

# Mobile computation: Spatiotemporal integration of the properties of objects in motion

**Patrick Cavanagh**

Department of Psychology, Harvard University,  
Cambridge, MA, USA, &  
Laboratoire Psychologie de la Perception,  
Université Paris Descartes,  
Paris, France



**Alex O. Holcombe**

School of Psychology, University of Sydney,  
Sydney, Australia



**Weilun Chou**

Department of Psychology, National Taiwan University,  
Taipei, Taiwan



We demonstrate that, as an object moves, color and motion signals from successive, widely spaced locations are integrated, but letter and digit shapes are not. The features that integrate as an object moves match those that integrate when the eyes move but the object is stationary (spatiotopic integration). We suggest that this integration is mediated by large receptive fields gated by attention and that it occurs for surface features (motion and color) that can be summed without precise alignment but not shape features (letters or digits) that require such alignment. Rapidly alternating pairs of colors and motions were presented at several locations around a circle centered at fixation. The same two stimuli alternated at each location with the phase of the alternation reversing from one location to the next. When observers attended to only one location, the stimuli alternated in both retinal coordinates and in the attended stream: feature identification was poor. When the observer's attention shifted around the circle in synchrony with the alternation, the stimuli still alternated at each location in retinal coordinates, but now attention always selected the same color and motion, with the stimulus appearing as a single unchanging object stepping across the locations. The maximum presentation rate at which the color and motion could be reported was twice that for stationary attention, suggesting (as control experiments confirmed) object-based integration of these features. In contrast, the identification of a letter or digit alternating with a mask showed no advantage for moving attention despite the fact that moving attention accessed (within the limits of precision for attentional selection) only the target and never the mask. The masking apparently leaves partial information that cannot be integrated across locations, and we speculate that for spatially defined patterns like letters, integration across large shifts in location may be limited by problems in aligning successive samples. Our results also suggest that as attention moves, the selection of any given location (dwell time) can be as short as 50 ms, far shorter than the typical dwell time for stationary attention. Moving attention can therefore sample a brief instant of a rapidly changing stream if it passes quickly through, giving access to events that are otherwise not seen.

Keywords: visual attention, spatiotemporal integration, color, motion

Citation: Cavanagh, P., Holcombe, A. O., & Chou, W. (2008). Mobile computation: Spatiotemporal integration of the properties of objects in motion. *Journal of Vision*, 8(12):1, 1–23, <http://journalofvision.org/8/12/1/>, doi:10.1167/8.12.1.

## Introduction

What properties can be computed for an object while it is moving across the retina? Motion detectors certainly integrate over a portion of an object's trajectory but this spatiotemporal integration is hard-wired (Burr, 1981) and may be a special case. For other object properties, like basic features and their relations, 3D shape, and transparency, there may be no hard-wired units that can combine these properties across locations (although see Nishida, 2004). Therefore, when stimuli move rapidly over the retina, analyses at each location may not have time to reach completion before the stimulus moves on.

This will degrade any processing that depends on a completed analysis in retinotopic coordinates. Nevertheless, numerous studies show that observers can process visual features of moving objects (e.g., Bex, Dakin, & Simmers, 2003; Blaser, Pylyshyn, & Holcombe, 2000; Brown, 1972; Chung & Bedell, 2003; Levi, 1996; Morgan & Watt, 1982) suggesting that further processing may occur that is not in retinal coordinates but involves object-specific accumulation of information (Kahneman, Treisman, & Gibbs, 1992; Nishida, 2004; Nishida, Watanabe, Kuriki, & Tokimoto, 2007; Oğmen, Otto, & Herzog, 2006; Shimozaki, Eckstein, & Thomas, 1999; Watanabe & Nishida, 2007). Neurons in higher visual areas show the basic requirements for this non-retinotopic step because a single cell with a

large receptive field can continue to respond to an object independently of its location, as long as the object remains within its receptive field.

Large receptive fields have been documented in high-level visual areas in human fMRI recordings (e.g., Kanwisher, 2003; Kim & Tong, 2005; Tong & Kim, 2005) and monkey single cell recordings (e.g., Ito, Tamura, Fujita, & Tanaka, 1995). Evidence that these areas still show some degree of retinotopy (receptive fields are large but still cover much less than the entire visual field) has been reported for LOC and FFA in human fMRI (e.g., Avidan, Hasson, Malach, & Behrmann, 2005; Larsson & Heeger, 2006), in human face aftereffects (Afraz & Cavanagh, 2008) and in IT in monkeys (DiCarlo & Maunsell, 2003; Rolls, Aggelopoulos, & Zheng, 2003). The size of these large receptive fields places an upper limit on the spatial extent over object-specific analyses may occur.

Like the “object files” concept (Kahneman et al., 1992), mobile computation requires a cataloguing of properties acquired locally, but in addition, it may also offer continued visual analysis based on partial information picked up on the fly. The simplest approach to test the perception of a moving object would be to present a single object moving smoothly around the fixation point as shown in Figure 1a. However, determining the extent of local analysis for the moving object is problematic so we have devised a different stimulus that neatly controls the extent of local processing. Multiple stimuli are presented at fixed locations around a circle (Figure 1c) and at each location two patterns alternate rapidly. The same two patterns are alternating at every location, but the alternation is out of phase so that odd locations show pattern 1 while even locations show pattern 2, and this arrangement then reverses in the next display array. If attention can track across the locations in synchrony with the alternation, a

single object is seen in apparent motion, moving around the circle, stepping through a ring of unmoving, flickering stimuli. To help guide attention at higher speeds, we use a ring that encircles the intended target in each array of the sequence (Figure 1c).

