Psychonomic Bulletin & Review
2007, 14 (6), 1199-1204

Space and time, not surface features, guide object persistence

Stephen R. Mitroff
Duke University, Durham, North Carolina

and

George A. Alvarez
Massachusetts Institute of Technology, Cambridge, Massachusetts

Successful visual perception relies on the ability to keep track of distinct entities as the same persisting objects from one moment to the next. This is a computationally difficult process and its underlying nature remains unclear. Here we use the object file framework to explore whether surface feature information (e.g., color, shape) can be used to compute such object persistence. From six experiments we find that spatiotemporal information (location as a function of time) easily determines object files, but surface features do not. The results suggest an unexpectedly strong constraint on the visual system's ability to compute online object persistence.

An important component of successful visual perception is the ability to keep track of information as belonging to the same entities from one moment to the next. As an observer moves about the environment, and the environment about the observer, objects need to somehow be linked into persisting representations. While this process of *object persistence* is critical, it is not obvious how best it should be accomplished. On one hand, it would be too taxing on our visual system to remember every detail from one view of the world to the next, such that each view can be compared with the previous one. From what is known of attentional processing and memory, even a limited version of such a system is just not feasible. However, on the other hand, it seems maladaptive to encode nothing, simply letting the outside world act as its own memory. While this is a philosophical stance (e.g., O'Regan & Noe, 2001) and can be used as a successful strategy when implementing scene perception for robotic locomotion (Michels, Saxena, & Ng, 2005), clearly some amount information is stored in memory at any given moment. Yet, the questions then become, how much information, and of what sort?

A useful framework for addressing the nature of object persistence is the *object file* theory (e.g., Kahneman, Treisman, & Gibbs, 1992). According to this theory, information about various visual entities is bound together into *object files*—episodic visual representations that use spatiotemporal information to track entities over time and motion and that store (and update) information about the objects. That is, object files provide a means to bind visual features (e.g., shape, color) into units and to keep track of bound features as "the same unit" from one moment to the next. Note there are a few critical assumptions to this definition: (1) Object files are episodic in nature, temporarily representing information in the here and now; (2) object files are indexed, or tracked, via spatiotemporal information (location and time); and (3) object files store surface feature information.

Object files are most directly explored through the *object-reviewing paradigm*—a simple visual task that allows for an assessment of persisting objecthood over time (Kahneman et al., 1992). Observers view a preview display in which visual information is presented on various objects. For example, the letter S appears in a square on the left side of a display and the letter G appears in a square on the right. The letters disappear and the squares move, (e.g., the left square moves to the top and the right square moves to the bottom) and a target letter then appears in one of the squares and observers simply name this letter as quickly as possible. Typically, observers are slightly quicker to respond when the target letter appears on the same object in which it had originally appeared than when it appears on the other object. So observers would be quicker to say "S" if the letter S appeared on the final top square than on the bottom square. Critically, this response time benefit cannot be explained by general priming: Both letters were present in the original display, equidistant from the target letter location. The only thing underlying this response time benefit is *objectness*: Observers respond faster when the target appears on the same object.

Since the introduction of the object file theory, a number of aspects have been explored in depth. Several studies examined the nature of the information stored in object files (e.g., Gordon & Irwin, 1996, 2000; Henderson, 1994; Henderson & Anes, 1994). For example, it was shown that

S. R. Mitroff, mitroff@duke.edu

the stored information is abstract and flexible: If observers see a picture of a fish during the preview display, they will be quicker to read the word *fish* if it appears on the object the picture was previewed on than if it appears on a different object (Gordon & Irwin, 2000). Other studies have explored the rules that guide the operation of object files, examining how they are created, maintained, and destroyed (e.g., Mitroff, Scholl, & Wynn, 2004, 2005a, 2005b; Noles, Scholl, & Mitroff, 2005). For example, one study found that when an object split into two, the associated object file was maintained, although it was attenuated (Mitroff et al., 2004).

