Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Cognition 96 (2005) 67–92
COGNITION
www.elsevier.com/locate/COGNIT

The relationship between object files and conscious perception

Stephen R. Mitroff*, Brian J. Scholl, Karen Wynn

Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, USA

Received 10 September 2003; accepted 29 March 2004

Abstract

Object files (OFs) are hypothesized mid-level representations which mediate our conscious perception of persisting objects—e.g. telling us 'which went where'. Despite the appeal of the OF framework, not previous research has directly explored whether OFs do indeed correspond to conscious percepts. Here we present at least one case wherein conscious percepts of 'which went where' in dynamic ambiguous displays diverge from the analogous correspondence computed by the OF system. Observers viewed a 'bouncing/streaming' display in which two identical objects moved such that they could have either bounced off or streamed past each other. We measured two dependent variables: (1) an explicit report of perceived bouncing or streaming; and (2) an implicit 'object-specific preview benefit' (OSPB), wherein a 'preview' of information on a specific 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 object), beyond display-wide priming. When the displays were manipulated such that observers had a strong bias to perceive streaming (on over 95% of the trials), there was nevertheless a strong OSPB in the opposite direction—such that the object files appeared to have 'bounced' even though the percept 'streamed'. Given that OSPBs have been taken as a hallmark of the operation of object files, the five experiments reported here suggest that in at least some specialized (and perhaps ecologically invalid) cases, conscious percepts of 'which went where' in dynamic ambiguous displays can diverge from the mapping computed by the object-file system.

Keywords: Conscious perception; Object files; Visual processing

* Corresponding author. Tel.: +1 203 432 2518; fax: +1 203 432 7172.
E-mail addresses: stephen.mitroff@yale.edu (S.R. Mitroff), brian.scholl@yale.edu (B.J. Scholl).
URL: http://www.yale.edu/perception/

0022-2860/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2004.03.008

1. Introduction

Psychologists regularly appeal to two types of visual representations: early lower-level perceptual features ("It's red", "It's round") and later higher-level recognized types ("It's a duck", "It's a rabbit"). However, these two types of representations cannot fully account for many aspects of visual processing. We can easily track an object as the *same* persisting individual over time and motion, for example, despite dramatic changes to both types of representations; in such situations, there is no doubt that the transformation from frog to prince involves a single persisting object (Kahneman, Treisman, & Gibbs, 1992). Without a computation of such persisting objecthood—over time, motion, and property changes—visual experience would be incoherent. To account for such abilities, an intermediate level of representation is required, and the object-file framework is perhaps the most popular theory of such representations (Kahneman & Treisman, 1984; Kahneman et al., 1992).

An 'object file' is a mid-level visual representation which 'sticks' to a moving object over time on the basis of spatiotemporal properties, and stores (and updates) information about that object's surface properties. In this way, object files help to mediate our conscious perception of persisting objects—e.g. telling us 'which went where' or allowing us to perceive object persistence despite featural changes or occlusion. Object files are thought to underlie object representations through a three-stage process consisting of: (a) a *correspondence operation*, which uses spatiotemporal information to determine for each visual object whether it is novel or whether it moved from a previous location; (b) a *reviewing operation*, which retrieves previously stored object properties (e.g. color, shape); and (c) an *impletion operation*, which uses both current and reviewed information to create an evolving conscious percept, perhaps of object motion.

1.1. The object-reviewing paradigm

Direct evidence for the operation of object files comes from the object-reviewing paradigm (Kahneman et al., 1992). Theoretically, this paradigm involves both the 'correspondence' and 'reviewing' operations. When the features of two object-views at different times match the correspondence computed by spatiotemporal factors—i.e. when the features are similar across two object-views which 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.

Two typical examples of this paradigm are depicted in Fig. 1. In the initial preview display a number of distinct objects are presented and then letters briefly appear on some or all of them. 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

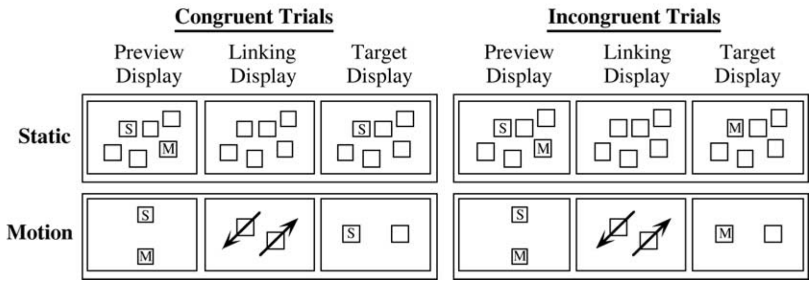

Fig. 1. Sample displays used in the original object-reviewing experiments of Kahneman et al. (1992). In the static displays, the target is seen as the same object as one of the previews, because it appears on the same object, in the same location. Objecthood and location are unconfounded in the moving displays. In each case, congruent information facilitates target naming on the same object, relative to incongruent information. (These experiments also involved No-Match trials, not depicted here.)

congruent trial), compared to when the final letter initially appeared on a different object (in an *incongruent*, trial see Fig. 1). 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.

1.2. Object files and conscious perception

A basic assumption of the object-file framework is that object files mediate the construction of conscious percepts—that these mid-level representations serve as a critical connection between lower-level visual processing and higher-level visual experience. Indeed, the impletion operation explicitly presupposes a direct connection between object files and conscious perception: the whole point of this operation is to combine the spatiotemporal mapping of 'which went where' (computed by the correspondence operation) with any featural changes (computed by the reviewing operation), and to realize this combination as a visual percept, which may involve both motion and featural change. In this way the object-file framework (in general) and the impletion operation (in particular) are thought to underlie our conscious perception of both change and persistence.

Despite this tight theoretical coupling between object files and conscious perception, there is actually very little experimental work focused on this connection. The seminal experiments of Kahneman and Treisman and their colleagues (Kahneman et al., 1992) did demonstrate several cases in which OSPBs seemed to coincide with conscious percepts: they reported several experiments wherein both the OSPBs and the conscious percepts seem to trace the same unambiguous paths. For example, in the lower panels of Fig. 1, observers consciously perceive one object start in the upper position and then move to the left, and an OSPB also involves this temporal pair. However, the visual events involved in such displays are especially unambiguous, and thus do not constitute a strong test of the coupling between object files and conscious percepts. Indeed, Kahneman et al. did not even bother to collect reports of the conscious percepts involved in these events, and it would have been silly to do so: with such clear independent paths, there seems little doubt that all observers would see exactly the same thing on every trial. In contrast, real-world trajectories can be more complicated: not only can object paths frequently overlap, but objects themselves often occlude one another. Such situations, unlike the simpler events used in the initial object-reviewing experiments, give rise to correspondence problems—which went where?—and it remains unclear whether the object-file system's solutions to such correspondence problems will always match conscious perception.

Many other studies since the seminal work of Kahneman and Treisman have used the object-reviewing paradigm, but these other studies have also tended to use especially simple and unambiguous motion events. Most of the recent work on object files has addressed the question of what types of information can be stored in object files and can elicit OSPBs. The initial work with simple flashed letters is ambiguous in this regard: the reviewing process could operate over matches between the abstract identities of the letters, or their specific visual forms. More recent work has shown that object files are not always tied to the specific physical form or properties of an object, but rather can be indexed at a more abstract level, addressed through post-categorical identity information (Gordon & Irwin, 1996, 2000; Henderson, 1994; Henderson & Anes, 1994). Thus, for example, OSPBs are obtained even when the preview is a picture (e.g. of a fish) and the target is a word (e.g. the word 'fish') (Gordon & Irwin, 2000).

Other recent work with the object-reviewing paradigm has explored the processes which mediate the construction, maintenance, and destruction of object files. One recent project, for example, explored the duration for which object files can persist, and found that they last for up to 8 s—at least five times longer than the previously tested values (Noles, Scholl, & Mitroff, in press). Other studies have explored the ways in which the maintenance of object files is affected by principles such as cohesion—the idea that objects must always maintain a single bounded contour. When an object gradually splits into two, for example, the associated object file cannot also split without incurring significant performance costs (Mitroff, Scholl, & Wynn, 2004).

