

# Motion tracking modulates capacity allocation of visual working memory

JOO-HYUN SONG and YUHONG V. JIANG  
*Harvard University, Cambridge, Massachusetts*

Previous studies on visual working memory (VWM) have primarily investigated memory for an array presented for a single moment. Here, we examined VWM for two arrays separated by a 1,100-msec interval. We focused on the allocation of VWM capacity to the two arrays as a function of dynamic events inserted between them. During the interval, irrelevant dots moved to form three types of motion: (1) coherent (apparent) motion that connected Arrays 1 and 2, (2) jumpy motion, or (3) coherent motion with two disconnected segments. Results showed that VWM for Array 2's locations was better than for Array 1's, especially when the arrays were connected by coherent motion. We suggest that coherent motion between two temporally disparate arrays connects the arrays into a single visual event. VWM is then biased toward remembering the more recent state of the event.

We live in a constantly changing visual environment: People shift locations as they move around, objects move in and out of view as the wind sweeps across trees, and even a static scene transforms as our eyes land on one spot after another. At least two mechanisms have been proposed that may help us maintain a sense of temporal continuity in a dynamic visual world: tracking of moving objects and visual working memory (VWM). Motion tracking allows one to follow an object as it moves from one location to another (Pylyshyn & Storm, 1988). Even when tracking is not required for performing a task, observers tend to track items as they move (Kahneman, Treisman, & Gibbs, 1992). Motion tracking provides an effective mechanism for temporal updating as long as there is good spatiotemporal continuity between one visual object and another (Kahneman et al., 1992).

When spatiotemporal continuity is lacking, VWM supplies another mechanism to maintain temporal continuity. VWM refers to the momentary retention of visual information after its disappearance. It has a limited capacity—only about four visual objects can be retained in VWM simultaneously (Luck & Vogel, 1997; Phillips, 1974). This capacity dictates that only the most important information can be maintained across spatiotemporal disruptions. Together, motion tracking and VWM help us connect the history of visual objects to their present states. Other mechanisms also exist to maintain temporal continuity, including the retention of a high-fidelity, high-resolution iconic memory shortly after stimulus disappearance (Sperling, 1960), and visual long-term memory, which is particularly suited

to maintaining the gist of a scene (Hollingworth & Henderson, 2003; Standing, 1973). This study focuses on the interaction between motion tracking and VWM.

Specifically, we ask: To what degree does spatiotemporal continuity between two visual arrays affect VWM for these arrays? This question has rarely been addressed by previous studies of VWM, which have focused primarily on memory for displays presented for a single moment. Although some studies have examined VWM for sequentially presented arrays (Broadbent & Broadbent, 1981; Brockmole, Wang, & Irwin, 2002; Jiang & Kumar, 2004; Phillips & Christie, 1977), they often used arrays that lacked spatiotemporal continuity. This has left us with limited knowledge about how VWM interacts with motion tracking.

A few studies on motion tracking have examined the kind of memory that results from active tracking. In multiple-object tracking, when items are put into apparent motion and then briefly occluded, people typically have good memory for objects' spatial locations and directions of motion, but poor memory for nonspatial properties (Scholl, Pylyshyn, & Franconeri, 1999). Even when participants succeeded at tracking target objects' locations, they failed to tie nonspatial properties such as color and shape to the correct locations (Pylyshyn, 2004). This suggests that holding object identities in VWM is not the same as tracking moving objects' locations.

Tracking nonspatial properties of objects depends on maintaining coherence in object motion. When the direction of motion—clockwise or counterclockwise—is perceptually ambiguous, participants have difficulty binding object colors to their locations, even when they are instructed to treat the objects as moving clockwise (Saiki, 2003). Object identities are retained to a certain degree on moving displays, at least when the number of objects is small and motion is unambiguous. For example, Kahneman et al. (1992) briefly previewed two letters in four boxes before rotating the ensemble of boxes by 90°. Following the rotation, another letter was presented in one of the boxes. Identification

---

This study was supported by NSF Grant 0345525 to Y.V.J. The Korea Foundation for Advanced Studies provided support for J.-H.S. We thank James Brockmole, Livia King, Won Mok Shim, Rachel Sussman, Tim Vickery, and Amanda Willis for comments. Correspondence concerning this article should be addressed to J.-H. Song, Smith-Kettlewell Eye Research Institute, 2318 Fillmore St., San Francisco, CA 94115 (e-mail: jhsong@ski.org).

of the letter was facilitated when it matched the letter previewed in the same box, but was delayed when it matched the other previewed letters (see also Henderson, 1994; Henderson & Anes, 1994). Thus, spatiotemporal continuity is important for maintaining stable representations of objects, with “object files” keeping various properties of an object bound together (Rensink, 2000).