One issue that has not been explored more deeply though, is the assertion that object files are solely indexed through spatiotemporal information (Kahneman et al., 1992). All previous evidence for persisting object representations has come from studies in which an object is perceived to move from one place to another such that a spatiotemporal correspondence is apparent. When discussing object file representations, there must exist a necessary reliance upon visual memory such that specific representational information is needed to maintain objects as being the same. In the seminal paper on object files by Kahneman et al., the authors were clear that spatiotemporal information is the one and only means with which objects files can be maintained. Similar to the concept of *FINST* (e.g., Pylyshyn, 1989), the object file theory suggests that an object's surface feature and identity information may be stored independent of where the object happens to be and a spatiotemporal pointer allows for the tracking of the object. When an object moves or changes in some fashion, it is reassessed based upon its location with respect to time and if there is a sufficient spatiotemporal match, it is deemed to be the same object, and only then are the contents accessed.

Spatiotemporal correspondence is not an issue unique to the object file theory. It has played an important role, for example, in multiple object tracking (e.g., Scholl, Pylyshyn, & Franconeri, 1999), apparent motion (e.g., Dawson, 1991), and within the infant visual cognition literature (e.g., Wilcox & Chapa, 2004). As such, it is critical to understand how privileged spatiotemporal information really is. Similar to what has been asked in other literatures (e.g., Flombaum & Scholl, 2006; Navon, 1976; Scholl, 2001), here we ask whether nonspatiotemporal information can drive the computation of object persistence. Kahneman et al. (1992, Experiment 6) used a pseudoapparent motion paradigm to explore whether color alone could drive response time differences in the object-reviewing paradigm. Observers viewed a preview display of two letters, drawn in different colors, presented in white squares (one above and one below center). The colored letters disappeared and then two new squares appeared, one to the left and one to right of center. After a brief delay, the first two squares disappeared and a target letter appeared in one of the remaining squares. Observers were no quicker to respond if the target letter was the same color as when it was previewed than if it was a different color, suggesting that color alone cannot drive the computation of object files. However, this was not necessarily a fair test: By having the four squares all briefly present at the same time, it was clear that the objects at the end were not the same objects as those at the beginning. If object files keep track of entities as being the same objects, then there was no reason to expect any effect in this situation, regardless of whether or not color can underlie the computation of object files.

In three experiments and three subsequent control experiments, we reexamine whether surface feature information can underlie object persistence in the object-reviewing paradigm. In Experiment 1, we test the single surface feature of color. In Experiment 2, we examine color, shape, size, luminance, and topology (whether or not there is a hole in the object). In Experiment 3, we test color, shape, size, luminance, topology, and polarity. If we find that surface features can drive object files, it will expand the current understanding of the nature of object persistence. Alternatively, if we corroborate the original claims of Kahneman et al. (1992), we will provide evidence for a rather strict constraint upon the visual system's ability to keep track of objects over time.

METHOD

The three primary experiments use the same methods, except where noted below.

