way, distance and duration were perfectly confounded, so that (for example) the frames in the 1,000-msec condition moved twice as far as those in the 500-msec condition. The observers completed 16 practice trials, the results of which were not recorded, followed by 256 trials—64 trials of each duration, presented in a separate random order for each observer.

Results and Discussion

On average, the observers were 95.31% accurate ($SD = 2.44\%$), and analyses were conducted on only those trials with correct responses. Accuracy did not differ by duration condition [$F(3,88) = 2.02, p = .12$; accuracies for the 500-, 1,000-, 2,000-, and 4,000-msec conditions were 95.04%, 95.11%, 94.43%, and 96.67%, respectively]. The observers' overall mean RT was 720.73 msec. RTs for the congruent and incongruent match trials are depicted in Figure 4, broken down by linking-display duration. OSPBs were obtained for each duration, although the effect was only marginally significant at 1,000 msec {at 500 msec, 82.06 msec OSPB [$t(22) = 3.22, p < .01$]; at 1,000 msec, 35.33 msec OSPB [$t(22) = 1.81, p = .08$]; at 2,000 msec, 63.95 msec OSPB [$t(22) = 2.11, p = .046$]; and at 4,000 msec, 38.71 msec OSPB [$t(22) = 2.68, p = .01$]}.[2]

These results confirm that object files can persist for longer durations than those tested in previous experiments—and more than twice as long as the maximum previously tested value. Object files may thus be the "glue" that makes visual experience coherent not just in online moment-by-moment processing on a scale of milliseconds, but also on the scale of seconds that characterizes our everyday perceptual experiences.
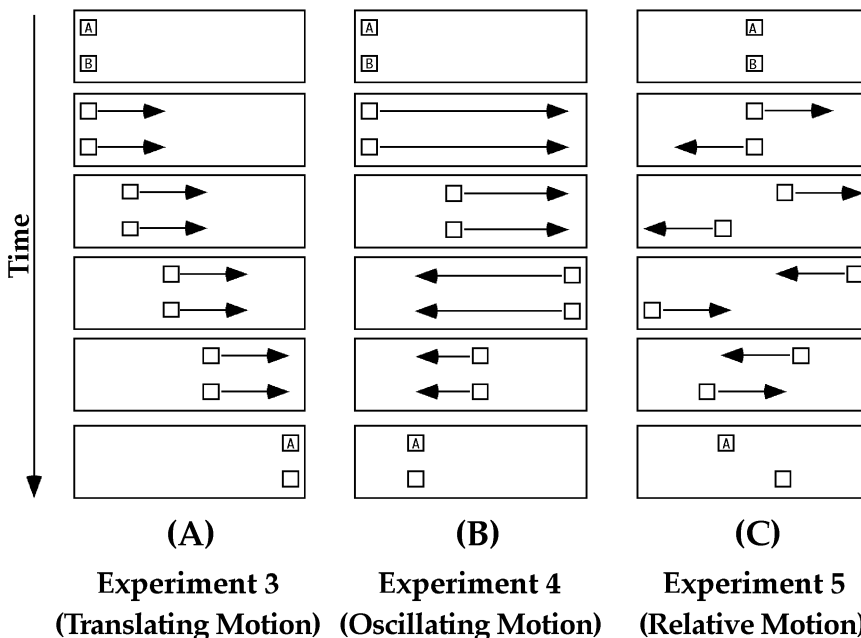