Though all of this recent work has added to our understanding of the object-file framework in important new ways, none of these studies directly test the hypothesized link between object files and conscious perception. There has yet to be an exploration of the underlying *nature* of object files, in the sense of just how such "mid-level" representations relate to higher-level aspects of visual processing, and to our eventual conscious percepts. Must they always correspond, or may higher-level processes sometimes revise or even overrule the mid-level correspondences computed by the object-file system? Here we address such questions, by exploring the behavior of object files in ambiguous motion events.

1.3. Bouncing vs. streaming

Perhaps the ideal ambiguous stimulus with which to explore such aspects of object files is the *bouncing/streaming* display, first studied by Metzger (1934), Michotte (1946/1963, Experiments 24 and 97), and Julesz (1959, cited in Julesz, 1995, p. 50). In one variant of this phenomenon (depicted in Fig. 2), two objects (X and Y) start out at the upper corners of a display, and then each move diagonally downward, meeting in the center, after which they continue to move to the lower corners of the display. This event can be perceived in two distinct ways, each of which provides a different answer to the question of 'which went where?': either the two objects are seen to travel rotated-V-shaped paths, colliding in the center as if they had 'bounced' off each other (Fig. 2b); or, both objects are seen to traverse entirely linear diagonal paths, and are seen to 'stream' past each other (Fig. 2c).

Though most observers tend to perceive streaming in such displays (Bertenthal, Banton, & Bradbury, 1993; Sekuler & Sekuler, 1999), several factors can increase the likelihood of seeing bouncing. For example, observers are more likely to perceive bouncing: when the objects decelerate before overlapping (Sekuler & Sekuler, 1999); when unambiguous properties such as color or shape explicitly bounce (Feldman & Tremoulet, under review); when attention is drawn away from the point of overlap (Watanabe & Shimojo, 1998); when the objects pause briefly while fully overlapped (Sekuler & Sekuler, 1999); when the objects' speed profiles contain nonlinearities (Bertenthal et al., 1993); when an additional unrelated event occurs suddenly at the moment of overlap (Watanabe & Shimojo, 2001a); or when an auditory event coincides with the moment of overlap (Scheier, Lewkowicz, & Shimojo, 2003; Sekuler, Sekuler & Lau, 1997; Watanabe & Shimojo, 2001b).

The bouncing/streaming phenomenon is an ideal ambiguous stimulus with which to study the connections between object files and conscious percepts for several reasons: (a) It is an inherently dynamic stimulus, in which object-based effects are easily distinguished from location-based effects; (b) It is perceptually bistable, allowing for two possible answers to the question of 'which went where?'; (c) Relatively straightforward manipulations can increase the incidence of conscious bouncing or streaming percepts; and (d) though initially developed as a probe only of conscious percepts, it is readily adaptable for use with the object-reviewing paradigm.

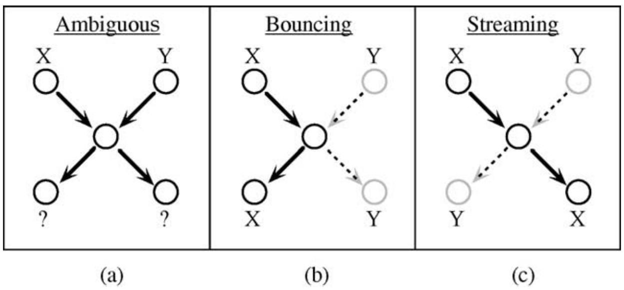

Fig. 2. (a) The ambiguous bouncing/streaming stimulus, which can be perceived as (b) *bouncing*, wherein objects X and Y move towards each other, make contact and then bounce off each other; or as (c) *streaming*, wherein objects X and Y each move in linear paths across the display, passing in the middle.

1.4. The current studies

In the experiments reported below, we employ several variants of the bouncing/streaming phenomenon in which we collect two dependent measures: a conscious percept, and an OSPB obtained via the object-reviewing paradigm. Moreover, we often collect these two responses on the very same trials. This method allows us to directly measure the degree to which object files always solve correspondence problems in the same manner as is reflected in conscious percepts. To our knowledge, this is the first study which has explicitly contrasted these two measures in ambiguous motion events. In Experiment 1 we report an initial demonstration that object files and conscious percepts can diverge: OSPBs can bounce, even when conscious percepts stream. We then replicate and extend this unexpected finding in Experiments 2–5 to rule out several alternate explanations.

2. Experiment 1: perceived streaming

Our adaptation of the object-reviewing paradigm for use in bouncing/streaming displays was straightforward: On each trial, two discs initially appear above the center of the display, one to the left and one to the right. A single letter then briefly appears in each of the discs, after which the discs move downward on diagonal paths so that they fully overlap at the center of the display (Fig. 3a). Without any pause, the discs then continue downward on diagonal paths, stopping below the center with one to the left and one to the right—such that it is ambiguous whether they streamed past or bounced off each other. Subsequently, a single letter appears in one of the two discs and observers respond as quickly as possible whether this target letter is the same as *either* of the two initial preview letters.

This task represents a modification of the object-reviewing paradigm, first employed by Kruschke and Fragassi (1996), and subsequently used successfully in other experiments in our lab (Mitroff et al., 2004; Noles et al., in press). The voice-key response method used in the initial object-reviewing studies yielded especially small and fragile effects (often less than 10 ms). We think this was due to the unfortunate fact that this task allows observers to completely ignore the preview letters—something that may become increasingly likely as the experiment progresses. (Indeed, it is possible to perform with perfect accuracy on the voice-key version even if you shut your eyes during all preview displays.) Accordingly, the modification we employ here, wherein observers simply press a key on each trial to indicate whether the final letter matched *any* of the initially presented letters, requires observers to attend to the initially presented letters, but still allows us to measure OSPBs (since observers make the same 'match' response to congruent 'same-object' matches and incongruent 'different-object' matches). Moreover, this method allows us to *measure* whether observers attended to the initial display (via their accuracy scores), unlike voice-key responses which are typically always correct. These modifications result in larger and more robust OSPBs.

As applied to the bouncing/streaming stimulus, this modified object-reviewing paradigm essentially allows us to measure whether the object files corresponding to the two objects bounce or stream. As with all object-reviewing experiments, an OSPB

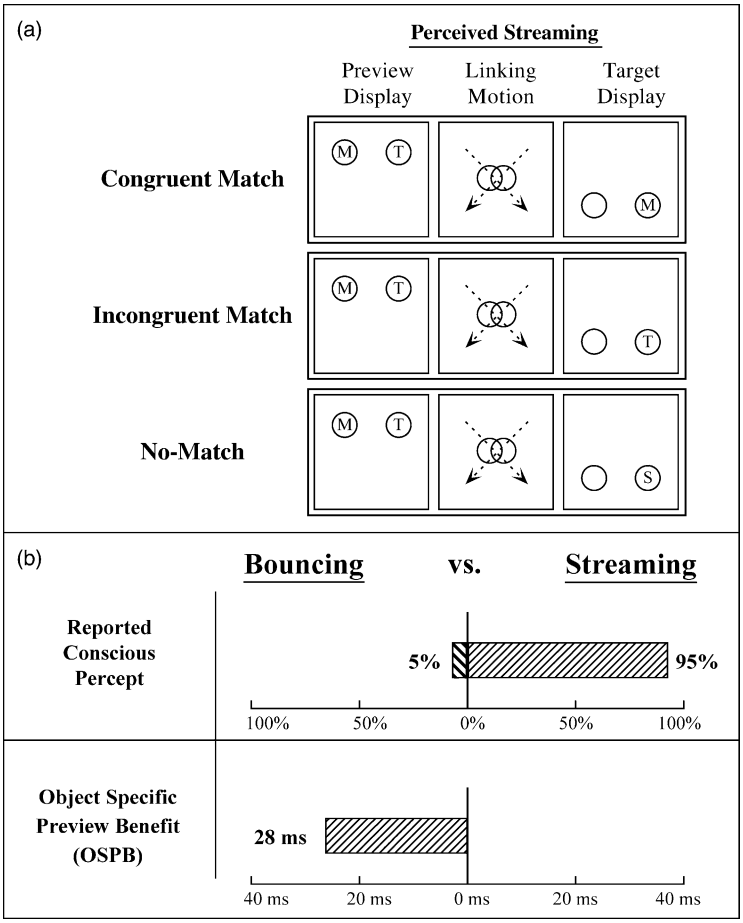

Fig. 3. Depictions of the trial types (not to scale) and results of Experiment 1, testing whether conscious percepts and OSPBs correspond or diverge in bouncing/streaming displays. (a) In each case observers simply responded as quickly as possible whether the final target letter appeared *anywhere* in the initial preview display on that trial. The preview displays were presented for 500 ms, the dynamic linking displays for 400 ms, and the target displays remained until response. Here we depict examples of Congruent-match, Incongruent-match, and No-match trials. (The left preview object is always probed in these depictions, assuming a streaming interpretation, but in the experiment the probe location was counterbalanced and randomized, and the final probe letter could appear in either target location.) (b) The top panel depicts the streaming bias in observers' conscious reports (95% of the trials) and the bottom panel depicts the robust OSPB for bouncing.

