

Research Report

Divide and Conquer

How Object Files Adapt When a Persisting Object Splits Into Two

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

Yale University

ABSTRACT—*Coherent visual experience requires not only segmenting incoming visual input into a structured scene of objects, but also binding discrete views of objects into dynamic representations that persist across time and motion. However, surprisingly little work has explored the principles that guide the construction and maintenance of such persisting object representations. What causes a part of the visual field to be treated as the same object over time? In the cognitive development literature, a key principle of object persistence is cohesion: An object must always maintain a single bounded contour. Here we demonstrate for the first time that mechanisms of adult midlevel vision are affected by cohesion violations. Using the object-file framework, we tested whether object-specific preview benefits—a hallmark of persisting object representations—are obtained for dynamic objects that split into two during their motion. We found that these preview benefits do not fully persist through such cohesion violations without incurring significant performance costs. These results illustrate how cohesion is employed as a constraint that guides the maintenance of object representations in adult midlevel vision.*

Visual processing begins with an undivided wash of unbound features, and results in the visual experience of discrete objects and events. Accordingly, a tremendous amount of research has explored the principles by which the visual system segments images into units. However, visual experience consists of more than individual snapshots of the world: Observers must also bind individual views of objects into dynamic representations that persist 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.

Address correspondence to Stephen R. Mitroff or Brian J. Scholl, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205; e-mail: stephen.mitroff@yale.edu or brian.scholl@yale.edu.

THE COHESION PRINCIPLE

What does it mean to be a persisting object? Although there has been relatively little exploration of this question in the adult perception literature, it has received a considerable amount of study by cognitive developmental psychologists interested in the nature of the infant's object concept (e.g., Carey & Xu, 2001; Chiang & Wynn, 2000; Huntley-Fenner, Carey, & Solimando, 2002; Spelke, 1990). One of the principles identified by these researchers as being critical is *cohesion*: "Objects are connected and bounded bodies that maintain both their connectedness and their boundaries as they move freely" (Spelke, Phillips, & Woodward, 1995, p. 45). Thus, according to this principle, a feature cluster that splits into two is not a persisting object. In fact, some such cohesion violations appear to destroy infants' enduring object representations (e.g., 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 (e.g., Bloom, 2000; Pinker, 1997).

Inspired by this developmental research, we asked in the present study whether mechanisms of adult midlevel vision—which also have the task of tracking objects over time and motion—also respect the cohesion principle.

OBJECT FILES

Psychologists regularly appeal to two primary types of visual representations: low-level visual features ("It's red," "It's round") and higher-level object types ("It's a duck," "It's a truck"). But observers can easily track persisting objects even when both sorts of information change: There is no doubt that the transformation of a frog into a prince involves a single persisting visual object (Kahneman, Treisman, & Gibbs, 1992). To account for this ability, an intermediate level of representation is required, and the *object-file* framework is perhaps the most popular theory of such representations (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.

Direct evidence for such representations comes from the *object-reviewing* paradigm (Kahneman et al., 1992), depicted in Figure 1. Initially, a number of objects are presented, and distinct letters appear