Figure 3. Schematic depictions (not to scale) of congruent match trials with dynamic linking displays of 4 sec and the types of motion employed in (A) Experiment 3, (B) Experiment 4, and (C) Experiment 5. (Note that 4 sec was the only linking-display duration common to all three experiments.)
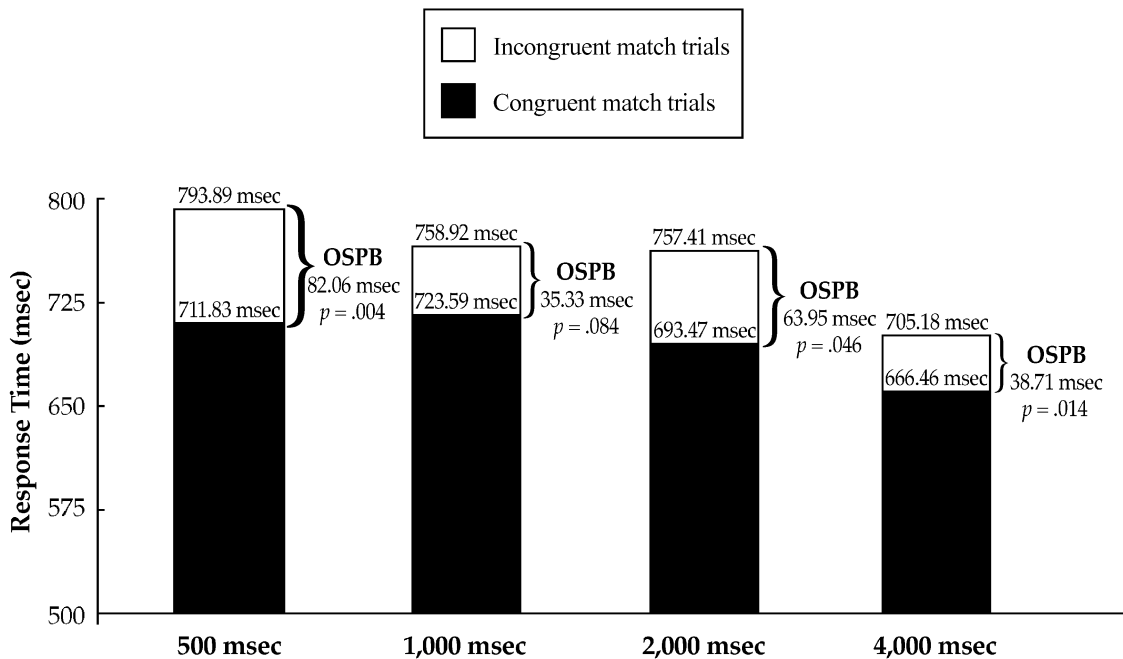
Figure 4. Mean response times for incongruent and congruent match trials, broken down by linking-display duration, in Experiment 3.

EXPERIMENT 4
Persistence Through Oscillating Motion

The fact that OSPBs were not attenuated with increasing linking-display durations in Experiment 3 suggested that we had not yet reached the temporal limits of object file representations. Accordingly, we also tested for the persistence of object files after 8-sec linking-display intervals—twice the maximum tested in the previous experiment, and more than 10 times the maximum tested in the original experiments of Kahneman et al. (1992). We also attempted to replicate the results of the previous experiment by testing durations of 2 sec and 4 sec as well.

Beyond these changes in duration, we also made one other critical change in the stimuli in this experiment in order to test the robustness of object file persistence. The previous experiment used simple translating motion, but in the real world objects frequently alter their trajectories, and often do not maintain perfectly linear paths for long durations. Thus, this experiment involved oscillating motion: At the shortest duration (2 sec), the objects simply translated to the right as in Experiment 3. For the longer two durations, however, the objects oscillated back and forth along this path segment, always moving at the same speed (see Figure 3B). This manipulation allowed us to determine whether object files persist through trajectory changes in addition to persisting for particular durations.

Method

This experiment was identical to Experiment 3, except as noted here. Forty-two observers participated in individual 45-min testing

sessions. During the linking display, the pair of frames translated horizontally at 11.60°/sec for 2,000, 4,000, or 8,000 msec. The locations of the frames at the end of each duration's motion were varied, so that observers could not simply wait and monitor a single location for all trials. On 2,000-msec trials, the frames simply moved 23.20° to the right before coming to a stop. On 4,000-msec trials, the frames translated 27.84° to the right, then immediately reversed their direction (with no pause) and translated leftward for 18.56° (coming to a stop 9.28° to the right of their starting position). On 8,000-msec trials, the frames translated to the right for 27.84°, to the left for 27.84°, to the right once again for 27.84°, and finally to the left for 9.28° (coming to a stop 18.56° to the right of their starting position). As in Experiment 3, distance and duration were perfectly confounded, so that (for example) the frames in the 8,000-msec condition moved twice as far as those in the 4,000-msec condition. Observers completed 24 practice trials, the results of which were not recorded, followed by 216 trials—72 trials of each duration, presented in a separate random order for each observer.

