The onset of receding motion captures attention: Comment on Franconeri and Simons (2003)

RICHARD A. ABRAMS and SHAWN E. CHRIST
Washington University, St. Louis, Missouri

Franconeri & Simons (2003) reported that some but not all types of motion capture attention in a visual search task, ostensibly because some types of motion are behaviorally significant. In the present article, we argue that a more parsimonious explanation of their results is that the *onset* of motion, but not motion per se, captures attention. This conclusion helps to resolve inconsistencies between the Franconeri and Simons findings and earlier reports from other investigators and is consistent with results that we have recently reported (Abrams & Christ, 2003). The Franconeri and Simons interpretation rests largely on their failure to find attentional capture by one type of motion—simulated receding motion. We report here the results of two experiments that demonstrate that the onset of receding motion does indeed capture attention when the motion is produced using stereo depth cues. As we have argued elsewhere, the capture of attention in displays containing motion appears to be dependent on the onset of the motion, rather than on the mere presence of specific types of motion as suggested by Franconeri and Simons. One possible explanation is that the onset of motion captures attention because it serves as a strong cue to animacy—and the detection of nearby animals can have important consequences for survival.

In a recent paper, Franconeri and Simons (2003) reported attentional capture by several different types of motion. In particular, unidirectional translational motion, oscillatory translation, and simulated looming motion all captured attention. However, simulated receding motion did not.[1] Franconeri and Simons suggested that their pattern of results was consistent with the hypothesis that behaviorally urgent stimuli can capture attention. According to this hypothesis, receding motion, as if an object were moving away from the observer, is less urgent and hence less likely to capture attention.

There is much that is appealing about the behavioral urgency hypothesis offered by Franconeri and Simons, and it is similar to the explanation that we have offered for our recent results (Abrams & Christ, 2003). Building on previous work by others (Hillstrom & Yantis, 1994; Yantis & Egeth, 1999), we studied the attentional consequences associated with motion and various movement transitions. In particular, using a visual search paradigm similar to that employed by Franconeri and Simons, we compared attentional capture for items that recently began to move, recently stopped moving, were continuously moving, and never moved. All of our results were consistent with a single conclusion: Motion per se does not capture attention, but the onset of motion does. A potential explanation that we offered is that the onset of motion is likely to be indicative of animacy, thus signaling

the presence of a possible predator or prey. The ability to detect such animals may indeed be behaviorally urgent, and one's survival may depend on it.

We believe that capture by motion onset, not by specific types of motion, is also consistent with the results reported by Franconeri and Simons, and that a motion onset explanation is more parsimonious than the behavioral urgency hypothesis they offered. One reason for this argument is that the motion onset explanation is a testable and potentially falsifiable hypothesis, whereas the behavioral urgency hypothesis does not permit one to determine a priori what sorts of stimuli are likely to be behaviorally urgent. Capture by motion onset in the Franconeri and Simons experiments seems likely because they presented the search array in each of their experiments for only 150 msec after the onset of the various types of motion. If attention had been captured by the onset of motion, then target detection would have been facilitated in each of their conditions involving moving stimuli, as Franconeri and Simons observed (with one exception to be addressed below).

Note also that this reinterpretation of the Franconeri and Simons data helps to explain the discrepancy between their findings and those of earlier researchers such as Hillstrom and Yantis (1994); Folk, Remington, and Wright (1994); and Yantis and Egeth (1999). These earlier researchers failed to find capture by some of the same sorts of motion studied by Franconeri and Simons. The reason for this failure, we believe, is that the methods used in the earlier experiments did not create the conditions necessary for motion *onset* to capture attention. In particular, in the Hillstrom and Yantis (1994) and

Yantis and Egeth (1999) experiments, the search stimuli were moving from the moment at which they appeared in the display. Similarly, in the Folk et al. experiments, the cues and targets that were moving were in motion from the time at which they first appeared. As we have noted previously (Abrams & Christ, 2003), sensitivity to motion onset appears to be suppressed around the time of appearance of a new object (Torriente, Valdes-Sosa, Ramirez, & Bobes, 1999). In the Franconeri and Simons experiments, however, a 1,000-msec preview period was provided before the motion began, thus temporally separating the appearance of the placeholders from the onset of motion and permitting the capture of attention by the motion onset.