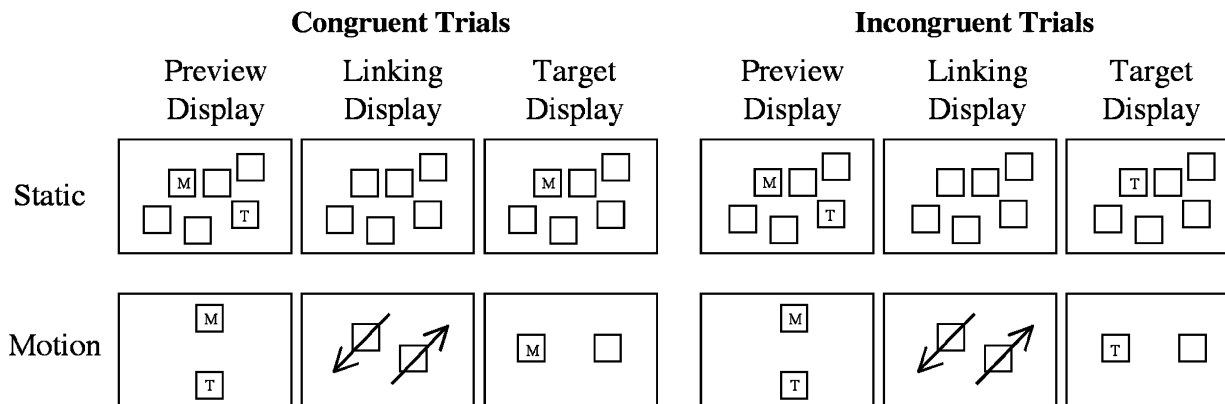


Fig. 1. Sample displays used in the object-reviewing paradigm (Kahneman, Treisman, & Gibbs, 1992). Each trial consists of a sequence of preview, linking, and target displays. In the static cases, the object on which the target letter appears is seen as one of the preview objects, because it is the same object in the same location. The target letter is either the same letter that appeared on that object in the preview display (congruent trials) or the letter that appeared on a different object in the preview display (incongruent trials). In the moving displays, objecthood and location are unconfounded. In each type of display, target naming is facilitated on congruent relative to incongruent trials.

briefly on some of them. The objects then move about the visual display for a brief period, after which a single letter appears on just one of the objects. The subject's task is to name the final letter as quickly as possible. Typically, this response is slightly faster when the letter matches one of the initially presented letters than when it does not (a type of displaywide priming effect). However, subjects are even faster to name the final letter when it is the same letter that initially appeared on that same object, rather than a letter that initially appeared on a different object—an *object-specific preview benefit* (OSPB). This effect can thus be used as an index of persisting objecthood: Manipulations that degrade enduring object representations will result in attenuated OSPBs.

Researchers in both adult vision and cognitive development have often appealed to the object-file framework (for reviews, see Carey & Xu, 2001; Scholl, 2001), and several studies have explored the types of information that can be stored in object files (e.g., Gordon & Irwin, 1996, 2000; Henderson, 1994; Henderson & Anes, 1994), how long object files last (Noles, Scholl, & Mitroff, 2003), and the relation between object files and conscious perception (Mitroff, Scholl, & Wynn, 2003). However, no studies since Kahneman and Treisman's seminal experiments (Kahneman et al., 1992) have explored the principles that cause object files to be constructed, destroyed, or updated. These principles are critical, because they essentially define what "counts" as a persisting object in the object-file framework.

THE CURRENT PROJECT

Given the central place of cohesion in theories of object persistence, we have begun to study the principles that guide the construction and maintenance of object files by focusing on the simplest possible cohesion violation: a moving object that splits into two separate objects (see Fig. 2). In the object-reviewing paradigm, what happens to the object-file representation of an object that splits in this way? There are several possibilities: (a) The cohesion violation might obliterate the object file's contents (i.e., the identity of the initially presented preview letter in that object); (b) the information about the preview letter

might survive the splitting intact, but stay bound to only one of the two resulting objects, indicating that object files cannot themselves split into two; or (c) the object file's contents might essentially be "copied" to both of the two resulting objects. Each of these outcomes would be revealing about the degree to which cohesion is an important determinant of object persistence in midlevel visual processing.

METHOD

Thirty-six Yale University undergraduates participated for course credit or payment. The displays were presented on a Macintosh iMac computer using custom software written using the VisionShell graphics libraries (Comtois, 2003). Observers sat without head restraint approximately 50 cm from the monitor.