between a temporal pair of objects indicates that the visual system has treated them as stages of the same persisting individual. Thus if observers are faster to respond when the preview letter from the upper right reappears as the target letter in the lower right—relative to either (a) when it reappears in the lower left, or (b) when the other preview letter reappears in the lower right—we can infer that the object files bounced. Alternatively, if observers respond faster when the upper right preview letter reappears as the target in the lower left—relative to either (a) when it reappears in the lower right, or (b) when the other preview letter reappears in the lower left—we can infer that the object files streamed. OSPBs are implicated in either case, since these comparisons all involve target letters which appeared in the preview display, thus factoring out any display-wide priming.

In this first experiment, after each object-reviewing response, observers also reported via a keypress whether they consciously perceived that trial as bouncing or streaming. In fact we induced a strong bias to consciously perceive streaming, by using smooth, constant, and reasonably fast motion. This allows us to provide the clearest possible situation in which to evaluate the relationship between conscious percepts and object files. (In contrast, it is much more difficult to tune the displays to produce exactly half bouncing and streaming percepts, and even then there will be considerable variance both within and across observers, as well as unwanted temporal dependency effects in sequential responses, etc. Thus we opted simply to push conscious percepts to one extreme, and then test whether the object files were similarly affected.) The critical question we ask in this study is then whether object files will also stream, along with observers' conscious percepts.

2.1. Methods

2.1.1. Participants

Thirteen Yale University undergraduates participated in a single 30-min testing session for course credit or pay. The data from one observer were eliminated because the average response time to the speeded 'Match/No-match' response was more than two standard deviations from the overall mean. All analyses were performed with the data from the remaining 12 observers.

2.1.2. Apparatus and materials

The displays were presented on a Macintosh iMac computer using custom software written using the VisionShell graphics libraries (Comtois, 2003). 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 discs (each 2° in diameter) presented as black outlines on a white background, with their centers 2.49° above the center of the display, and 2.49° to the right or left of the display, respectively. When the discs moved, they traveled along linear diagonal paths at 17.5°/s, moving a total of 7.03° such that they ended with their centers 2.49° below the center of the screen, and 2.49° to the right or left of the display, respectively. (Thus the two discs started on the two upper corners of a square and ended on its two lower corners, as in Fig. 2.) The discs were always drawn as outlined rings, such that neither occluded the other when they were superimposed. The preview and target letters were drawn in a black monospaced font subtending 1°, centrally located in the discs. The preview letters were drawn without replacement from the set 'K, M, P, S, T, V' and the target letter was either one of the two preview letters from the given trial (on congruent and incongruent Match trials), or was drawn from the remaining letters from the above set (on No-match trials).

2.1.3. Procedure

The observer pressed a key to start each trial, causing the two discs to immediately appear. After 500 ms, a different preview letter appeared for 500 ms in each of the two discs (Fig. 3a). The discs then began their diagonal motion downward meeting at the center of the display after 200 ms and ending vertically below the center of the display after a total motion duration of 400 ms. The single target letter appeared in one of the two discs (equally often in each) immediately after the motion ended. 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 trials, and half were No-match trials. The target letter matched the right preview letter on half of the Match trials, and the left preview letter on the other half. The target letter appeared equally often in the lower right or left, regardless of condition. Immediately after the speeded Match/No-match response, observers reported via a second unspeeded keypress whether they perceived that trial as bouncing or streaming.[1]

Detailed verbal and written instructions were provided to each observer. Observers were instructed to make the Match/No-match responses as quickly as possible while remaining accurate, and to complete this response before considering their conscious perceptual report. The instructions emphasized that the motion was identical on every trial, but that two percepts were nevertheless possible in principle. We further stressed that observers should never feel compelled to report any particular distribution of percepts: observers knew that they were thus free to report different percepts from trial to trial, or to provide the same response on nearly all trials—whatever matched their visual experience. (In fact the displays were constructed so as to yield a strong streaming bias, and thus in practice the reported percepts were highly uniform. We nevertheless measured such responses on each trial just to be certain that all observers perceived the displays in the intended manner.) After 20 practice trials, observers participated in 240 test trials, presented in a different random order for each observer.

2.2. Results

On average, observers were 96.88% accurate on the Match/No-match question (SD = 2.08%), and reported a 'streaming' percept on 94.99% of trials (SD = 6.86%). All analyses were conducted on only those trials with correct responses made within two standard deviations of that observer's overall mean response time; on average 4.27% of trials per observer were eliminated due to such outlying response times (SD = 1.27%).

Object-specific preview benefits (OSPBs) were defined in terms of the streaming bias, such that trials with streaming preview letters (i.e. a preview letter which reappeared as the target on the opposite side of the display) were categorized as *Congruent* (the mean RT for

[1] Conscious percepts were reported via a dichotomous keypress simply because all percepts fell into one of these categories. Though various manipulations can influence the overall *proportion* of bouncing vs. streaming responses, each individual event is always clearly seen as one or the other.

which was 666.93 ms), and trials with bouncing preview letters (i.e. a preview letter which reappeared as the target on the same side of the display) were categorized as *Incongruent* (the mean RT for which was 638.75 ms). OSPBs are then computed as the difference between Congruent and Incongruent response times: positive OSPBs indicate streaming object files, while negative OSPBs indicate bouncing object files. The data from those trials which were consciously perceived as streaming yielded a significant *negative* OSPB, as depicted in Fig. 3b (Mean = −28.18 ms; $t(11) = 2.98$, $P = .012$; all tests two-tailed), suggesting that the object files had bounced even though the percepts streamed. (Analyzing all trials regardless of the reported conscious percept also yielded a significant negative OSPB, but in this and all other experiments we report analyses for only those trials on which observers reported consciously perceived 'streaming'—always at least 93% of all trials— since doing so is a more conservative test of whether conscious percepts and OSPBs can pull in opposite directions.)

2.3. Discussion

Despite the overwhelming bias to consciously perceive streaming in this experiment, robust preview benefits were seen for *bouncing*. Thus object files and conscious perception can diverge, in at least one type of ambiguous motion pattern. Note that these results do not merely reflect a failure to find significant OSPBs with conscious 'streaming' percepts—as might be the case, for example, if the object-reviewing paradigm simply was not sensitive enough to measure such effects. Rather, the results reflect two opposing highly significant effects: each type of response reveals a robust competing solution to the underlying correspondence problem posed by the ambiguous bouncing/streaming phenomenon—which went where? Frankly, this result surprised us: we had initially expected the two measures to coincide, and were then intending to use OSPBs as the first implicit quantitative way to study the various manipulations that influence the bouncing/streaming phenomenon.

This finding thus poses a challenge to the view that object files and conscious perception are always intimately linked, and raises the question of just how these mid-level representations relate to higher-level aspects of visual processing. Do conscious percepts of streaming reflect further elaboration of 'bouncing' correspondences initially computed by the object-file system—a type of revision or over-riding by later yet-to-be-specified processes? Or, do these two systems simply function independently here, with the conscious percepts being computed by entirely different processes? Before returning to such critical questions in the General Discussion, we first sought to replicate this effect, and to ensure that these results do in fact reflect object-specific processing.