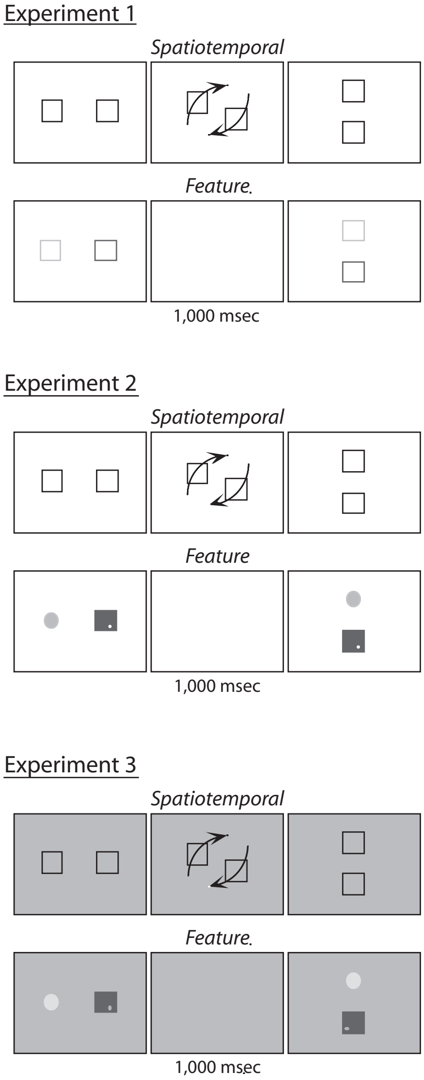
Participants

Observers were members of the Duke University community and were paid or received course credit for their participation. All had normal, or corrected-to-normal, vision and none reported being colorblind. If an observer's mean response time or accuracy rate was more than two standard deviations from the group mean (calculated for each experiment independently), their data were removed from all additional analyses. Ten observers participated in Experiment 1; 33 participated in Experiment 2, with the data from 4 outliers removed (2 for response time and 2 for accuracy); and 21 observers participated in Experiment 3, with data from 2 removed (both for accuracy).

Apparatus and Materials

Stimulus presentation and data collection were accomplished with a G4 Macintosh computer with a 19-in. CRT monitor, using custom software written with the VisionShell graphics libraries (Comtois, 2006). Observers sat, without head restraint, approximately 50 cm from the monitor in a dimly lit room (all visual angle calculations are estimated with this distance). Stimuli were presented against a solid white background (except for in Experiment 3, which had a medium gray background). Preview and target letters were drawn in a black monospaced font, subtending 1º. The letters were drawn without replacement from the set [K-M-P-S-T-V] for Experiments 1 and 2 and from [K-M-P-S-T-V-X-F-H-R] for Experiment 3.

Within each experiment there were two trial types—*spatiotemporal* and *feature*. In all three experiments, the spatiotemporal trials were as follows (see Figure 1): Each trial began with two identical squares (1.90 deg^2), presented as black outlines, 4.75º to the left and right of the center of the display (all distances are calculated to the center of the objects). When the squares moved, they traveled a curvilinear path, either clockwise or counterclockwise, such that they ended 4.75º above and below center. The feature trials differed by experiment, and the stimuli were as follows (see Figure 1): In Experiment 1, the preview display consisted of two squares (1.90 deg^2), 4.75º to the left and right of center. One of the two squares was colored blue and the other was red (color assignment varied across trials and was counterbalanced across the experiment). The squares reappeared in the target display 4.75º above and below center with one blue and the other red. In Experiment 2, the spacing and lo-



	Response Times (in Milliseconds)		
	Congruent Match	Incongruent Match	Object-Specific Preview Benefit
Experiment 1 – Spatiotemporal	523.00	548.24	25.24 msec* $t(9) = 3.75$ $p = .005$
Experiment 1 – Feature	563.81	564.78	0.96 msec $t(9) = 0.11$ $p = .914$
Experiment 2 – Spatiotemporal	555.58	569.04	13.47 msec* $t(26) = 2.42$ $p = .023$
Experiment 2 – Feature	597.57	602.35	4.78 msec $t(26) = 0.84$ $p = .406$
Experiment 3 – Spatiotemporal	567.10	581.68	14.58 msec* $t(18) = 2.19$ $p = .042$
Experiment 3 – Feature	593.76	595.12	1.36 msec $t(18) = 0.20$ $p = .846$

Figure 1. Display depictions, response times for congruent and incongruent match trials, and OSPBs (with their significance values) for the spatiotemporal and feature trials of Experiments 1–3. Feature trials contained two distinct objects. In Experiment 1, a blue square versus a red square; in Experiment 2, a blue, large, dark, square with a small hole versus a red, small, light, circle; and in Experiment 3, a blue, large, dark, square (with a small hole) that was relatively darker than the background, versus a red, small, light, circle that was relatively brighter than the background. The figure only depicts one motion direction for the spatiotemporal trials (objects could move clockwise or counterclockwise), only one object configuration for the feature trials (the "red" object could also appear on the right at the start and/or on the bottom at the end), and does not depict the preview and target displays wherein letters were drawn on the objects.

cations were the same as in Experiment 1, but the objects' surface features differed. One was blue, square, large (2.40 deg^2), relatively dark, and had a small hole in the bottom right quadrant. The other was red, circular, small (1.40º in diameter), relatively bright, and did not have a hole. Other than the hole in the square, both were solid objects. In Experiment 3, the objects were identical to those from Experiment 2, except that the hole was relatively larger and nonuniform (see Figure 1). Note also that in Experiment 3 the red object was brighter than the gray background and the blue object was darker, resulting in a polarity difference.