Each trial began with two circles (2° in diameter), presented as black outlines on a white background (see Fig. 2). They were drawn 2.5° to the left of the vertical midline, with one 2.5° above and the other 2.5° below the horizontal midline. (Distance measures were calculated from the circles' centers.) After 500 ms, a letter (subtending 1° , drawn in a black monospaced font) appeared in each circle, drawn without replacement from the set *K, M, P, S, T, and V*. After 1 s, these *preview letters* disappeared, and the circles began their motion (always at $5^\circ/\text{s}$, for a total of 1 s). Regardless of condition, each circle first moved 2.5° to the right. At this point, one of the circles (chosen randomly) continued to move in a straight motion for another 2.5° . The other circle's motion depended on the condition. In the *splitting* condition (75% of trials), it also continued to move 2.5° to the right, but at the same time gradually split into two identical circles (one moving upward and one moving downward); until the two resulting circles had completely separated, only their outermost shared contour was drawn. When the motion ended, the three resulting circles were arranged vertically, equally spaced 3.34° apart. In the *trajectory-change* condition (25% of trials), the remaining circle simply followed one of the upward or downward paths from the splitting condition (i.e., moving gradually to the right and either upward or downward, chosen randomly). This condition served as a

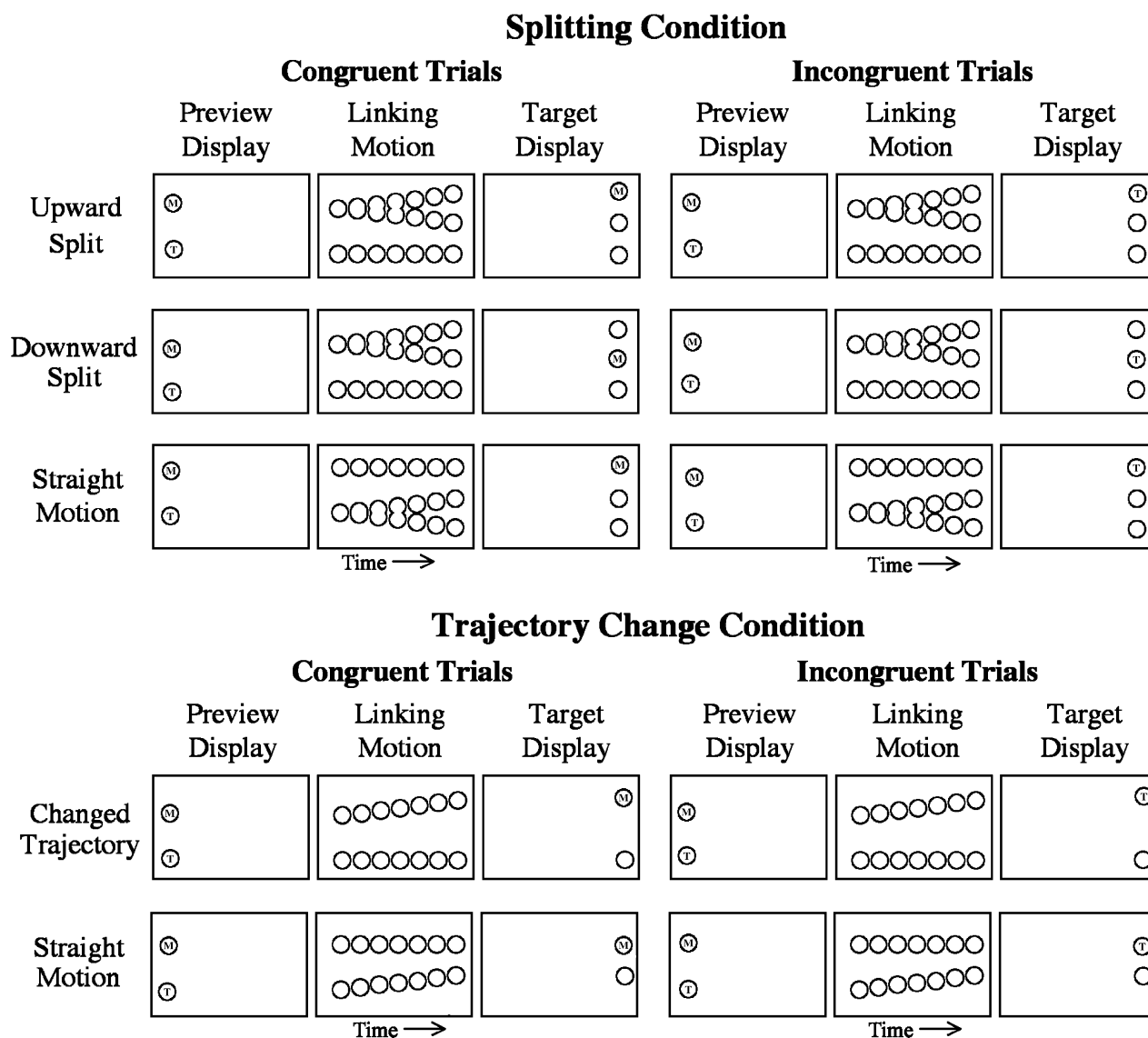


Fig. 2. Depictions of the five trial types in the present study (not to scale). The observers' task was to indicate as quickly as possible whether the final letter had appeared anywhere in the initial preview display on that trial. Objects traveled a straight trajectory, traveled a curved trajectory, or split into two separate objects. The preview and linking displays were presented for 1 s each, and the target displays remained until response. Congruent-match and incongruent-match trials for each trial type are depicted. This illustration simplifies the actual experiment in four ways: First, there are no examples of the no-match trials, in which the final probe letter was not one of the preview letters. Second, in the actual experiment, the final probed letter could appear on any final object, whereas the topmost initial object is probed in all these depictions. Third, upward and downward motions were equally likely in the trajectory-change condition, though only upward motions are depicted here. Fourth, the splitting or trajectory change did not begin until the object had already traversed half of its total path length; these initial linear motions are not depicted in the figure because they did not differ between conditions.

control, involving a trajectory change that was equivalent to the trajectory change in the splitting condition, but without a cohesion violation.