The studies reviewed above provided some hints about the interaction between motion tracking and VWM. However, the emphasis of these studies was primarily motion tracking, rather than VWM. Our study was designed to study the property of VWM, with or without motion tracking between two target arrays. Because coherent motion connects two separate visual arrays into two instances of a single event, it changes the number of events that one must remember in VWM. In turn, the overall VWM capacity may increase because a bigger chunk is formed from two separate displays. This is not the only possible outcome, however. With coherent motion tracking, VWM capacity can decrease because the intermediate motion may produce interference. Alternatively, motion tracking can change the distribution of VWM capacity to the leading and trailing arrays without affecting the overall capacity.

## EXPERIMENT 1 Coherent (Apparent) Motion Versus Jumpy Motion

Two dot arrays, each containing three colored dots, were presented with a short interstimulus interval (ISI). The participants were told to remember the six locations from the two arrays, and determine which one of two testing locations matched their memory. During the ISI, several irrelevant arrays containing white dots were presented to form two motion conditions. The irrelevant dots either moved continuously, to produce a sense of coherent (apparent) motion between the target arrays, or they moved at random locations, to produce a sense of jumpy motion. How would coherent motion affect VWM for the target arrays?

### Method

**Participants.** In this study, 13 students from Harvard University (18–30 years old) participated for payment or course credit. They all had normal color vision and normal or corrected-to-normal visual acuity.

**Equipment.** The participants were tested individually in a room with normal interior lighting. They viewed a computer screen from an unrestrained distance of about 57 cm, at which distance 1 cm corresponded to 1° of visual angle.

**Stimuli and Procedure.** Figure 1 illustrates the trial sequence used in this experiment. After participants fixated a white cross shape (0.2°) for 200 msec, two visual arrays, each lasting 100 msec, were presented with an ISI of 1,100 msec. A blank retention interval of 1,000 msec followed the second target array, after which the test display was presented until the participants made a response. The two visual arrays each contained three colored dots (0.8° × 0.8°); the dots in Array 1 were red and those in Array 2 were green. The test display included two white dots marked by the digits 1 and 2, which were randomly assigned. The participants were told to determine which test dot matched one of the locations of the memory arrays. On half of the trials, one test location (the “target”) matched one of Array 1’s locations; on the remaining trials, it matched one of Array 2’s locations. The other test items (the

“filler”) were about 2.5° away from another memory location selected from the same memory array as the matched location. The participants were not cued to the relevant array. Accuracy feedback was displayed immediately after each response. One sec later, the next trial began.

The locations of the dots within the two arrays were randomly selected from a 3 × 3 invisible grid matrix that subtended 23.4° × 15.6°. Three cells were randomly chosen from the matrix, and within each cell, two locations (at least 2.5° apart) were randomly selected: one to contain Array 1’s position and the other to contain Array 2’s position. The locations were chosen in this manner to leave the path of apparent motion unambiguous. Five frames (100-msec presentation + 100-msec blank interval), each containing three white dots, were presented 100 msec after Array 1’s termination. In the coherent motion condition, the distance between corresponding dots on Arrays 1 and 2 was divided into five equal units. Each of the five irrelevant frames then occupied each successive unit, resulting in a clear perception of coherent (apparent) motion along the straight line connecting the two memory arrays. All distractor items were at least 2.5° away from the memory items and test locations. The minimal speed of motion was 2°/sec. In the jumpy motion condition, the three dots in the third distractor frame were presented at the same locations as those in the third frame of the coherent motion condition (i.e., halfway between Arrays 1 and 2). Dots in the remaining four distractor frames were presented at randomly selected locations within three different randomly selected cells. Thus, both motion conditions contained distractor frames, allowing us to control for low-level visual masking effects.