3. Experiment 2: unambiguous translation

One possible alternate explanation is that the results of Experiment 1 do not in fact reflect an *object-specific* preview benefit, but rather a '*side-specific*' preview benefit: observers are responding faster not when the preview and target letters occur on the same object per se, but when they occur on the same side of the display. Thus target letters on the right side engender faster responses when they match right-side preview letters, and likewise target letters on the left side engender faster responses when they match left-side preview letters. Even though our displays do not have perceptually salient vertical axes and, to our knowledge, no such 'side-specific' priming effects have been reported, this remains a possible alternative. We sought to test this hypothesis, given the counterintuitive nature of the results of Experiment 1. In this experiment the two discs still moved along the corners a square, but their paths did not intersect: each disc translated horizontally—one from the upper-right position to the upper-left, and the other from the lower-left to the lower-right (Fig. 4a).

If the response time differences of Experiment 1 reflect only a 'side-specific' preview benefit, then a similar pattern of results might be found here: observers should respond

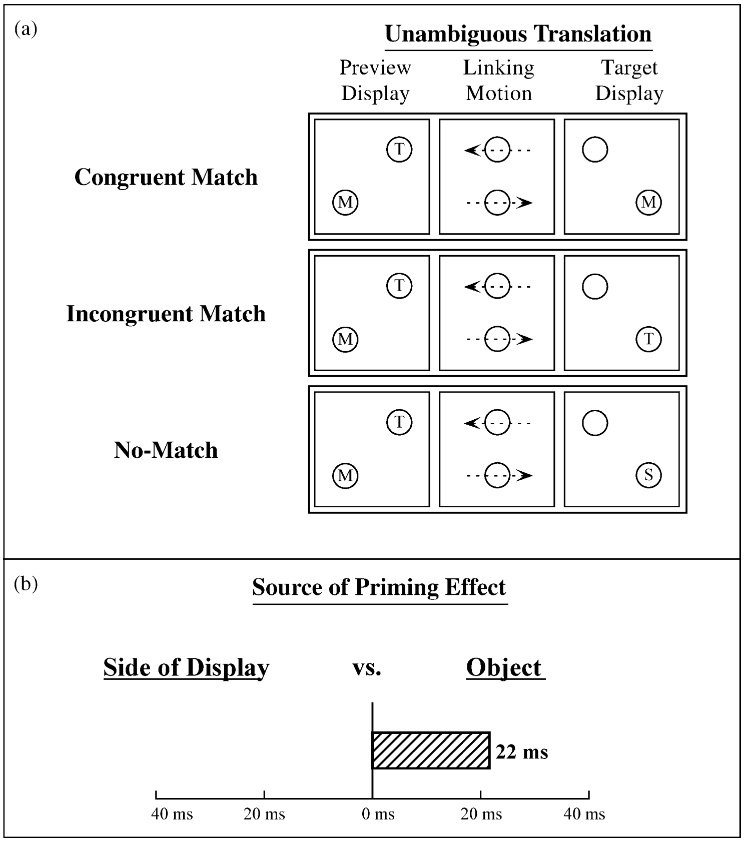

Fig. 4. Depictions of the trial types (not to scale) and results of Experiment 2, testing for side-specific processing effects. (a) Here we depict examples of Congruent-match, Incongruent-match, and No-match trials. (Congruency is always depicted here in terms of the lower-left preview object being probed, but in the experiment the probe location was counterbalanced and randomized, and the final probe letter could appear in either target location.) (b) The displays were constructed such that objecthood and side-of-screen were pitted against each other, and the obtained preview benefit was determined by objecthood, not by the side of the display.

faster (a) when a target letter in the upper-left position matches the preview letter from the lower-left (relative to the preview letter from the upper-right); and (b) when the target letter in the lower-right matches the preview letter from the upper-right (relative to the lower-left). Alternatively, if the results of Experiment 1 do reflect bona-fide object-specific effects, then object-specific effects should also exist here: responses to target letters should always be faster when they match their horizontal counterparts, given that all motion was in the horizontal plane. (Note that with this display it makes little sense to measure conscious percepts, since there is no ambiguity.)

3.1. Method

This experiment was identical to Experiment 1 except as noted here.

3.1.1. Participants

This experiment was run with a new set of 13 observers, each of whom participated in a 20-min testing session. The data from one observer were eliminated since the overall accuracy on the Match/No-match questions was more than two standard deviations below the overall group mean. All analyses were based on the data from the remaining 12 observers.

3.1.2. Materials and procedure

One disc started in the upper right position and the other started in the lower left, and each disc moved in a linear horizontal path across the display such that the upper-right disc ended in the upper-left position and the lower-left disc ended in the lower-right. As depicted in Fig. 4a, this resulted in nonintersecting motion trajectories; because these paths were unambiguous, observers did not report their conscious percepts.

3.2. Results and discussion

On average, observers were 96.15% accurate on the Match/No-match question (SD = 2.08%). Only correct responses made within two standard deviations of each observer's overall response time were used for the remaining analyses; on average 3.92% of trials per observer were eliminated due to such outlying response times (SD = 0.93%).

Observers responded faster when the target letter matched the preview letter presented in the same object (552.09 ms), relative to when it matched the preview letter presented in the other object (574.01 ms; $t(11) = 3.05$, $P = .011$). Because side-of-screen and objecthood were in perfect opposition in this experiment, this reflects a 21.92 ms OSPB (Fig. 4b) which cannot be explained as a side-specific preview benefit—suggesting that the results of Experiment 1 do in fact reflect object-specific processing rather than side-specific processing. Furthermore, this positive OSPB provides a context in which to evaluate the results of Experiment 1, suggesting that the negative OSPB from Experiment 1 does not reflect some sort of 'reversed' preview benefit somehow elicited by our paradigm, but rather object files tracking the bouncing objects.

4. Experiment 3: rotated perceived streaming

Though the previous experiment ruled out a simple alternate explanation based on 'side-specific' preview benefits, it also differed from Experiment 1 in another possibly important way: the motion paths in Experiment 2 were completely unambiguous, whereas those Experiment 1 were ambiguous and gave rise to a correspondence problem. Thus it remains possible in principle that a side-specific bias could operate only under conditions of ambiguity. Though nobody to our knowledge has ever observed such an effect (with or without ambiguity), it nevertheless seemed worth ruling out this possibility, given the counterintuitive nature of our primary finding in Experiment 1.

To test this 'ambiguity-induced side-specific processing' alternate explanation, we simply rotated the displays of Experiment 1 counterclockwise by 90°. Thus we still employ an ambiguous bouncing/streaming stimulus, but now with the objects moving from left-to-right rather than from top-to-bottom (Fig. 5a). This display removes all possibility of 'left-to-left' or 'right-to-right' benefits, of the sort that might be expected due to a hemispheric bias, since the two discs in both the initial and final configurations are vertically centered. In all cases, we still expect observers to consciously perceive streaming in this display, and the critical question concerns the nature of the resulting OSPB. If side-specific processing did indeed fuel the effects in Experiment 1, we would expect to find no OSPB in this experiment. If the results of Experiment 1 are in fact due to object-specific processing, however, we would still expect to observe a significant OSPB—either for streaming (if object files now match conscious percepts) or for bouncing (which would replicate Experiment 1).

4.1. Method

This experiment was run with a new set of 12 observers and was identical to Experiment 1 except that the displays were rotated 90° counterclockwise such that the preview letters appeared on the left of the display and the target letter on the right (see Fig. 5a).

4.2. Results and discussion

On average, observers were 96.56% accurate on the Match/No-match question (SD = 1.85%), and reported a 'streaming' percept on 96.50% of trials (SD = 7.41%). As in Experiments 1 and 2, all analyses were conducted on only those trials with correct responses made within two standard deviations of that observer's overall mean response time; on average 4.13% of trials per observer were eliminated due to such outlying response times (SD = 1.48%).