Results and Discussion

On average, the observers were 93.40% accurate ($SD = 4.79\%$), and analyses were conducted on only those trials with correct responses. Accuracy did not differ by duration condition [$F(2,123) = 0.532, p = .59$]. The observers' overall mean RT was 752.82 msec. RTs for the congruent and incongruent match trials are depicted in Figure 5, broken down by linking-display duration. OSPBs were obtained for each duration, although the effect was only marginally significant at 2,000 msec {at 2,000 msec, 25.88 msec OSPB [$t(41) = 1.80, p = .09$]; at 4,000 msec, 43.25 msec OSPB [$t(41) = 2.88, p = .01$]; and at 8,000 msec, 29.99 msec OSPB [$t(41) = 2.30, p = .04$]}.

These results replicate those of Experiment 3 and extend them in two ways. First, this experiment demonstrates that object files last even longer—for at least 8 sec, which is more than 10 times the value tested in the original object file experiments. Second, this experiment demonstrates that object files can persist for this long even when the corresponding objects undergo severe trajectory changes during their motion. Whereas other experiments have used smooth curvilinear motion (Kahneman et al., 1992) or slight angular inflections (Mitroff, Scholl, & Wynn, 2004), to our knowledge this is the first demonstration that object files can persist through such severe violations of inertia. Thus, we see that object files are robust in the face of both simple display manipulations and temporal delays.

EXPERIMENT 5
Persistence Through Relative Motion

The OSPBs obtained in object reviewing experiments with dynamic stimuli must reflect the maintenance of object-specific information (and not just location-based priming), since the preview and target letters are presented in different locations. Because the motion in Experiments 3 and 4 consisted of parallel translation, however, it is possible that the preview and target letters could have appeared in the same retinal locations (although not in the same environmental locations), if observers happened to keep their eyes positioned perfectly between the two frames throughout the translating or oscillating motions. We have already demonstrated—using the rotational motion in Experiment 2—that this interpretation cannot be true in general. However, as a check to ensure that the maintenance of object files across longer durations cannot reflect location-specific priming, here we sought to demonstrate that object files persist even when the two frames move relative to each other (oscillating in opposite directions), so that no pattern of eye movements could result in the preview and target letters being in the same retinal positions. We tested for the persistence of OSPBs through such manipulations using 4-sec linking displays.