Procedure

Each trial began when the observer pressed the space bar, causing the preview objects to appear. After 500 msec, the preview letters appeared and remained visible for 500 msec. In the spatiotemporal trials, the objects then began their motion, either clockwise or counterclockwise, and moved for 1,000 msec. In the feature trials, the objects disappeared and then reappeared 1,000 msec later at the target locations. Although some observers reported experiencing apparent motion between the preview and target object locations on feature trials, this was not a necessary component to the experimental logic; the critical aspect of the feature trials was that the surface feature properties were the only information that could be used to link either target object to one of the preview objects.

After the linking phase, the target letter appeared either in the top or in the bottom object and observers responded as quickly as possible whether the target letter was the same as either of the preview letters. The target was neither of the preview letters 50% of the time (*no-match* trials) and was one of the preview letters the other 50% of the time (*match* trials). On match trials, 50% of the time the target letter matched the letter that had been previewed in that object (*congruent match* trials), and 50% of the time it matched the preview letter from the other object (*incongruent match* trials). Congruency was determined by continuity in the spatiotemporal trials and by surface features in the feature trials. Observers pressed the "1" key on the number pad to respond "match" and the "2" key for "no-match." This modified version of the object-reviewing paradigm (using a matching task rather than naming) has been successfully used numerous times (e.g., Mitroff et al., 2004, 2005a, 2005b; Noles et al., 2005) and has the advantage that it requires observers to attend to, and remember, the preview letters. Here we are testing the limits of object persistence, asking whether surface features can underlie object files, so using such a method is particularly appropriate.

All variables were counterbalanced such that the target letter was equally likely to appear in the top or bottom object, the objects were equally likely to rotate clockwise or counterclockwise in the spatiotemporal trials, and either object was equally likely to appear on the left or right and top or bottom in the feature trials. There were 288 test trials, half of which were spatiotemporal, and the other half feature trials. Trial type was randomly varied throughout the experiment, and the trial order was different for each observer. Prior to the experiment, observers received written and oral instructions and completed 20 practice trials. Observers were prompted to rest every 50 trials and were able to do so after any trial.

RESULTS

Observers were quite accurate on the match/no-match questions, both for the spatiotemporal and feature trials (*spatiotemporal*—Experiment 1, 95.97%; Experiment 2, 97.66%; Experiment 3, 96.38%; *feature*—Experiment 1, 96.81%; Experiment 2, 97.40%; Experiment 3, 96.71%), and there were no differences in accuracy by trial type (all p values $> .25$). Only trials with an accurate response within two standard deviations of the observer's mean response time were included in further analyses. Few trials were removed based upon these criteria (Experiment 1, 4.45% ($SD = 0.98\%$); Experiment 2, 4.13% ($SD = 0.91\%$); Experiment 3, 4.06% ($SD = 1.21\%$).

The critical measure of the object-reviewing paradigm is the *object-specific preview benefit* (OSPB), a response time benefit for congruent match over incongruent match trials. Separate OSPBs were calculated for spatiotemporal and feature trials for each observer, and the group results are presented in Figure 1. A significant OSPB was found for the spatiotemporal trials in each experiment, but not for the feature trials. An ANOVA performed over the difference data (OSPBs—incongruent match minus congruent match trials), with experiment as a between-subjects factor and trial type (spatiotemporal versus feature) as a within-subjects factor, resulted in a significant main effect of trial type [$F(1,54) = 6.08$, $p = .017$] and no interaction between trial type and experiment ($F < 1$).

ELIMINATING POTENTIAL CONCERNS

Before discussing the implications of these results, we briefly offer supporting evidence from three additional experiments.