The participants were told to give equal priority to the two target arrays and to completely ignore the white dots inserted between them. Each participant completed 240 experimental trials, with conditions randomly intermixed (2 motion distraction conditions × 2 arrays × 60 instances).

### Results

We measured accuracy separately for the coherent motion and jumpy motion conditions. Array 1’s memory was calculated on the basis of trials in which the test location matched one of Array 1’s locations, and Array 2’s memory was calculated on the basis of the other trials. Figure 2 shows the percent correct for each array. Accuracy in all conditions was significantly above chance (all  $ps < .05$ ). An ANOVA on array and motion coherence showed a significant main effect of array [ $F(1,12) = 47.57, p < .001$ ]. Performance was better for Array 2’s locations than for Array 1’s locations, showing a recency effect. The main effect of motion coherence was not significant ( $F < 1$ ), suggesting that motion coherence did not affect overall VWM capacity. However, the interaction between array and motion coherence was significant [ $F(1,12) = 6.20, p < .05$ ]. The recency effect was significant in both the jumpy motion condition [ $t(12) = 3.63, p < .005$ ] and the coherent motion condition [ $t(12) = 7.76, p < .001$ ]. However, the significant interaction effect showed that the latter was stronger. Coherent motion appeared to have enhanced Array 2’s memory at the cost of Array 1’s memory, although the two simple effects—Array 2 enhancement and Array 1 decrement—failed to reach significance ( $ps > .09$ ).

### Discussion

How did the presentation of irrelevant moving stimuli affect VWM for the locations of two target arrays spanning the motion? Experiment 1 showed that the overall memory accuracy was relatively unaffected by whether the inserted motion was coherent or jumpy. However, the

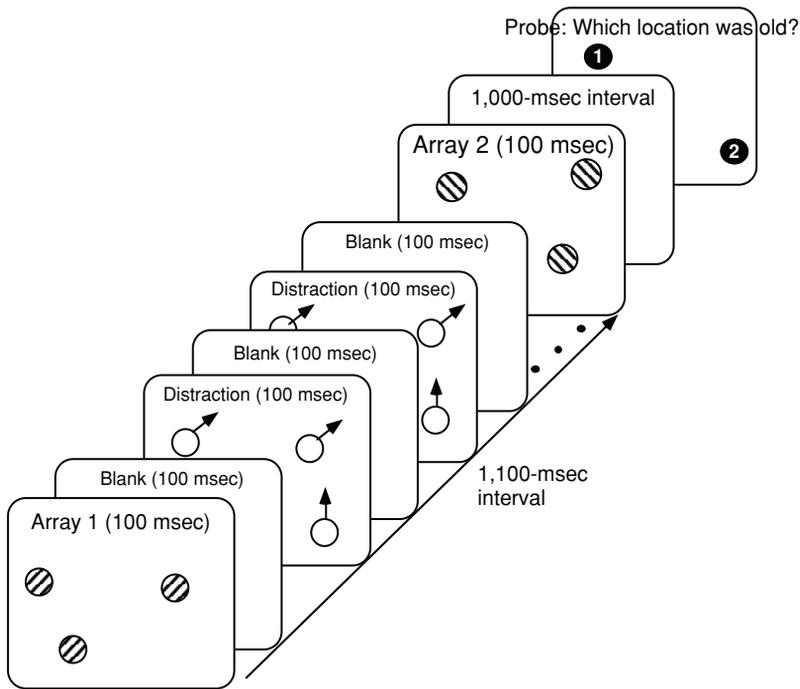


Figure 1. A schematic demo of the trial sequence in Experiment 1.

type of motion changed how this capacity was distributed. In both cases, VWM was better for the trailing array than for the leading array, resulting in a recency effect. The recency effect became stronger with coherent motion. This finding is consistent with the idea that VWM places greater emphasis on more recent visual information, especially when the leading and trailing arrays belong to two instances of a single event.