As in Experiment 1 (and Fig. 5), OSPBs were defined in terms of the streaming bias, such that trials with streaming preview letters (e.g. a preview letter which initially appeared on top, then later reappeared on the bottom) were categorized as *Congruent*, and trials with bouncing preview letters (e.g. a preview letter which initially appeared on top, then later reappeared on the top) were categorized as *Incongruent*. OSPBs are then again computed as the difference between Congruent and Incongruent response times:

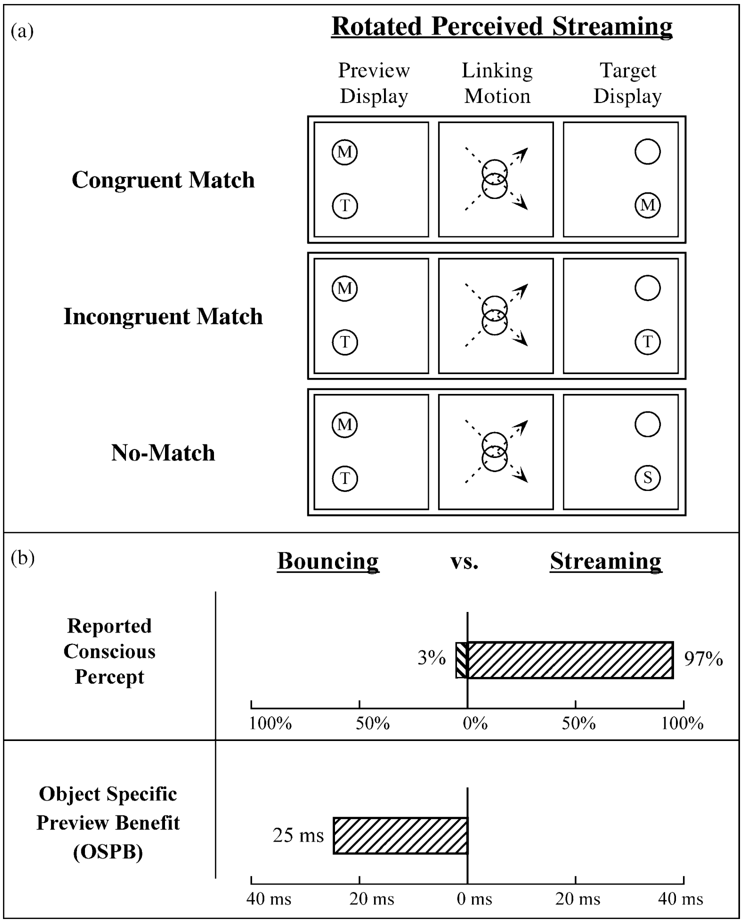

Fig. 5. Depictions of the trial types (not to scale) and results of Experiment 3, testing for side-specific processing under conditions of ambiguity. (a) Here we depict examples of Congruent-match, Incongruent-match, and No-match trials. (The top preview object is always probed in these depictions, assuming a streaming interpretation when computing congruency, but in the experiment the probe location was counterbalanced and randomized, and the final probe letter could appear in either target location.) (b) The top panel depicts the streaming bias in observers' conscious reports (97% of the trials) and the bottom panel depicts the robust OSPB for bouncing.

positive OSPBs indicate streaming object files, while negative OSPBs indicate bouncing object files. The data from those trials which were consciously perceived as streaming yielded faster responses for Incongruent trials (611.69 ms) than for Congruent trials (636.38 ms), resulting in a significant *negative* OSPB (Mean = −24.70 ms; $t(11)=4.93$, $P<.001$), suggesting that while observers consciously perceived streaming, the object files nevertheless bounced (Fig. 5b). This result, combined with that of Experiment 2, rules out an alternate explanation based on 'side-specific' processing, even one which operates only under conditions of ambiguity.

5. Experiment 4: extended unambiguous translation

Experiment 1 suggested that in some situations object files can 'bounce' while conscious percepts 'stream', and Experiments 2 and 3 ruled out an alternate explanation for this pattern of results based on 'side-specific' processing, which could either occur in all circumstances (Experiment 2) or only under conditions of ambiguity (Experiment 3). In the remaining two experiments we address a second alternate explanation, based on *distance*. Because the objects in our bouncing/streaming displays always move on diagonal paths to and from the corners of a square, the distance between the initial and final positions was always smaller for 'bouncing' paths than for 'streaming' paths. (This can be readily seen, for example, in Fig. 2: bouncing interpretations span one leg of the relevant triangle, while streaming interpretations span the hypotenuse.) So perhaps the results of Experiment 1 do not reflect an *object-specific* preview benefit, but rather a *distance-driven* preview benefit: observers are responding faster not when the preview and target letters occur on the same object per se, but simply when they are presented closer to each other.

In fact, this absolute difference in distances is actually quite small, and the actual distance *traveled* is always identical across conditions. Nevertheless, we test this possibility here simply by replicating Experiment 2 (with unambiguous horizontal translation), with a horizontally extended display—such that the distance traveled by the objects on their horizontal paths is greater than the vertical distance between the two paths (Fig. 6a). If the results of Experiments 1 and 3 were driven by distance and not objecthood, then observers should respond faster (a) when a target letter in the upper-left matches the preview from the lower-left (relative to the preview from the upper-right) and (b) when a target letter in the lower-right matches the preview from the upper-right (relative to the preview from the lower-left). However, if the previous findings reflect true object-specific processing, then observers should respond faster when the target letter matches the preview letter from the same object, in the same horizontal plane.

5.1. Method

This experiment was identical to Experiment 2 except as noted here.

5.1.1. Participants

A new set of 13 observers participated in a single 20-min testing session and one observer's data were eliminated because the overall accuracy was more than two standard deviations below the group mean. All analyses were based upon the data from the remaining 12 observers.

5.1.2. Materials and procedure

As depicted in Fig. 6a, the two discs started and ended 3.54° horizontally from the center of the display such that the upper-left target location was closer to the lower-left preview location (4.98°) than the upper-right preview location (7.07°) and the lower-right

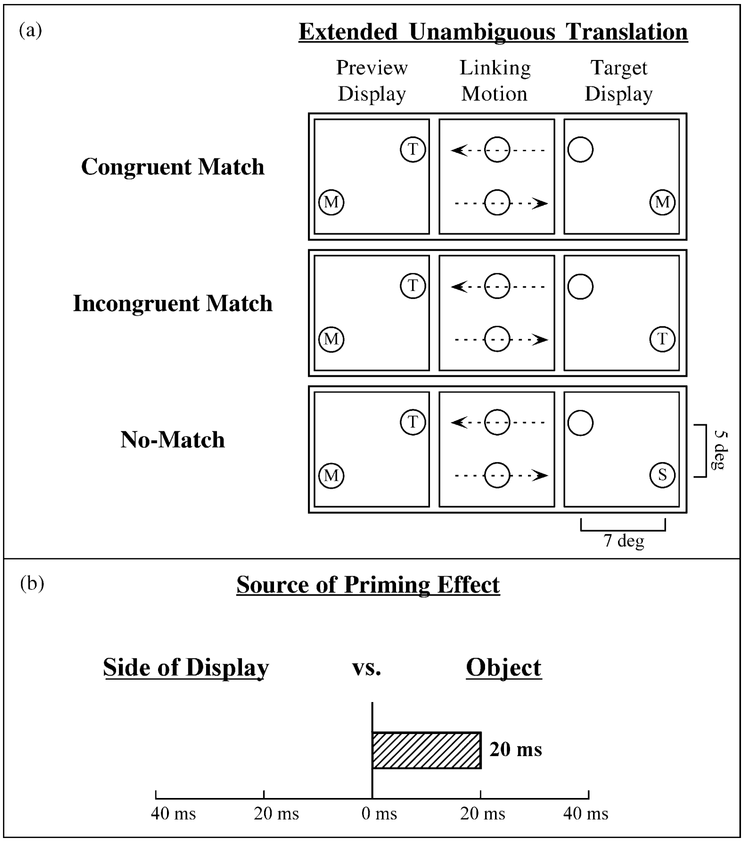

Fig. 6. Depictions of the trial types (not to scale) and results of Experiment 4, testing for distance-based effects. (a) Here we depict examples of Congruent-match, Incongruent-match, and No-match trials. (Congruency is always depicted here in terms of the lower-left preview object being probed, but in the experiment the probe location was counterbalanced and randomized, and the final probe letter could appear in either target location.) The horizontal separation between the objects was greater than their vertical separation. (b) The preview benefit was determined by objecthood, not by distance.

target location was closer to the upper-right preview location (4.98°) than the lower-left preview location (7.07°).