Immediately after the motion ended, a single target letter appeared in a randomly chosen final circle and remained until the participant responded. We used an adaptation of the object-reviewing paradigm (drawn from Kruschke & Fragassi, 1996): Observers made a speeded response, pressing one key to indicate that the target letter was the same as either of the preview letters or another key to indicate that it did not appear in the preview display. (Note that this task requires observers to remember the display, but does not prioritize any object-

based information—as would be the case if they had to base their response on only the letter that had originally appeared in the probed object.) Fifty percent of trials were *no-match* trials, in which the target letter (drawn from the same set as the preview letters) did not appear in either of the original circles. Of the remaining *match* trials, 50% were *congruent matches* (in which the target letter was the same as the preview letter that initially appeared on that circle), and 50% were *incongruent matches* (in which the target letter was the same as the preview letter that initially appeared on the other initial circle). After 20 practice trials, 320 test trials were presented in a different random order for each observer.

TABLE 1
Response Times and Object-Specific Preview Benefits

Condition and motion	Response time (ms)		Object-specific preview benefit (OSPB)
	Incongruent match (different object)	Congruent match (Same object)	
Splitting			
Straight motion	562.97	553.03	No effect (9.94 ms) $t(35) = 1.45, p = .155$
Upward split	551.62	538.58	13.04 ms ^a $t(35) = 2.29, p = .030$
Downward split	560.20	545.40	14.80 ms ^a $t(35) = 2.49, p = .028$
Trajectory change			
Straight motion	572.19	552.15	20.04 ms ^b $t(35) = 2.45, p = .020$
Changed trajectory	570.70	533.95	36.74 ms $t(35) = 4.06, p < .001$

^aThe upward-split and downward-split OSPBs did not differ ($p = .848$). ^bThis OSPB was larger than the OSPBs in both split motions of the splitting condition ($ps < .05$).

RESULTS

Overall accuracy was high (96.99%) and did not differ between conditions (97.08% in the splitting condition and 96.70% in the trajectory-change condition), $t(35) = 0.93, p = .357$ (all tests are two-tailed). All analyses were limited to trials with accurate responses. Trials on which response times were more than 2 standard deviations from the observer's mean were eliminated (4.14% of the trials). Response times indicated a displaywide priming effect: As in many studies of object reviewing, responses on no-match trials (577.04 ms) were significantly slower than responses on either incongruent-match trials (561.68 ms), $t(35) = 2.53, p = .016$, or congruent-match trials (545.19 ms), $t(35) = 5.48, p < .001$, and the magnitude of this general priming did not differ between the splitting and trajectory-change conditions, $t(35) = 0.23, p = .819$.