Method

This experiment was identical to Experiment 3, except as noted here. Fourteen observers participated in individual 30-min testing sessions. The two square frames were initially vertically aligned, with their centers 2.32° above and below the center of the display. After the preview display, the empty square frames moved 9.28° in opposite directions (one moving left and the other to the right, with the motion of the boxes randomly assigned on each trial). The frames then changed direction (with no pause, as in Experiment 4) and moved an additional 13.92° in the opposite direction from their initial motions before coming to a stop 4.64° to the right and left of their starting locations, respectively (see Figure 3C). The frames moved at a constant 5.80°/sec for the 4,000-msec duration of each linking display. We also included a static fixation point (an outlined square subtending 0.58°) located in the center of the screen, directly between the initial locations of the two frames, and the observers were encouraged to maintain fixation throughout each trial rather than tracking either of the motions. (Note, however, that even if observers did move their eyes, the two boxes were still in different relative positions during the presentation of the preview and target letters—such that no pattern of eye movements could have resulted
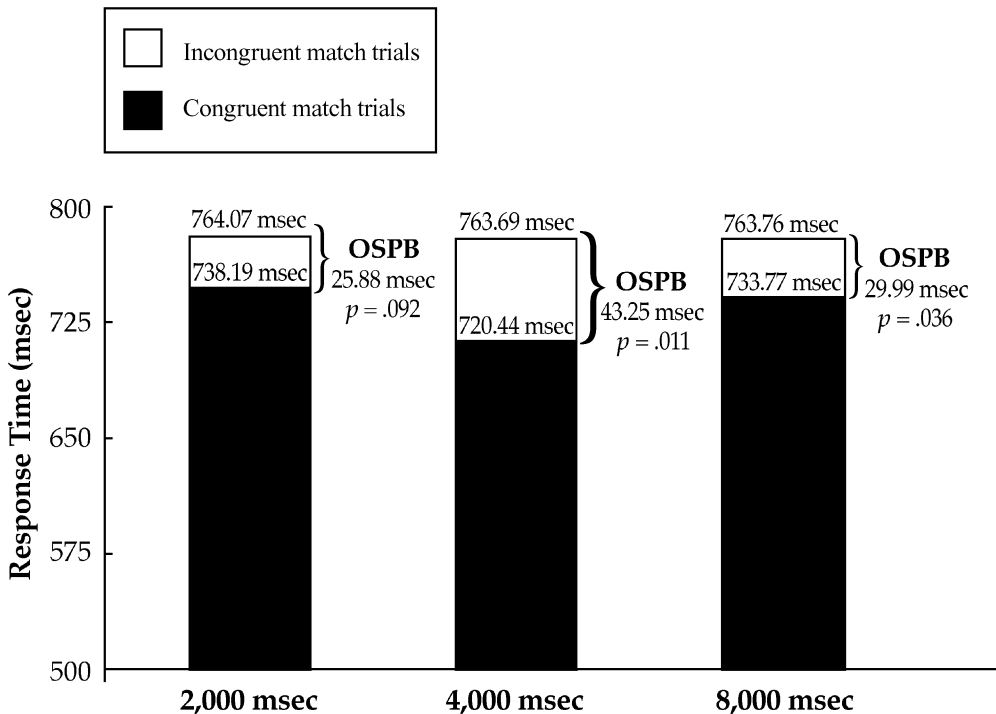

Figure 5. Mean response times for incongruent and congruent match trials, broken down by linking-display duration, in Experiment 4.

in retinally stable frame positions.) Observers completed 16 practice trials, the results of which were not recorded, followed by 128 test trials.

Results and Discussion

On average, observers were 93.48% accurate (SD = 10.13%), and all analyses were conducted on only those trials with correct responses. The observers' overall mean RT was 673.52 msec, and a significant OSPB of 39.15 msec was obtained [see Figure 6; $t(13) = 2.46$, $p = .03$]. This result constitutes a third demonstration (replicating Experiments 3 and 4) that object files can persist for at least 4 sec. More importantly, this control experiment demonstrates that the persistence of OSPBs through such durations must reflect object-specific processing, since the observers were instructed to maintain fixation and no pattern of eye movements could have resulted in retinal location–specific effects.

Although we take these results to have eliminated the possibility of retinally based spatial effects, it is important to note that there are still several ways in which the object-based nature of this processing could be realized. One possibility, for example, is that our effects could be partially supported by the *relative* positions of the two translating objects—that is, by the fact that the same object is above the other in both the initial and final displays, despite their differing retinal positions. This idea has been noted in passing in other object file research, and it potentially affects the interpretation of other studies that have used translating motion (see Henderson & Anes, 1994), but to our knowledge it has never been directly explored. However, this possibility seems unlikely to us, since we have also obtained OSPBs of similar magnitudes using this method with displays in which the

two objects traveled along the diagonals of a square and thus did not maintain the same relative positions, either horizontally or vertically (Mitroff, Scholl, & Wynn, 2005). In any case, the present effects do rule out any simple spatial explanations based on the objects' final positions matching their initial retinal or environmental positions.