There is, however, one set of results from the Franconeri and Simons paper that is inconsistent with the explanation that we have offered here. According to our explanation, the onset of motion in any direction might be expected to capture attention, because motion onset would be expected to be indicative of an animate object regardless of the direction of movement. But Franconeri and Simons failed to find capture by receding motion. Their explanation was that receding motion is less behaviorally urgent than the other types of motion studied. It is important to note, however, that Franconeri and Simons simulated receding motion by using relative size as the only cue to depth. Thus, the placeholder in their display shrank as it "receded." Whereas relative size can provide cues to depth, an alternative interpretation of their display is that the placeholder was maintaining its distance from the observer but shrinking in size. As such, the Franconeri and Simons experiment does not provide a strong test of the possibility that receding motion (or the onset of receding motion) can capture attention—only that perhaps shrinking does not capture attention. We provide such a test by using displays in which binocular disparity served as a depth cue.

EXPERIMENT 1

In this experiment, subjects searched for letters in displays in which one of several placeholders moved away from the observer, using binocular disparity as a depth cue. The target was no more likely to appear in the moving placeholder than in a static one. If the onset of receding motion captures attention, target identification should be quickest on those trials for which the target happened to coincide with the receding element.

Method

Subjects. Eleven undergraduates participated in the experiment. All were experimentally naive and received course credit in return for their efforts.

Apparatus and Procedure. The subjects were seated in front of a cathode ray tube in a dark room. The subjects viewed the display through a stereoscopic viewer mounted on the table at which they sat. The viewer constrained each eye's view to one half of the display and permitted the use of binocular disparity to present stimuli at a range of depths. The sequence of events on each trial is shown in Figure 1. At the beginning of each trial, the display contained a central fixation cross and three or six placeholders. The placeholders were figure eights that were 2.89° high, 1.44° wide, and located at equal intervals along an imaginary circle with a 5° radius centered at fixation. When three placeholders were present, they appeared at the 0° (directly above fixation), 120°, and 240° positions along the circle. When six placeholders were present, the positions were 0°, 60°, 120°, 180°, 240°, and 300°. All elements on the display were presented at zero disparity, with the exception of one randomly selected placeholder that was displayed so as to appear to be in front of the others by 6.5°. The retinal size of the "floating" placeholder was adjusted so as to be consistent with the disparity cues. After a preview of 3,000 msec, the floating placeholder moved in depth back to the plane that contained the remaining placeholders. The motion was completed in 150 msec, at which point segments from each of the placeholders were removed to reveal letters. One of the placeholders became the letter *S* or *H,* representing the target stimulus. All remaining placeholders were replaced by distractor letters (either all *E*s or all *U*s). Subjects were instructed to respond to the target's identity as quickly as possible by pressing one of two keys (*z* or "/") on the keyboard. The presence of a preview period, the 150-msec duration of receding motion, and the presentation of the search array at the end of the motion were identical to corresponding features in the Franconeri and Simons experiment.

The search array remained visible until the subject responded or 3,000 msec had elapsed. If the subject responded incorrectly, a brief tone followed by the message *Wrong Response* was presented. A tone and a relevant message (i.e., *Too Early* or *Too Slow*) was presented if a subject responded less than 300 msec after array onset or failed to respond within 3,000 msec, respectively. The intertrial interval was 1,000 msec.