The comparison of interest was the OSPB, defined as the difference between congruent-match and incongruent-match trials. Faster responses to congruent than incongruent trials would reveal the maintenance of object-specific information, above and beyond displaywide priming. Thus, the critical analyses in this study determined which conditions produced significant OSPBs (see Table 1). In the trajectory-change condition, significant OSPBs were observed for both the object with the changed trajectory and the object that moved in a straight horizontal motion. In the splitting condition, significant OSPBs were observed for both upward-split and downward-split objects, but not for the object that moved with a straight motion.¹ Additional comparisons indicated that the OSPBs for the upward-split and downward-split

motions did not differ, $t(35) = 0.19, p = .848$, though these OSPBs were significantly smaller than (indeed, less than half of) the OSPB for the changed-trajectory motion in the trajectory-change condition, $t(35) = 2.44, p = .020$, for the upward split and $t(35) = 2.12, p = .041$, for the downward split.

DISCUSSION

Many researchers in several areas of cognitive science appeal to cohesion as possibly the central principle that defines persisting objecthood (e.g., Chiang & Wynn, 2000; Huntley-Fenner et al., 2002): If a feature cluster dissolves into multiple independent bounded contours, then it cannot be an object. This presents an interesting problem in the context of visual cognition, however, because observers clearly can perceive objects that split into two. (Indeed, the only other study that employed cohesion violations in the context of adult visual cognition showed that adults had no difficulty attentively tracking through cohesion violations per se; vanMarle & Scholl, 2003.) Nevertheless, the possibility remains that such conscious percepts are not effortlessly constructed, and in the present study we demonstrated for the first time, in the context of the object-file framework, that mechanisms of adult midlevel vision are adversely affected by cohesion violations.

The Cost of Cohesion Violations for Object Files

Using the object-reviewing paradigm (Kahneman et al., 1992), we explored what happens when an object splits into two. The central result of this experiment was that at least some object-specific information was maintained through such manipulations, but that the resulting OSPBs were significantly smaller than (indeed, less than half of) the OSPBs when objects underwent similar motions without cohesion violations. This result is broadly consistent with two interpretations, both of which have the same ultimate conclusion: that the

¹Across all motion types, for both the splitting and the trajectory-change conditions, observers were faster to respond when the target letter originally appeared as the top preview letter than when it appeared as the bottom preview letter. (Moreover, we have consistently observed this pattern in many other object-reviewing experiments, 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, such that no OSPBs could be driven solely by such displaywide spatial effects.

maintenance of the object files of adult visual cognition is constrained by a principle of cohesion.

The first, and strongest, interpretation of our results is that the object file associated with the initial object (before it split into two) did survive the splitting, but went along with only one of the two resulting objects (the choice perhaps being random). Thus, each resulting object would receive a full OSPB on half the trials, and no OSPB on the other half. Under this interpretation, cohesion violations serve as an especially strong constraint on dynamic object persistence: Object files can track objects through time and motion to some degree, but cannot themselves split into two—even though objects in the world may occasionally do this, and observers are easily able to consciously perceive such events. In short, according to this interpretation, object files cannot undergo even simple cohesion violations by splitting into two.

A second interpretation of our data also indicates an effect of cohesion on the maintenance of object files, but in a different way. Perhaps the object file associated with the initially presented object was able to survive the cohesion violation and be copied into both resulting objects: This, too, would be consistent with the fact that significant OSPBs were obtained for both upward-split and downward-split objects. Even under this interpretation, however, the data still indicate two important costs. First, the resulting OSPBs were less than half the size of the equivalent OSPB in the control condition (which involved similar motions but with no cohesion violation). Second, maintaining the object-specific information through the cohesion violation significantly attenuated the object file associated with the other, nonsplitting object in the display: Indeed, no significant OSPB was observed for this additional object, whereas the identical motion in the control condition yielded robust maintenance of object-specific information for both objects in the display. (Note that this cost was not due to the increased number of objects in the final display—that is, three rather than two—because reliable OSPBs are observed even with four-object displays; Kahneman et al., 1992, Experiment 5.) Thus, although object files can survive cohesion violations, their doing so places intense demands on the system, limiting the ability to maintain information about other objects.²