A second possibility to consider is that the results could reflect retinally specific location-based processing due to the tracking of only a single object on each trial. This is an important concern that also potentially affects the interpretation of many (or even most) other object reviewing experiments (including those of Kahneman et al., 1992), but in fact we think there are several reasons to doubt that such tracking occurred. First, although we did not monitor eye movements in this experiment, the observers were instructed to fixate throughout the trial and not to track any of the objects. We also obtained similar patterns of results with psychophysically trained observers who are experienced at maintaining fixation. Second, this interpretation predicts that the observed OSPBs would be the result of a bimodal distribution of responses, with even more robust effects on half of the trials (when the observer happened to track the probed object) and no effect on the other half of the trials (when he or she tracked the unprobed object). However, no such pattern was apparent in our data: Each of the means is the result of a roughly Gaussian distribution. In any case, the results of this experiment demonstrate that object files can persist for at least several seconds when measured using displays and controls similar to those in both the initial object reviewing experiments (Kahneman et al., 1992) and several other recent reports (Mitroff et al., 2004, 2005).

GENERAL DISCUSSION

The five object reviewing experiments reported here yield a simple primary conclusion: Object files can persist for much longer than has been previously suspected or tested—for at least 8 sec, which is more than 10 times the maximum delay tested in the original object reviewing demonstrations of Kahneman et al. (1992). Moreover, we have also demonstrated that object files can persist for such durations through several types of motion, including sudden changes in direction more extreme than those tested in previous studies.

These results bear in important ways on both the nature of object files and the situations in which they operate. Object files have often been thought of in especially transient terms, as operating only on the scale of milliseconds that has been characteristic of previous object reviewing studies. However, whereas many aspects of lower level visual processing do in fact operate on a scale of milliseconds—including the "microgenesis" of individual static percepts such as amodal completion (e.g., Rauschenberger & Yantis, 2001)—our everyday conscious perception of the world unfolds and changes on the scale of several seconds. We are not typically aware
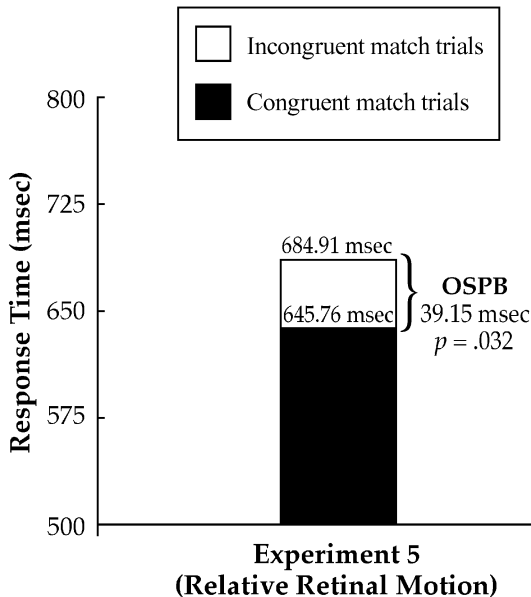

Figure 6. Mean response times for incongruent and congruent match trials in Experiment 5.

of the moment-by-moment details of our visual world. Rather, in most everyday activities—from driving to doing dishes to reading an academic paper—we often perceive (and remember) the world in terms of extended temporal events, each of which can last several seconds (e.g., Zacks & Tversky, 2001). As a result, it is natural to ask about the ways in which object files may underlie computations of object persistence in such activities on a scale of seconds, and our results suggest that object files may indeed operate in such contexts. (See Mitroff et al., 2005, for other studies of the relationship of object files to conscious visual perception.)

The persistence of object file representations for several seconds raises the possibility of using the object reviewing paradigm as a tool for exploring many other types of temporally extended visual events. In particular, many well-known visual phenomena also unfold over the course of several seconds, and we can thus ask about the roles that object files play in such situations. To take a single example, many researchers have explored object persistence using only simple perceptual reports on ambiguous dynamic displays, such as "bouncing versus streaming" (e.g., Bertenthal, Banton, & Bradbury, 1993; Sekuler & Sekuler, 1999; Watanabe & Shimojo, 1998). In such displays, two objects move on opposite diagonals of a square (meeting in the center before continuing their motion) and can be seen as either streaming past or bouncing off each other. What factors determine which of these possibilities is perceived? Previous research on such questions has had to rely on explicit perceptual reports, but the persistence of object files for several seconds allows us to directly explore the conditions under which object files influence the perception of bouncing or streaming by analyzing the resulting OSPBs (Mitroff et al., 2005). Doing so not only provides a new type of tool with which to study object persistence across temporally extended events, but also allows for exploring the influence of more complicated types of object transformations on the construction, maintenance, and destruction of object files (see also Mitroff et al., 2004).