Experiment 4: Trial Structure Concern

An intermixed trial design in the object-reviewing paradigm can bias observers' ability to process objecthood (Mitroff, Arita, & Fleck, 2006), so it is possible that the design here may have weakened the feature trials. To test for this, 38 observers participated in an additional experiment solely composed of 256 feature trials. The procedure and stimuli were identical to Experiment 2, except that the objects differed by three features: square, small (1.07 deg^2), and red vs. circular, large (2.90º in diameter), and blue. Overall accuracy was 95.37% and there was no significant feature-driven OSPB [congruent, $M =$ 499.88 msec; incongruent, $M = 504.01$ msec; OSPB = 4.13 msec, $t(37) = 1.12$, $p = .27$].

Experiment 5: Disappearance Concern

It is possible that the disappearance of the objects in the feature trials, per se, could have weakened the feature trials. To address this, 15 observers participated in an experiment wherein objects disappeared and reappeared, but moved along a spatiotemporally continuous path. The procedure was similar to Experiment 1 with the following changes: Two identical white outlined discs (2.52º in diameter) appeared vertically aligned (5.05º between their centers and drawn against a black background) 12.63º to the left of center. After a letter briefly appeared in each disc (1,000 msec), they moved to the right at a constant rate of 6.31 deg/sec for a total of 4 sec. Once they stopped (12.63º to the right of center), a single target letter appeared in one. At 650 msec into the motion, the discs disappeared, reappearing 2,700 msec later, further along the motion path, as if they had been moving the whole time. Each observer completed 160 trials, and overall accuracy was 91.20%. Critically, an OSPB [13.78 msec, $t(14) =$ 2.22, $p = .044$] was observed, even though the objects were physically gone for nearly 3 sec.

Experiment 6: Linking Duration Concern

It is possible that the 1,000 msec ISI in the feature trials of Experiments 1–3 may have attenuated any existing effects. The spatiotemporal trials always had information physically present, but the feature trials had a 1,000-msec gap. To address this, we had 21 observers participate in an additional experiment. (Two observers were removed based upon the criteria described above, 1 for response time and 1 for accuracy.) This experiment was identical to Experiment 3 in all aspects save one: The ISI was reduced from 1,000 msec to 0 msec. Observers were accurate on 96.69% of the trials and showed the same pattern of data as that shown in Experiments 1–3. There was a significant OSPB for the spatiotemporal trials [congruent $M =$ 594.83 msec, incongruent $M = 613.82$, OSPB = 18.99, $t(18) = 2.27$, $p = .036$], but not for the feature trials [congruent $M = 624.63$ msec, incongruent $M = 633.02$, OSPB = 8.39, $t(18) = 0.829$, $p = .418$].

DISCUSSION

Through the above six experiments, it was found that surface feature information alone did not allow for the maintenance of object file representations. Even when there were six featural differences between the two objects in Experiments 3 and 6, the observers were no quicker to respond when the target letter reappeared in the same object than when it reappeared in the other object. Importantly, though, on an intermixed set of trials with spatiotemporal continuity, the observers were quicker to respond on congruent than incongruent trials, demonstrating that the failure on the feature trials is not a general null result but rather a specific failure of surface feature continuity to yield an advantage for congruent trials. Furthermore, given that the feature trials showed significantly smaller object-specific preview benefits than the spatiotemporal trials, the important message of these experiments is not necessarily that features can *never* underlie the computation of object files, but rather that, if they can, they sure don't do so very well.

The present findings support the original predictions and claims of Kahneman et al. (1992) and do so multiple times and with a variety of tested features. While spatiotemporal information is known to play a special role in visual perception, it was still quite surprising that surface features wielded no significant influence here given that no other information relevant to object persistence was available. Considering the nature of the experimental design, this is particularly informative about where along the object file process the failure to use feature information arises. Consider that during the preview display, the visual system does not "know" whether there is going to be subsequent spatiotemporal information to follow. When there happened to be spatiotemporal information, object specific preview benefits were observed, so the objects must be processed, at least to some extent, at this stage. Furthermore, we know that objects can temporarily disappear and reappear and still produce OSPBs, as long as they are accompanied by clear spatiotemporal information. Thus, the only major change here is that the spatiotemporal information is not available to be used by the object file system's "reviewing" operation to determine whether a currently viewed object is the same as a previously viewed one. It appears that once the objects reappear in the target display, the object file system cannot use the surface feature information (even though it is presumably available) to address which object is which.