Design. Following 24 practice trials, the subjects served in 336 experimental trials. One half of the trials had an array size of three elements, and the other trials had an array size of six. The target was the receding element on one third of the trials in the array-size three condition and on one sixth of the trials in the array-size six condition. Thus, the receding element was not more likely to be the
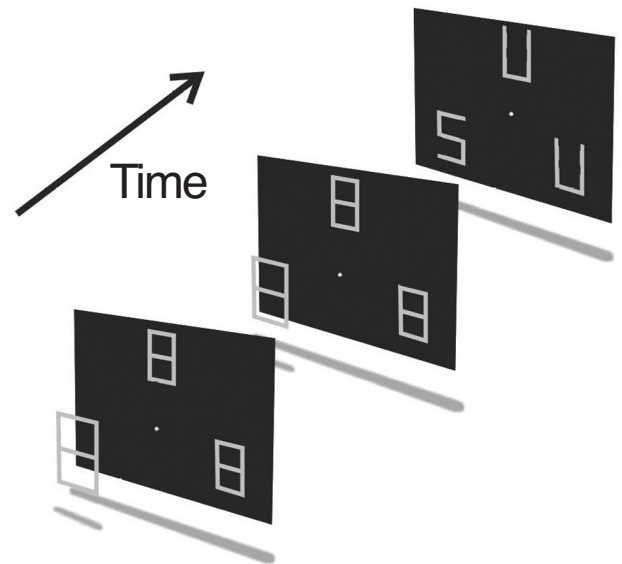

Figure 1. Perspective rendering of the sequence of events during Experiment 1. One randomly selected placeholder was seen as floating in front of the others by means of binocular disparity and relative size. That placeholder receded until it was in the same plane as the other elements, at which time the search array was presented. Additional details are provided in the text.

target than were any of the other elements. The target was equally likely to be an S or H in each condition, and the distractor letters were equally likely to be Es or Us for each type of target. The target-to-response key mapping was counterbalanced across subjects. Trial types were randomly mixed. At intervals of 48 trials, the subjects were given the opportunity to take a break.

Results

Mean reaction times are shown Figure 2. As can be seen, reaction times increased considerably with increases in array size for the static condition (the search slope was 17.7 msec per item) but only slightly for the receding condition (slope = 4.3 msec per item). These slopes are consistent with the hypothesis that the receding item captured attention, so that the target was identified quickly when it was the receding item, regardless of the size of the search array. These observations were corroborated by the statistics: We obtained main effects of array size [$F(1,10) = 82.3, p < .001$] and target type [$F(1,10) = 50.0, p < .001$] and an interaction between array size and target type [$F(1,10) = 8.2, p < .05$]. Errors were low ($M = 7.4\%$) and did not depend on target type or array size [$Fs(1,10) < 2.0, p > .05$].

Discussion

The onset of receding motion in the present experiment captured attention, as evidenced by the quick reaction times and shallow search slopes for trials on which the target happened to be the receding element. This result contrasts with the failure to find capture by receding motion in the experiment reported by Franconeri and Simons. In their experiment, however, depth was cued by relative size only, leaving open the possible interpretation that their "receding" element actually appeared to be shrinking and not receding. In contrast, we used both relative size and binocular disparity to cue depth and obtained robust evidence of attentional capture.
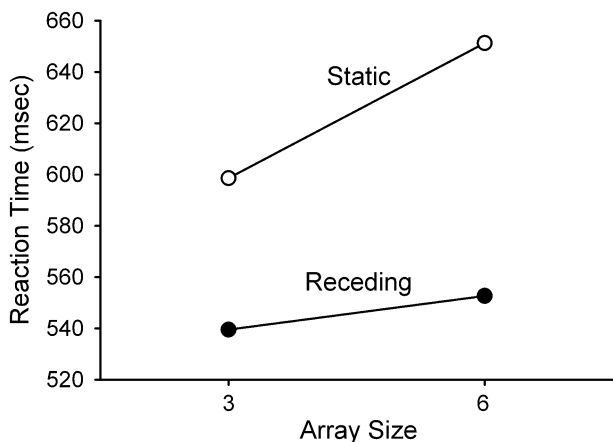

Figure 2. Mean reaction time to identify the target as a function of the size of the search array and the prior motion of the target element, from Experiment 1. Receding objects captured attention, as evidenced by the fast latencies and shallow search slope when the receding item happened to be the target.