Similarly, object files may also play important roles in other common experimental paradigms. For example, in multiple-object tracking (MOT) experiments, observers must track a subset of identical objects that follow haphazard unpredictable paths about the display. Some investigators (e.g., Kahneman et al., 1992; Scholl, 2001) have attempted to explain various aspects of MOT in terms of the operation of object files—suggesting, for example, that object files can survive occlusion on the basis of MOT results (Scholl & Pylyshyn, 1999). Such suggestions have always been tentative, since MOT trials last several seconds, but the present results suggest that object file representations could in fact underlie aspects of visual processing in such paradigms. Given the popularity of the object file framework, an interesting task for future work will be to explore whether and how object files play a role in many other temporally extended perceptual phenomena and paradigms (e.g., inattentional

blindness, object substitution masking, the detection of gradual changes, or the perception of causality).

Finally, the persistence of object files raises the possibility that they may play an important role in other traditionally independent research areas. For example, several researchers in the field of infant cognition have noted similarities between theories of dynamic object-based attention in adults and theories of the infant's "object concept." Using looking-time measures, infancy researchers have demonstrated that infants' "initial knowledge" of objects often seems to parallel principles of object persistence in adults' visual cognition. For example, both domains seem to involve similar numerical limits, object-based properties, spatiotemporal biases, and individual principles of object persistence (see Carey & Xu, 2001; Scholl & Leslie, 1999; vanMarle & Scholl, 2003). For the most part, however, this work has only raised analogies, since object files were thought to last for only a few hundred milliseconds, whereas the events characteristic of infancy experiments typically last several seconds. Indeed, the difference in the timescales of these processes has been explicitly identified as a primary objection to the integration of these previously separate research areas (Leslie, Xu, Tremoulet, & Scholl, 1998; Scholl & Leslie, 1999). The present experiments dispatch this objection by demonstrating that object files do indeed persist long enough to be active in such experiments. This result clears the way for a continued investigation of the ways in which object files may underlie infants' appreciation of simple numerical and physical laws (see also Chiang & Wynn, 2000; Feigenson et al., 2002; Feigenson & Halberda, 2004).

In sum, the present study has resolved a lingering ambiguity about the nature of object files and how they fit in with other types of temporally extended perceptual processing. Object files may thus be the "glue" that binds discrete perceptual snapshots into the persisting representations that make visual experiences coherent not just in online moment-by-moment processing, but on the scale of our everyday perceptual experiences.

REFERENCES

BAILLARGEON, R. (2002). The acquisition of physical knowledge in infancy: A summary in eight lessons. In U. Goswami (Ed.), *Blackwell handbook of child cognitive development* (pp. 47-83). Oxford: Blackwell.

BERTENTHAL, B. I., BANTON, T., & BRADBURY, A. (1993). Directional bias in the perception of translating patterns. *Perception*, **22**, 193-207.

BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

CAREY, S., & XU, F. (2001). Infant knowledge of objects: Beyond object files and object tracking. *Cognition*, **80**, 179-213.

CHIANG, W.-C., & WYNN, K. (2000). Infants' representation and tracking of multiple objects. *Cognition*, **77**, 169-195.

COMTOIS, R. (2003). *Vision Shell PPC* [Software]. Watertown, MA: Author.

DRIVER, J., DAVIS, G., RUSSELL, C., TURATTO, M., & FREEMAN, E. (2001). Segmentation, attention, and phenomenal visual objects. *Cognition*, **80**, 61-95.

FEIGENSON, L., CAREY, S., & HAUSER, M. (2002). The representations

underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, **13**, 150-156.