5.2. Results and discussion

On average, observers were 94.62% accurate on the Match/No-match question (SD = 3.37%). Only accurate responses made within two standard deviations of each observer's overall response time were used for the remaining analyses; on average 4.27% of trials per observer were eliminated due to such outlying response times (SD = 0.85%).

Observers responded faster when the target letter matched the preview letter presented in the same object (527.18 ms), relative to when it matched the preview letter presented in the other object (546.96 ms), yielding a significant OSPB of 19.78 ms ($t(11) = 3.20$,

$P = .008$; see Fig. 6b). Because objecthood was in direct opposition to proximity in this experiment, the response time differences must reflect an object-specific preview benefit rather than a simple distance-based effect.

6. Experiment 5: asymmetric perceived streaming

Though the previous experiment ruled out an alternate explanation based on proximity, it also differed in another possibly important way: as in Experiment 2, the motion paths were completely unambiguous, whereas those in Experiments 1 and 3 were ambiguous and gave rise to correspondence problems. Thus it remains possible in principle that a simple bias based on proximity could only operate under conditions of ambiguity. To test this 'ambiguity-induced distance-driven bias' alternate explanation, we modified the bouncing/streaming display from Experiment 1 such that the two streaming-paths were of different lengths, as in Fig. 7. Which diagonal had the shorter path was randomly varied across trials, such that the distance between the initial and final positions of the *shorter* streaming-path was smaller than the distance between the initial and final positions of either bouncing-path—and all of the analyses in this experiment compared the bouncing paths to only the shorter streaming path (ignoring the longer streaming path). Thus, in contrast to the earlier experiments, the preview and target positions are now farther away in the bouncing interpretation than in the streaming interpretation. The actual sequence of each trial was otherwise identical to that of Experiment 1, as depicted in Fig. 8a.

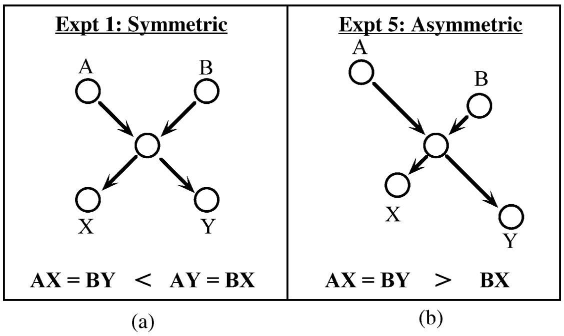

Fig. 7. A comparison between the displays used in Experiments 1 and 5. (a) The discs in Experiment 1 moved along symmetrical paths with equal vertical and horizontal distances between the preview and target locations, such that the initial and final positions were closer under a bouncing interpretation than under a streaming interpretation. (b) In Experiment 5, one of the preview locations and its corresponding 'streaming' target location were further from the center of the display than the other pair. (Half of the actual displays looked like this depiction, and half were its mirror image, such that the other diagonal had the longer streaming-path.) This asymmetry produced Incongruent, or 'bouncing', paths (i.e. AX and BY) that were longer than one of the two Congruent, or 'streaming', paths (i.e. BX). For all OSPB measures the Incongruent-match trials were contrasted with only those trials with the shorter of the two Congruent-match paths, in order to unconfound objecthood and proximity under conditions of ambiguity.

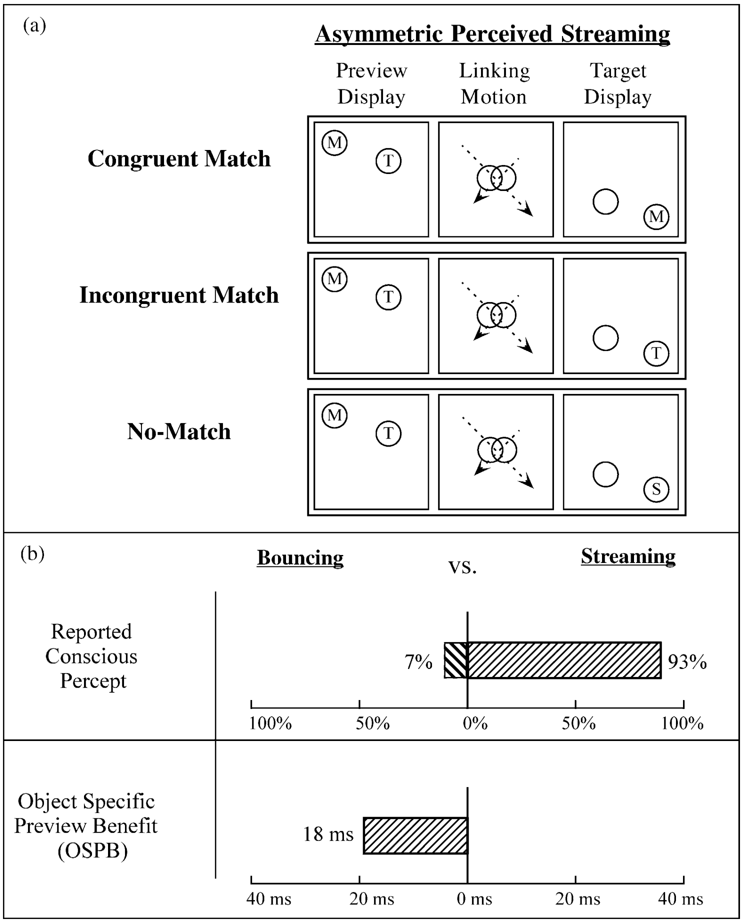

Fig. 8. Depictions of the trial types (not to scale) and results of Experiment 5, testing for distance-based effects under conditions of ambiguity. (a) Here we depict examples of Congruent-match, Incongruent-match, and No-match trials, with two simplifications: (1) the left preview object is always probed in these depictions, assuming a streaming interpretation when computing congruency, but in the experiment the probe location was counterbalanced and randomized, and the final probe letter could appear in either target location; and (2) the left preview location and the right target location appear farther from the center of the display in these depictions, but in the experiment the right preview location and the left target location could also appear farther away (the choice being counterbalanced and randomized across trials). (b) The top panel depicts the streaming bias in observers' conscious reports (93% of the trials) and the bottom panel depicts the robust OSPB for bouncing.

If distance-based processing did indeed fuel the effects in Experiment 1, we would expect to find OSPBs in this experiment which correspond to shorter streaming interpretations—which of course could also indicate an object-specific effect which perfectly tracked observers' conscious percepts, since this display also gave rise to a strong streaming bias. OSPBs corresponding to the longer bouncing interpretations, however, cannot be so explained, and must instead reflect object-based processing and a replication of the divergence between object files and conscious percepts observed in Experiments 1 and 3.

6.1. Method

This experiment was identical to Experiment 1 except as noted here.

6.1.1. Participants

A new set of 15 observers participated in a single 30-min testing session and one observer's data were eliminated because the overall response time was more than two standard deviations from the group mean. All analyses were based upon the data from the remaining 14 observers.

6.1.2. Materials and procedure

The bouncing/streaming display from Experiment 1 was altered such that one streaming-path was shorter than the other—and, critically, was shorter than either bouncing-path. On half of the trials, the left preview disc appeared farther from the center of the display (offset by 4.73° vertically and horizontally) than the right preview letter (offset by 2.49° vertically and horizontally) and the right target disc traveled to a correspondingly farther point from the center of the display than the left target disc (as depicted in Fig. 7). For the other half of the trials, the right preview disc and the left target disc were farther from the center of the display than the left preview and right target discs. Thus the distance between the initial and final positions of bouncing-paths was always longer than for one of the two streaming paths on each trial (7.56 vs. 7.04°). All trials were presented in a random order. Each disc took 500 ms to reach the point of intersection and another 500 ms to reach its final position. Thus, the two discs moved at different speeds: the disc which traveled the shorter path moved at 7.04°/s, while the disc which traveled the longer path moved at 13.39°/s. This created the percept of one object moving both farther and faster than the other, and the display still induced a strong streaming bias.