So what happens to an object file's information in the face of a cohesion violation? Our results rule out the two most extreme possibilities: Cohesion violations do not destroy object files completely, nor do object files persist through such violations without any adverse effects. We are left with the two interpretations just discussed. Examination of the actual distributions underlying our data revealed no hint of the bimodal response patterns (50% large OSPBs averaged with 50% zero OSPBs) suggested by the first interpretation—wherein the object file cannot itself split, and goes with only one of the two resulting objects. This suggests that our second interpretation may be more likely: The preview information may have traveled to both objects resulting from the split (as if the object file had been copied), but this operation more than halved the strength of the resulting OSPBs, and also severely attenuated the OSPB for the other, nonsplitting object. We prefer this second interpretation both because it is in some

ways more conservative (indicating a less extreme effect of cohesion violations on object files) and because there is no hint of bimodality in our data. However, bimodality is notoriously difficult to detect in such distributions, so we remain agnostic about which interpretation is correct. In any case, both interpretations fuel the same ultimate conclusion: that the object files of adult midlevel visual cognition are constrained by a principle of cohesion.

Conclusion

Though the object-file framework has been appealed to extensively in both the perception and the cognitive development literatures, very little is known about when object files are allocated and released, and how they are sustained, interrupted, and updated. Here we have explored these issues in the context of cohesion, which has been appealed to as perhaps the most central defining constraint on persisting objecthood. The nuanced results suggest that although observers can easily perceive an object split into two, such transformations actually place intense demands on the underlying midlevel visual representations.

Acknowledgments—For helpful conversation and comments on earlier drafts, we thank George Alvarez, Erik Cheries, Hoon Choi, Marvin Chun, Jen DiMase, Jonathan Flombaum, Steve Franconeri, Valerie Kuhlmeier, Alexandria Marino, Koleen McCrink-Gochal, Nic Noles, Kristy vanMarle, Yaoda Xu, and two anonymous reviewers. We also thank Alexandria Marino for assistance with data collection. S.R.M. was supported by National Institute of Mental Health Grant F32-MH66553-01. B.J.S. and K.W. were supported by National Science Foundation Grant BCS-0132444.

REFERENCES

- 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' tracking of objects and collections. *Cognition*, *77*, 169–195.
- Comtois, R. (2003). *VisionShell PPC [Software libraries]*. Cambridge, MA: Author.
- 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.
- Henderson, J.M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General*, *123*, 410–426.
- Henderson, J.M., & Anes, M.D. (1994). Roles of object-file review and type priming in visual identification within and across eye fixations. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 826–839.
- 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*, 223–250.
- 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.

²To our knowledge, this is the first study to show that OSPBs in the object-reviewing paradigm can persist through nonuniform trajectories, and in displays in which different objects follow different types of trajectories in the same trial. In other published object-reviewing experiments, objects followed either uniformly translating or uniformly rotating motions.

- 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.
- Mitroff, S.R., Scholl, B.J., & Wynn, K. (2003). The relationship between object files and conscious perception. *Journal of Vision*, 3, Abstract 338. Retrieved from <http://journalofvision.org/3/9/338/>
- Noles, N.S., Scholl, B.J., & Mitroff, S.R. (2003). The persistence of object-file representations. *Journal of Vision*, 3, Abstract 324. Retrieved from <http://journalofvision.org/3/9/324/>
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Scholl, B.J. (2001). Objects and attention: The state of the art. *Cognition*, 80, 1–46.
- Spelke, E.S. (1990). Principles of object perception. *Cognitive Science*, 14, 29–56.
- 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, England: Clarendon Press.
- vanMarle, K., & Scholl, B.J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, 14, 498–504.

(RECEIVED 4/7/03; REVISION ACCEPTED 7/11/03)