Before proceeding, we should consider one potential alternative explanation for our results. The receding element in our experiment did differ from the other elements in one way other than the fact that it receded: At the beginning of each trial, the element that was to recede was also a singleton in depth, floating in front of the plane of the other objects. If attention had been drawn to that element prior to the onset of motion (presumably because of its status as a depth singleton), then it might be possible that our results were not due to the motion onset. There are, however, several arguments to discount this possibility: First, there is no empirical basis for the possibility that a depth singleton will capture attention. Indeed, even highly salient singletons do not capture attention in the absence of an appropriate attentional set (e.g., Yantis & Egeth, 1999). Second, if the depth singleton had attracted attention prior to motion, the timing of the events on the trial would lead to the expectation of *poorer* target detection performance for the receding element due to inhibition of return (IOR). In fact, the timing of trial events and the equivalent target probabilities across search elements presented ideal conditions for the generation of IOR. Theeuwes and Godijn (2002) recently showed such an effect for an irrelevant color singleton. Thus, we conclude that attention was captured by the onset of the receding motion.

EXPERIMENT 2

In this experiment, we sought to compare directly the attentional effects of the onset of receding motion with those of shrinking. To do so, we had subjects search for a target in a display similar to that used in Experiment 1. On each trial here, one element receded using stereo depth cues, as in Experiment 1. Also on each trial, one element shrunk, as in the experiment by Franconeri and Simons. The target occurred equally often in each of these two types of elements. If shrinking provides only a weak depth cue, as we have suggested, target identification should be faster for the receding stimulus than for the shrinking one.

Method

Subjects. Twelve undergraduates participated in the experiment. All were experimentally naive and received course credit in return for their efforts. None had participated in Experiment 1.

Procedure. The method was similar to that used in Experiment 1, with differences noted here. Each trial began with a 3,000-msec preview of a placeholder display that contained three figure-eight placeholders. The fixation point and two of the placeholders were presented at zero disparity, but the other placeholder was floating 6.5º in front of the plane of the monitor. After the preview period, the floating placeholder receded using stereo depth cues and relative size, as in Experiment 1. At the same time, a second placeholder shrank, with the change in retinal size matching that of the receding placeholder (from 3.25º high and 1.66º wide to 2.89º high and 1.44º wide). The movement and shrinking lasted 150 msec, after which segments were removed from all three placeholders to reveal the search array. The target was equally likely to appear in the static, shrinking, or receding element. The subjects pressed one key if

an S was present and another if an H was present. Each subject completed 24 practice trials followed by 336 test trials, with breaks offered every 48 trials.

Results

Mean reaction times for static, shrinking, and receding trials were 631.6 msec, 593.5 msec, and 577.2 msec, respectively, with the condition having a significant effect on the reaction times [$F(2,22) = 23.1$, $p < .001$]. Both shrinking and receding targets were identified more quickly than static ones [for static vs. shrinking, $t(11) = 4.5$, $p < .001$; for static vs. receding, $t(11) = 5.9$, $p < .001$]. Importantly, receding targets led to faster responses than shrinking ones [$t(11) = 2.4$, $p < .05$]. Errors occurred on 5.2% of trials overall and did not depend on condition [$F(2,22) < 1$].

Discussion

This experiment yielded two important results: First, we replicated the results from Experiment 1 showing that a substantial advantage is enjoyed by targets undergoing motion onsets, even in the case of motion that is away from the observer. The benefit for motion defined by stereo depth cues was 54 msec. Second, we showed that the benefit of the stereo depth receding motion exceeded that of the shrinking stimulus, supporting our earlier assumption that the attentional impact of a depth change may depend upon the manner in which the depth change is rendered. The finding that latencies in the shrinking condition were faster than those in the static condition is similar to the main effect of target location reported by Franconeri and Simons. This effect alone cannot be used to argue for attentional capture because, in the case of Franconeri and Simons's experiment, it occurred in the absence of a difference in search slopes (and in the present experiment we did not measure search slopes). Importantly, the present results do confirm the assumption upon which Experiment 1 rested: Stereo depth–defined receding motion produces a more profound attentional effect than recession simulated through shrinking.

GENERAL DISCUSSION

In two experiments, we presented evidence suggesting that the onset of receding motion captures attention. In Experiment 1, we found fast latencies and shallow search slopes for stimuli that receded via binocular disparity depth cues—strong evidence of attentional capture. In Experiment 2, we obtained faster latencies for stereo depth–defined stimuli (like those from Experiment 1) relative to stimuli that simulated receding motion by shrinking (similar to those used by Franconeri and Simons). These results support our assumption that stereo depth cues are more likely to lead to the perception of motion.