6.2. Results and discussion

On average, observers were 95.98% accurate on the Match/No-match question (SD = 2.74%), and reported a 'streaming' percept on 93.20% of trials (SD = 11.96%). All analyses were conducted on only those trials with correct responses made within two standard deviations of that observer's overall mean response time; on average 4.08% of trials per observer were eliminated due to such outlying response times (SD = 1.10%).

As in Experiments 1 and 3, OSPBs were defined in terms of the streaming bias, such that trials with streaming preview letters (i.e. a preview letter which reappeared as the target on the opposite side of the display) were categorized as *Congruent*, and trials with bouncing preview letters (i.e. a preview letter which reappeared as the target on the same side of the display) were categorized as *Incongruent*. OSPBs are then again computed as the difference between Congruent and Incongruent response times, but now using only the shorter of the two streaming interpretations: positive OSPBs indicate streaming object files, while negative OSPBs indicate bouncing object files. Observers responded faster on the Incongruent trials

(AX and BY in Fig. 7b; Mean = 534.52 ms) than on the shorter-distance Congruent trials (BX in Fig. 7b; Mean = 552.39 ms), resulting in a significant *negative* OSPB of −17.87 ms ($t(11) = 2.33$, $P = .036$, see Fig. 8b). Because these Congruent trials involved a shorter distance than the Incongruent trials, persisting objecthood and proximity were in direct opposition, and the results could not have been driven by proximity. Rather, these results must reflect object-specific processing, and as such they provide further evidence for the divergence between object files and conscious perception.

7. General discussion

In the past decade, researchers in several areas of psychology have increasingly realized the importance of mid-level visual representations of persisting objects, and have often appealed to object files as a framework within which to theorize about the nature of mid-level vision. Despite the wealth of theorizing about object files, their underlying nature has remained obscure, in the sense that we do not have a clear understanding of how this mid-level computation of persisting object representations relates to our eventual conscious percepts of persisting objects in the world. In this sense, research on object files has rarely fleshed out the precise meaning of "mid-level", and has often simply assumed that object files determine our conscious percepts. Here we directly tested this assumption, and discovered at least one instance in which object files diverge from conscious perception in the representation of 'which went where'.

In Experiment 1, we manipulated the spatiotemporal parameters of a bouncing/streaming display such that observers consciously perceived streaming on a vast majority of trials. Nonetheless, the same observers showed robust object-specific preview benefits (OSPBs) for bouncing in the very same trials: their percepts streamed, even while the object files bounced.[2] In Experiments 2–5 we replicated this surprising result two more times, and ruled out alternate explanations based on possible hemispheric biases (Experiments 2 and 3) and proximity (Experiments 4 and 5), both with and without ambiguous correspondence problems.

7.1. How do object files relate to conscious percepts?: four possibilities

The five experiments in this paper serve as a case study of how object files and conscious perception can diverge, and thus show that the impletion operation is not always a direct link from mid-level representations to conscious visual experience. So what is

[2] In another study we induced a conscious bouncing percept and found that the object files also bounced (OSPB = 20.95, $t(10) = 2.41$, $P = .037$), but such experiments prove more problematic to interpret. Below we consider the possibility that the results of Experiments 1, 3 and 5 reflect a global bouncing bias (driven perhaps by a solidity constraint) which operates in the object-file system. Thus, finding that object files bounce when conscious percepts also bounce can either indicate a match between object files and conscious percepts, or simply the operation of the same bouncing bias—operating independently of the conscious percepts. In any case, the current experiments stand alone as a case study of how object files and conscious percepts *can* diverge in at least one situation.

the relationship between object files and conscious perception? Our results are consistent with at least four possibilities, which we briefly discuss in turn: complete divergence, downstream overriding, divergence only under ambiguity, and divergence only under solidity violations.

7.1.1. Complete divergence

The assumption that object files determine conscious percepts may be entirely mistaken: the representation of persisting objects and the computation of 'which went where' may occur twice, in independent processing streams—once with object file representations and once in a distinct system which determines our conscious percepts. Such a scheme is certainly possible, and similar dissociations have been observed in other areas of visual processing. For example, some researchers have provided evidence for the operation of separate processing streams in 'vision for action' and 'vision for perception', which may sometimes compute conflicting interpretations of the same image data—such that illusions arise in one system, but not the other (Haffenden, Schiff, & Goodale, 2001; Milner & Goodale, 1995). However, these other proposed dissociations rest on a wealth of evidence and principled distinctions between action and awareness, whereas object files are already tightly theoretically linked to conscious perception. Without some independent evidence for such a dissociation, it seems unlikely that—and entirely unclear why—the same underlying computations of persisting objecthood should always be duplicated in this way.

Most object-reviewing experiments do not directly speak to this possibility, since they typically involve unambiguous object trajectories: different solutions to a correspondence problem will not arise when there is no correspondence problem. However, one previous study from our lab (Mitroff et al., 2004) does suggest that object files and conscious percepts do not always diverge in complex motion displays. In this study, a single object split into two new objects during its motion, such that observers clearly perceived a single object splitting into two. Consistent with this conscious percept, an OSPB was found for both resulting objects—such that the object file seemed to have, in some sense, survived the split and become associated with both resulting objects. This wasn't a preordained result: the object file could have remained bound to only one of the objects, even while conscious perception represented the 'split'. Instead, this is one case where object files and conscious percepts converged. Of course, this is consistent with both convergence and with separate systems (which in this case happen to compute object persistence in similar ways). In any case, it seems premature to conclude that object files have nothing whatsoever to do with conscious percepts, even if that possibility is consistent in principle with the current results.

7.1.2. Downstream overriding

Another possibility is that object files do contribute to conscious percepts, but not directly: they may play a critical role in computing heuristic solutions to correspondence problems, but other processes can still intervene before the formation of our conscious percepts—and in certain cases these processes may tune or even override the matches computed in the object-file system. Note that this would still be entirely consistent with the characterization of object files as 'mid-level' processes—just not as 'penultimate-level' processes, as is sometimes assumed.

This interpretation highlights the fact that visual representations are themselves dynamic entities, which are not always computed in a single step, but rather evolve in the course of online visual processing. This online evolution of visual representations is a ubiquitous feature of visual processing. For example, the amodal completion of a partially occluded disc behind a square does not occur in a single step: rather, it is first represented as a mosaic 'pac-man' which is then only gradually completed into a disc (Sekuler & Palmer, 1992). (Note that even this is a 'mid-level' representation, however, since the initial segmentation of the scene into distinct coherent regions has already occurred.) Moreover, studies which interrupt this processing at various times can reveal the continuous evolution of this representation, as the contour is gradually completed over time (Rauschenberger & Yantis, 2001). In such studies, the critical stimulus is represented in two conflicting ways (an uncompleted 'mosaic' and a completed disc) at different points in visual processing, and only one of these corresponds to our conscious visual percept of amodal completion.

Perhaps object files work in a similar way: they are mid-level representations which compute initial heuristic solutions to visual correspondence problems, and while these solutions may often survive to determine our conscious percepts, they may also sometimes be overridden by downstream processing. However, this interpretation rings hollow without a more detailed theory of just when such overriding will occur. In fact, it seems possible that our bouncing/streaming displays lead to overriding for either (or both) of two unique principled reasons: their use of ambiguity, and the apparent violation of solidity constraints.