However, there are alternative interpretations. In particular, recency might have increased simply because the locations of Array 2 were predictable in the coherent but not the jumpy motion conditions. Although predictability may be a characteristic of motion tracking, it captures a more general effect. That is, predictability of Array 2's locations can be provided with nonmotion cues. Furthermore, although we tried to equate the distance between the distractor frames and the target arrays, the first distractor frame was closer to Array 1, and the last distractor frame was closer to Array 2 in the coherent motion condition. Such increased proximity can impair memory for Arrays 1 and 2 as a result of visual masking. Thus, the predictability of Array 2's locations and the increased proximity between the two target arrays and distractors may singly or jointly account for findings from Experiment 1. To narrow down the role of motion tracking in VWM, we needed to control for these confounding factors. Experiment 2 was designed to achieve this goal.

## EXPERIMENT 2

### Connected Versus Disconnected Motion

In this experiment, we compared VWM for two sequential arrays separated by two types of motion stimuli: con-

nected motion and disconnected motion. The connected motion condition was the same as the coherent motion condition of Experiment 1, in which five distractor frames moved continuously to bridge Array 1 with Array 2 through apparent motion. The disconnected motion condition also contained five distractor frames plus two target arrays, but the motion was broken down into two segments. Array 1 and the first distractor frame were placed close together to form one short motion segment, allowing us to control for proximity. Distractor Frames 2–5 and Array 2 were also arranged close together to form a short coherent motion segment. This made the locations of Array 2 predictable.

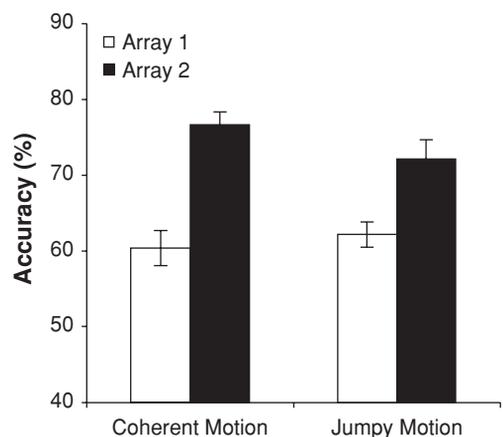


Figure 2. Results from Experiment 1. Error bars represent the between-subjects standard error for each condition.

The critical feature of this condition, however, was that Distractor Frames 1 and 2 were relatively far apart, resulting in the percept of two disconnected motion segments. With proximity and predictability controlled for, any difference between the connected and disconnected motion conditions could be attributed only to motion continuity.

## Method

**Participants.** Thirteen Harvard participants were recruited for payment or course credit. They were 18–30 years old and had normal color vision and normal or corrected-to-normal visual acuity.

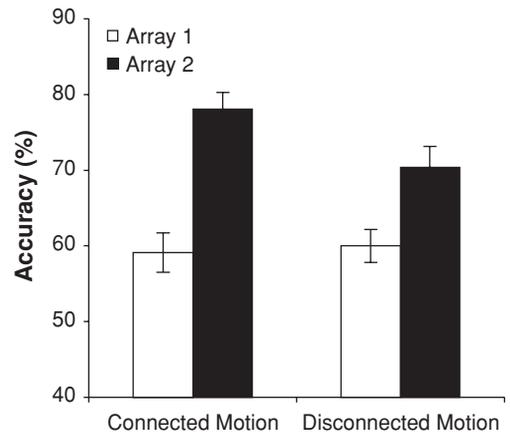
**Procedure.** This experiment was the same as Experiment 1 except that the jumpy motion condition was replaced by a disconnected motion condition. The connected motion condition was the same as Experiment 1's coherent motion condition, in which five distractor frames were inserted between two target arrays to form coherent (apparent) motion. In the disconnected motion condition, Distractor Frame 1 was placed adjacent to Array 1 to control for proximity between the two. The two frames formed the first motion segment that moved at 2°/sec. In addition, Distractor Frames 2–5 formed coherent (apparent) motion that led to Array 2 along a straight line with a minimal speed of 2°/sec. These five frames (including Array 2) formed the second motion segment, allowing us to control for predictability of Array 2. The critical manipulation was that a relatively large spatial distance (about 8°, compared with 0.5° in the coherent motion condition) separated Distractor Frames 1 and 2. The resulting percept was the presence of two disconnected motion segments.