Before concluding, it is worth considering why the simulated recession through shrinking studied in Experiment 2 and by Franconeri and Simons did not result in attentional capture, despite the fact that the shrinking was accompanied by considerable retinal motion. If indeed the onset of motion in general captures attention, as we have suggested, why was attention not captured by the retinal motion of the shrinking objects' contours? The reason, we believe, is that attention is attracted not by low-level changes in luminance-defined contours, such as that which occurs when an object moves and also when an object shrinks, but instead by a change in the perceived location of the object. Thus, for example, changes that are interpreted as shape or size changes but not as movement of the object would not be expected to capture attention. Support for this distinction has been provided by Rauschenberger (2003). He argued that featural changes may capture attention only when they are interpreted by the perceptual system as indicating the emergence of a new object. Additional support for these ideas comes from several studies that have shown that the perceptual system can compute the movement of an apparent object center that is distinct from motion of the object's contours. For example, cells in a brain area sensitive to movement (the medial superior temporal area) respond selectively not merely to local components of the stimulus motion, but also to a higher order, derived representation of the orientation of the depicted object (Sugihara, Murakami, Shenoy, Andersen, & Komatsu, 2002). Also, Masson and Stone (2002) showed that pursuit eye movements are not driven by the retinal motion of the contours of the pursued object, but instead by a computed representation of the movement of the object itself. Given the close links between attention and eye movements (e.g., Kustov & Robinson, 1996; Shepherd, Findlay, & Hockey, 1986), it is perhaps not surprising that attention may also be guided by the perceived object motion as distinct from retinal motion at the object's contours, in agreement with our present results.

Franconeri and Simons's failure to detect capture by receding motion was the only evidence they offered that was inconsistent with our hypothesis that motion *onset* captures attention. It is now possible to interpret their results, our earlier results (Abrams & Christ, 2003), and the results of researchers such as Folk et al. (1994), Hillstrom and Yantis (1994), and Yantis and Egeth (1999) with a single, simple explanation: The onset of motion of preexisting objects captures attention, whereas motion per se does not. Motion onset is important, we believe, because it is indicative of the presence of an animate object, an animal, possibly predator or prey, whose efficient detection might be important for survival.

REFERENCES

ABRAMS, R. A., & CHRIST, S. E. (2003). Motion onset captures attention. *Psychological Science*, **14**, 427-432.

FOLK, C. L., REMINGTON, R. W., & WRIGHT, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 317-329.

FRANCONERI, S. L., & SIMONS, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, **65**, 999-1010.

HILLSTROM, A. P., & YANTIS, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics*, **55**, 399-411.

KUSTOV, A. A., & ROBINSON, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, **384**, 74-77.

MASSON, G. S., & STONE, L. S. (2002). From following edges to pursuing objects. *Journal of Neurophysiology*, **88**, 2869-2873.

RAUSCHENBERGER, R. (2003). When something old becomes something new: Spatiotemporal object continuity and attentional capture. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 600-615.

SHEPHERD, M., FINDLAY, J. M., & HOCKEY, R. J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology*, **38A**, 475-491.

SUGIHARA, H., MURAKAMI, I., SHENOY, K. V., ANDERSEN, R. A., & KOMATSU, H. (2002). Response of MSTd neurons to simulated 3D orientation of rotating planes. *Journal of Neurophysiology*, **87**, 273-285.

THEEUWES, J., & GODIJN, R. (2002). Irrelevant singletons capture attention: Evidence from inhibition of return. *Perception & Psychophysics*, **64**, 764-770.

TORRIENTE, I., VALDES-SOSA, M., RAMIREZ, D., & BOBES, M. A. (1999). Visual evoked potentials related to motion-onset are modulated by attention. *Vision Research*, **39**, 4122-4139.

YANTIS, S., & EGETH, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 661-676.

NOTE

1. Franconeri and Simons also reported results from two additional conditions that revealed attentional capture by the appearance of a new object (their "onset" and "disocclusion" conditions). Those results are not relevant to the present discussion.