7.1.3. Divergence under ambiguity

One possibility is that object files and conscious perception will diverge only for objects whose spatiotemporal paths are ambiguous—i.e. whose paths present correspondence problems. This is consistent with the characterization of object files as a quick-and-dirty heuristic system: their computations of object persistence may nearly always survive to determine our conscious percepts, *except* in those cases in which two otherwise-identical objects completely intersect during their motions. Object files may not be able to reliably handle such situations, such that downstream processing (gated by focal attention?) may be required in just these cases to alter the interpretation of 'which went where'.[3]

Note that this interpretation is consistent with all object-reviewing experiments to date. This is the first study to pose such an ambiguity, and the first study to find a divergence between object files and conscious percepts. All previous object-reviewing studies—including the original experiments of Kahneman et al. (1992) and Experiments 2 and 4 of

[3] This interpretation rhetorically assumes that the match reflected in our conscious percepts is always 'correct', whereas the object-file system's match is wrong. Note that this needn't be the case, though: there could be situations in which the object-file system's solution is actually correct, and the conscious percepts are mistaken. Indeed, this is what happens in action/awareness dissociations, in which you are fooled by visual illusions, but your hands are not (Haffenden et al., 2001).

this paper—involve unambiguous trajectories, such that the object-file system's computation of object persistence over time survives to determine our conscious percepts.

If this interpretation is correct, then our displays in Experiments 1, 3 and 5 in this paper are ecologically invalid in an important sense, since such pure ambiguous situations—with perfect spatial overlap and no featural differences—rarely occur in everyday visual experience. In this way our results are consistent with the possibility that object files will nearly always correspond to conscious percepts both in the real world and in most experiments. The divergence we observe here may thus be theoretically important to our conception of the underlying nature of object files, but may have little practical upshot.

7.1.4. Divergence under solidity violations

Another more specific possibility is motivated by the use of object files in the study of cognitive development. Though the object-file framework originated in the study of adults' visual perception, it has become increasingly popular as a potential explanation of some results in the infant cognition literature. Some researchers have suggested that object files may underlie infants' abilities to track objects over time and behind occluders; to track and update the numerosity of small sets of items; and in general to perceive persisting objects (Feigenson, Carey, & Hauser, 2002; Leslie, Xu, Tremoulet, & Scholl, 1998; Scholl, 2001; Wynn & Chiang, 1998). Such explanations are fueled by several surprising analogies between the operation of mid-level vision in adults and infants' behavior on object-cognition tasks (for reviews of these analogies see Carey & Xu, 2001; Scholl & Leslie, 1999).

Within the infant cognition literature, several specific principles have been proposed as being critically involved in the representation of persisting objecthood (Spelke, 1990; Spelke, Phillips, & Woodward, 1995), and recently these principles have been explicitly tested with adults in mid-level vision experiments. For example, one critical principle of object persistence is that of *cohesion*: "objects are connected and bounded bodies that maintain both their connectedness and their boundaries as they move freely" (Spelke et al., 1995, p. 45). In fact, some violations of cohesion appear to destroy infants' enduring object representations (Chiang & Wynn, 2000; Huntley-Fenner et al., 2002), and many cognitive scientists have explicitly taken cohesion to be the most important constraint on what it means to be a persisting object (Bloom, 2000; Pinker, 1997). Recently, we have studied the operation of this principle using object-reviewing (Mitroff et al., 2004) and multiple-object tracking (vanMarle & Scholl, 2003), and found that even adults' visual processing is affected in critical ways by cohesion violations.

Perhaps we can think of the current studies in a similar way, as evidence of how object files are disrupted by violations of the principle of *solidity*: that "objects move only on unobstructed paths: no parts of two distinct objects coincide in space and time" (Spelke, Breinlinger, Macomber, & Jacobson, 1992, p. 606). Infants as young as 5 months of age show an appreciation of this solidity constraint and will tend to look longer in certain paradigms when it is violated (Baillargeon, 1999; Baillargeon, Spelke, & Wasserman, 1985). Perhaps the object-file system also realizes a principle of this type: the system may embody a structural constraint that prevents two object files from corresponding to the same place at the same time. (Remember that our objects were always drawn as transparent rings, without any of the T-junctions or other depth cues that would prevent the objects from being seen to overlap in 2D interpretations.) If the object-file system also adheres to an indirect solidity principle of this type (as it does to the cohesion principle), and interprets the current displays as two objects being in the same plane, then it *should* compute a ‘bouncing’ interpretation, since streaming would involve two solid objects passing through each other. This interpretation remains an intriguing focus for further study.[4]

7.2. Conclusions

Whichever of these four interpretations turns out to be correct, this study has demonstrated for the first time that object files and conscious percepts needn’t always converge: at least in some circumstances, additional processes beyond the object-file system can influence our conscious percepts of ‘which went where’. In this sense, this study is a first step in clarifying the underlying nature of object files, by determining their relationship to other aspects of visual processing and awareness.

Acknowledgements

For helpful conversation and/or comments on earlier drafts, we thank Erik Cheries, Hoon Choi, Valerie Kuhlmeier, Alexandria Marino, Nic Noles, Alison Sekuler, Kristy vanMarle, Pamela Yee, and two anonymous reviewers. We also thank Anne Dudley and Alexandria Marino for assistance with data collection. Portions of this work were presented at the 2003 meeting of the *Vision Sciences Society*, Sarasota, FL. BJS and KW were supported by NSF #BCS-0132444. SRM was supported by NIMH #F32-MH66553-01.

References

Baillargeon, R. (1999). Young infants’ expectations about hidden objects: A reply to three challenges. *Developmental Science*, *2*, 115–132.

Baillargeon, R., Spelke, E., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*, 191–208.

Bertenthal, B., Banton, T., & Bradbury, A. (1993). Directional bias in the perception of translating patterns. *Perception*, *22*, 193–207.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

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). VisionShell PPC. [Software libraries]. Cambridge, MA: Author.

---

[4] It is tricky to unconfound solidity violations (Section 7.1.4) and pure spatiotemporal ambiguous correspondence problems (Section 7.1.3), since both require—and are defined by—two objects being in the same place at the same time. However, we are currently pursuing this by using bouncing/streaming displays with featural differences and occlusion, with both infants and adults.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, *13*(2), 150–156.

Feldman, J., & Tremoulet, P. D. (under review). Individuation of objects over time. Submitted for publication.

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

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

Haffenden, A. M., Schiff, K. C., & Goodale, M. A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: Nonillusory effects of pictorial cues on grasp. *Current Biology*, *11*, 177–181.

Henderson, J. (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.

Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of distinct entities. *Cognition*, *85*(3), 203–221.

Julesz, B. (1995). *Dialogues on perception*. Cambridge, MA: MIT Press.

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

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 174–219.

Kruschke, J. K., & Fragassi, M. M. (1996). The perception of causality: Feature binding in interacting objects. In *Proceedings of the Eighteenth 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*(1), 10–18.

Metzger, W. (1934). Beobachtungen über phänomenale Identität. *Psychologische Forschung*, *19*, 1–60.

Michotte, A. (1946/1963). *La perception de la causalite*. Louvain: Institut Superior de Philosophie, 1946. English translation of updated edition by T. Miles & E. Miles, *The perception of causality*, Basic Books. 1963.

Milner, A., & Goodale, M. (1995). *The visual brain in action*. Oxford: Oxford University Press.

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.

Noles, N. S., Scholl, B. J., & Mitroff, S. R. (in press). *The persistence of object-file representations*. Perception & Psychophysics.

Pinker, S. (1997). *How the mind works*. New York: Norton.

Rauschenberger, R., & Yantis, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature*, *410*, 369–372.

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, *6*(3), 233–244.

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.

Sekuler, A., & Palmer, S. (1992). Perception of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General*, *121*, 95–111.

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

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*, 308.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605–632.

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: Clarendon Press.

vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects vs. substances. *Psychological Science*, *14*(5), 498–504.

Watanabe, K., & Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception*, *27*, 1041–1054.

Watanabe, K., & Shimojo, S. (2001a). Postcoincidence trajectory duration affects motion event perception. *Perception and Psychophysics*, *63*, 16–28.

Watanabe, K., & Shimojo, S. (2001b). When sound affects vision: effects of auditory grouping on visual motion perception. *Psychological Science*, *12*(2), 109–116.

Wynn, K., & Chiang, W.-C. (1998). Limits to infants' knowledge of objects: The case of magical appearance. *Psychological Science*, *9*, 448–455.