Note that the jump between Distractor Frames 1 and 2 in the disconnected motion condition resulted in increased distance between Arrays 1 and 2. We therefore ran a control experiment in which the distractor frames were eliminated, leaving only Arrays 1 and 2. The two arrays were either separated by a short distance, as in the connected motion condition, or a long distance, as in the disconnected motion condition. Accuracy for Array 1 was 55% in the short-distance condition and 59% in the long-distance condition, and accuracy for Array 2 was 68% in the short-distance condition and 70% in the long-distance condition. Statistical analysis revealed no effect of distance on overall accuracy or the recency effect ( $ps > .20$ ). Thus, any difference between connected motion and disconnected motion observed in Experiment 2 cannot be attributed to increased distance between the two target arrays.

## Results

Figure 3 shows percent correct for each array. Here, too, accuracy in all conditions was significantly above chance ( $ps < .005$ ). An ANOVA on array and motion connectivity showed a significant main effect of array [ $F(1,12) = 40.79, p < .001$ ], revealing a recency effect. The main effect of motion connectivity was not significant [ $F(1,12) = 1.16, p > .30$ ], but the interaction between array and motion was [ $F(1,12) = 6.28, p < .028$ ]. The recency effect was significant in both the disconnected motion condition [ $t(12) = 3.21, p < .01$ ] and the connected motion condition [ $t(12) = 7.75, p < .001$ ], but was stronger in the latter. Thus, even when extraneous factors such as proximity and predictability were controlled for, the recency effect was still greater with connected motion. Reminiscent of the results of Experiment 1, coherent motion resulted in a small reduction in accuracy for Array 1 and an enhanced accuracy for Array 2, although neither simple effect was significant on its own ( $ps > .09$ ).

When we pooled data across the two experiments, coherent motion again did not significantly affect the overall VWM performance [ $F(1,24) = 1.52, p > .20$ ]. It did



**Figure 3. Results from Experiment 2. Error bars represent the between-subjects standard error for each condition.**

enhance the recency effect [ $F(1,24) = 10.41, p < .005$ ], suggesting that the distribution of VWM across the two target arrays changed. However, of the two simple effects, only Array 2 memory was significantly enhanced by coherent motion [ $t(25) = 2.5, p < .02$ ]. Array 1 memory was slightly but not significantly impaired by coherent motion [ $t(25) < 1$ ]. Thus, although coherent motion tipped the balance of VWM toward the more recent array, the modulation was subtle.

## GENERAL DISCUSSION

This study examined the interaction between motion tracking and VWM. We inserted different types of motion between two successive visual arrays. The inserted motion stimuli were irrelevant to the VWM task and were perceptually distinctive from the target arrays. Our results showed that, under some conditions, such stimuli could affect VWM for the relevant memory arrays.

First, the inserted motion did not significantly change the overall VWM capacity. Averaged across the leading and trailing arrays, VWM accuracy was relatively constant whether the inserted motion was coherent, jumpy, or disconnected. In fact, a direct comparison between coherent motion and a no-motion condition in a preliminary study revealed no difference between the two in overall VWM performance. This result suggests that connecting two visual arrays into a single event does not reduce the competition between the two. It also suggests that the inserted stimuli did not directly compete with the target arrays for VWM capacity.

Second, in all motion conditions, VWM for the trailing array was better than that for the leading array. The recency effect is perhaps a joint result of memory decay and preference for more recent visual information. Because Array 1 was presented earlier than Array 2 was, it would have decayed more when memory was probed. However, when a single array was presented and the delay interval was systematically manipulated, VWM decayed only slightly over a 1- to 2-sec interval (Jiang, 2004; Luck & Vogel,

1997; Phillips, 1974). Thus, the dramatic difference between the two arrays cannot be completely accounted for by decay. Instead, the more recent target array appears to receive higher priority. This preference for more recent information is also seen in other paradigms, such as change detection (Beck & Levin, 2003) and substitution masking, in which a trailing stimulus “erases” perceptual awareness of a leading one (Di Lollo, Enns, & Rensink, 2000).

The most important finding of our study is that the recency effect in VWM was increased by irrelevant coherent motion inserted between the two arrays. Because the recency effect was smaller for jumpy motion sequences than for coherent motion sequences, the increase in the recency effect in VWM was not attributable to masking effects produced by the inserted stimuli. It was also not entirely due to other factors, such as the predictability of Array 2, because the recency effect was smaller when the apparent motion segment leading to Array 2 was disconnected from that originating from Array 1. The interaction between motion tracking and VWM suggests that VWM is part of the system that updates visual information across temporal lags.