FEIGENSON, L., & HALBERDA, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, **91**, 173-190.

GORDON, R. D., & IRWIN, D. E. (1996). What's in an object file? Evidence from priming studies. *Perception & 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, & Cognition*, **26**, 136-150.

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 & Performance*, **20**, 826-839.

HOFFMAN, D. D., & RICHARDS, W. A. (1984). Parts of recognition. *Cognition*, **18**, 65-96.

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.

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: Erlbaum.

LESLIE, A. M., XU, F., TREMOULET, P., & SCHOLL, B. J. (1998). Indexing and the object concept: Developing "what" and "where" systems. *Trends in Cognitive Sciences*, **2**, 10-18.

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.

NAKAYAMA, K., HE, Z. J., & SHIMOJO, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In D. N. Osherson (Series Ed.) & S. M. Kosslyn & D. N. Osherson (Vol. Eds.), *An invitation to cognitive science: Vol. 2. Visual cognition* (2nd ed., pp. 1-70). Cambridge, MA: MIT Press.

POLLATSEK, A., RAYNER, K., & HENDERSON, J. M. (1990). The role of spatial location in the integration of information across fixations. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 199-210.

RAUSCHENBERGER, R., & YANTIS, S. (2001). Masking unveils preamodal completion representation in visual search. *Nature*, **410**, 369-372.

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: Blackwell.

SCHOLL, B. J., & PYLYSHYN, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, **38**, 259-290.

SEKULER, A., & SEKULER, R. (1999). Collisions between moving visual targets: What controls alternative ways of seeing an ambiguous display? *Perception*, **28**, 415-432.

SPELKE, E. S., PHILLIPS, A., & WOODWARD, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber & D. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44-78). Oxford: Oxford University Press, Clarendon Press.

TARR, M., WILLIAMS, P., HAYWARD, W., & GAUTHIER, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, **1**, 275-277.

TREISMAN, A. M., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

VANMARLE, K., & SCHOLL, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, **14**, 498-504.

WATANABE, K., & SHIMOJO, S. (1998). Attentional modulation in perception of visual motion events. *Perception*, **27**, 1041-1054.

WYNN, K. (1992). Addition and subtraction by human infants. *Nature*, **358**, 749-750.

WYNN, K. (1998). Psychological foundations of number: Numerical competence in human infants. *Trends in Cognitive Sciences*, **2**, 296-303.

XU, F., & CAREY, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, **30**, 111-153.

ZACKS, J., & TVERSKY, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, **127**, 3-21.

NOTES

1. Note that because our modified paradigm requires observers to maintain some information across the linking display, we are essentially asking here how long object files *can* last—supported by attention and memory—and not necessarily how long such representations will always last in other situations, when attention is otherwise engaged. We do know that OSPBs in this paradigm are not completely driven by intentional strategies and task-driven processes, since in other recent work we have shown some striking limitations of object files using this method: In one study, we presented displays in which dynamic persisting objects split into two (Mitroff, Scholl, & Wynn, 2004), and in another, dynamic displays involving ambiguous "bouncing versus streaming" motions (Mitroff, Scholl, & Wynn, 2005). However, the role of attention in object reviewing has not previously been studied and remains an important topic for further research.

2. Across all durations and motion types that yielded significant OSPBs in each experiment reported in this article, the observers were faster to respond when the target letter originally appeared as the top rather than the bottom preview letter. Moreover, we have consistently observed this pattern in many other object reviewing experiments (e.g., Mitroff et al., 2004), as have other researchers (e.g., Gordon & Irwin, 2000). This location-based priming was necessarily factored out in all of our object-specific comparisons, which averaged across locations so that no OSPBs could be driven solely by such displaywide spatial effects. (If the putative OSPB did in fact reflect prioritized processing only for information initially presented in the top object, then we would observe a negative OSPB of exactly the same magnitude for the information originally presented in the bottom object. The fact that a significant OSPB persisted when both were averaged together indicates that object-specific processing combined with a location-specific benefit.) Additional experiments are in progress to determine just how many objects can simultaneously give rise to OSPBs.