Clearly, feature information, in the absence of spatiotemporal continuity, plays a key role in some aspects of visual processing (e.g., recognizing a friend you are picking up at the airport). However, an open question for future exploration is whether here we gave feature information a fair chance to guide online object persistence. While our spatiotemporal trials involved no ambiguity, the feature trials, in a sense, provided the visual system with a conundrum: Beyond there being no spatiotemporal information available to help support object file processing, there was also the surprising situation of objects magically disappearing and reappearing. Typically, motionless objects do not suddenly disappear and then reappear in new locations. Thus a possibility remains that under different circumstances, without conflicting spatiotemporal information, surface features could help guide object persistence.

The present experiments suggest that the computation of object persistence appears to rely so heavily upon spatiotemporal information that it will not (or at least is unlikely to) use otherwise available surface feature information, particularly when there is conflicting spatiotemporal information. This reveals a striking limitation, given various theories that visual perception uses whatever shortcuts, or heuristics, it can to simplify processing, as well as the theory that perception evolves out of a buildup of the statistical nature of our environment (e.g., Purves & Lotto, 2003). Instead, it appears that the object file system has "tunnel vision" and turns a blind eye to surface feature information, focusing on spatiotemporal information when computing persistence.

AUTHOR NOTE

G.A.A. was supported by NIMH Grant F32-EY016982-01. For helpful conversation and comments, we thank Mathias Fleck, Steven Franconeri, Andrew Hollingworth, Steven Luck, Brian Scholl, and an anonymous reviewer. We also thank Jason Arita and Tina Liang for assistance with data collection. Nicolaus Noles and Brian Scholl graciously provided a subset of the data for our Experiment 5. Correspondence concerning this article should be addressed to S. R. Mitroff, Center for Cognitive Neuroscience, Duke University, Box 90999, Durham, NC 27708-0999 (e-mail: mitroff@duke.edu).

REFERENCES

Comtois, R. (2006). VisionShell PPC. [Software libraries]. Cambridge, MA: Author.

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

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

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, & Cognition*, **26**, 136-150.

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

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

Michels, J., Saxena, A., & Ng, A. Y. (2005, August). High-speed obstacle avoidance using monocular vision and reinforcement learning. *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany.

Mitroff, S. R., Arita, J. T., & Fleck, M. S. (2007, November). *Stay-*

ing in bounds: Contextual constraints on object file coherence. Paper presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

MITROFF, S. R., SCHOLL, B. J., & WYNN, K. (2004). Divide and conquer: How object files adapt when a persisting object splits into two. *Psychological Science*, **15**, 420-425.

MITROFF, S. R., SCHOLL, B. J., & WYNN, K. (2005a, November). One plus one equals one: The effects of merging on object files. Poster presented at the annual meeting of the Psychonomic Society, Toronto, ON, Canada.

MITROFF, S. R., SCHOLL, B. J., & WYNN, K. (2005b). The relationship between object files and conscious perception. *Cognition*, **96**, 67-92.

NAVON, D. (1976). Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 130-138.

NOLES, N. S., SCHOLL, B. J., & MITROFF, S. R. (2005). The persistence of object-file representations. *Perception & Psychophysics*, **67**, 324-334.

O'REGAN, J. K., & NOE, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral & Brain Sciences*, **24**, 939-1031.

PURVES, D., & LOTTO, B. (2003). *Why we see what we do: An empirical theory of vision*. Sunderland, MA: Sinauer Associates.

PYLYSHYN, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, **32**, 65-97.

SCHOLL, B. J. (2001). Spatiotemporal priority and object identity. *Cahiers de Psychologie Cognitive*, **20**, 359-371.

SCHOLL, B. J., PYLYSHYN, Z. W., & FRANCONERI, S. (1999, May). When are spatiotemporal and featural properties encoded as a result of attentional allocation? Paper presented at the meeting of the Association for Research in Vision and Ophthalmology, Ft. Lauderdale, FL.

WILCOX, T., & CHAPA, C. (2004). Priming infants to attend to color and pattern information in an individuation task. *Cognition*, **90**, 265-302.

(Manuscript received April 21, 2006;
revision accepted for publication January 24, 2007.)