To find out whether the effect of coherent motion was restricted to VWM for spatial locations, we tested 13 participants in an additional experiment in which color, rather than spatial location, must be held in VWM. We found that VWM for colors was also affected by inserted coherent motion: Recency increased with coherent motion. However, this effect was smaller in color VWM than in location VWM. These results are consistent with the idea that although tracking carries some identity information, it primarily updates the spatial locations of visual objects.

What mechanism resulted in the modulation of VWM by coherent motion? We believe that the concept of object files (Henderson, 1994; Kahneman et al., 1992; Rensink, 2000) can help explain the present results. As in visual perception, VWM relies on the representation of objects (Luck & Vogel, 1997). Each time a few objects must be retained in VWM, an object file is opened for each object. As each object moves, or its surface properties change, information about the object's location or nonspatial properties is updated in the object file. When two successive arrays are presented for people to hold in VWM, new object files are opened for each item in each array. However, coherent motion between the two arrays makes it more likely that the same object files would be used for corresponding items on the two arrays. This can lead to increased recency effect.

As in earlier perceptual experiments (Kahneman et al., 1992), the present study did not require participants to actively track the motion of distractor objects. The fact that motion still affected the allocation of VWM suggests that tracking was to some extent obligatory. Whether the object-file mechanisms would still be used when participants actively track a subset of objects is an empirical question. Under such conditions, motion tracking is likely to compete directly with VWM for attentional capacity. This may make it more difficult to detect subtle modulations on recency.

To summarize, by inserting irrelevant motion between two visual arrays and studying how VWM for the two arrays is affected by motion, we have shown in our study that the capacity allocation of VWM can be modulated by motion tracking. We postulate that the effect of coherent motion on VWM originates from VWM's function in updating visual information across temporal lags.

## REFERENCES

- BECK, M. R., & LEVIN, D. T. (2003). The role of representational volatility in recognizing pre- and postchange objects. *Perception & Psychophysics*, **65**, 458-468.
- BROADBENT, D. E., & BROADBENT, M. H. P. (1981). Recency effects in visual memory. *Quarterly Journal of Experimental Psychology*, **33A**, 1-15.
- BROCKMOLE, J. R., WANG, R. F., & IRWIN, D. E. (2002). Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 315-334.
- DI LOLLO, V., ENNS, J. T., & RENSINK, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, **129**, 481-507.
- HENDERSON, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General*, **123**, 410-426.
- HENDERSON, J. M., & ANES, M. D. (1994). Roles of object-file review and type priming in visual identification within and across eye fixations. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 826-839.
- HOLLINGWORTH, A., & HENDERSON, J. M. (2003). Testing a conceptual locus for the inconsistent object charge detection advantage in real-world scenes. *Memory & Cognition*, **31**, 930-940.
- JIANG, Y. (2004). Time window from visual images to visual short-term memory: Consolidation or integration? *Experimental Psychology*, **51**, 45-51.
- JIANG, Y., & KUMAR, A. (2004). Visual short-term memory for two sequential arrays: One integrated representation or two separate representations? *Psychonomic Bulletin & Review*, **11**, 495-500.
- KAHNEMAN, D., TREISMAN, A., & GIBBS, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24**, 175-219.
- LUCK, S. J., & VOGEL, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, **390**, 279-281.
- PHILLIPS, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, **16**, 283-290.
- PHILLIPS, W. A., & CHRISTIE, D. F. M. (1977). Interference with visualization. *Quarterly Journal of Experimental Psychology*, **29**, 637-650.
- PYLYSHYN, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, **11**, 801-822.
- PYLYSHYN, Z. W., & STORM, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, **3**, 179-197.
- RENSINK, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, **7**, 17-42.
- SAIKI, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision*, **3**, 6-21.
- SCHOLL, B. J., PYLYSHYN, Z. W., & FRANCONERI, S. L. (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.
- SPERLING, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, **74**(11, Whole No. 498).
- STANDING, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, **25**, 207-222.