The rapid alternation at each location restricts the analysis that can be completed locally, allowing us to evaluate whether non-local accumulation is contributing to performance. The extent of local processing is measured in conditions where observers attend to and report the stimulus at a fixed location in the array. When the alternation rate is fast enough, the stimulus is no longer identifiable at any one location, even with extended scrutiny. If performance improves when tracking through the same display with a moving window of attention, this advantage must reflect continued analysis of object properties across locations in the moving frame. Our technique can evaluate the analysis of basic features like color, motion, and orientation, and their conjunctions as well as the construction of object descriptions from these features. It can also test for non-local analysis of high-level properties like transparency, 3D structure from junction cues and motion aperture effects. In this paper, we demonstrate the moving attention procedure and report evidence that color and motion features can be integrated across locations but letter and digit shapes cannot. We also demonstrate that interruption masking is specific to the attended trajectory (it is not retinotopic) whereas superposition masking is retinotopic.

Our technique focuses on object-based integration where information about the moving target may be accumulated in object-centered coordinates that are neither retinotopic nor spatiotopic. Previous studies have shown clear evidence for object-based spatiotemporal integration. For example, using moving objects visible

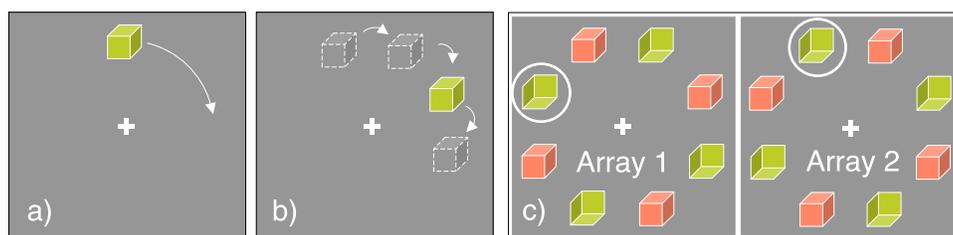


Figure 1. Apparent motion procedure. (a) A cube moves continuously around the fixation point at a speed too fast for local analysis of its structure to be completed at each location. If the perception of the cube’s 3D shape remains intact, it requires an accumulation of intermediate results in an object-based, position-independent representation. (b) In our procedure we do not use continuous motion but apparent motion where the target makes discrete steps around the fixation point. (c) To restrict the time available for processing at a single location, we present the target at every location alternating rapidly with a second pattern: the green cube in this example alternates with a red cube of a different orientation. We increase the alternation rate until neither cube can be identified (e.g., name the color of the cube whose front faces down to the left) when attending to any one location, ensuring that local processing cannot complete the analysis of the target. Now we add a guide, here a circle, that steps from location to location in phase with the local alternation so that the target (the green cube) is always within the guide which continues around the locations in the array (only two arrays in the sequence shown here). Observers are asked to track the guide with attention, recreating the apparent motion shown in (b) but now with verifiably insufficient information for identification at each location. Despite having only incomplete information at each location, the moving attention window in Experiment 1 clearly reveals the target, demonstrating accumulation of object-specific information across locations.

only through narrow slits (anorthoscopic presentation), Nishida (2004) showed integration of shape information along a motion path where no single frame of the motion sequence contained identifiable information. Information concerning the cortical location of this integration was found by Yin, Shimojo, Moore, and Engel (2002) who reported that perceptual integration of the moving anorthoscopic patterns was most evident in object areas within LOC. Other studies demonstrate how object information, presented at one location can be perceived at another location further along a motion trajectory that was perceived to pass through that location (Ogmen et al., 2006; Scharnowski, Hermens, Kammer, Ogmen, & Herzog, 2007). More recently, using apparent motion of targets that alternate between red and green at successive locations, Nishida et al. (2007) and Watanabe and Nishida (2007) have demonstrated integration of color information along a motion path. We are adding to this object-based integration literature by providing a general technique that can examine integration of many perceptual properties and that can explicitly control the amount of information available at each location on the path.

How does object-based integration relate to spatiotopic integration (De Graef & Verfaillie, 2002; Hayhoe, Lachter, & Feldman, 1991; Irwin, 1996; Melcher & Morrone, 2003; Rayner, 1998)? When the eyes move and objects in the scene do not, the objects remain fixed in spatiotopic coordinates and studies have suggested spatiotopic analyses of at least motion information in various sites in the visual cortices (d'Avossa et al., 2007; Melcher & Morrone, 2003). Spatiotopic analyses may be a special case of object-based processing that occurs when the eyes move but objects do not or it may depend on specific, eye movement contingent processes so it is not yet clear whether results for the two cases will be related.

Attention plays a critical role in our technique, linking adjacent locations into a single motion path. This makes our technique a poor choice to examine the importance of attention itself for the integration. Without attentive tracking, there is only local alternation with no net direction; a different technique would be required to adequately examine whether the integration changes without attention as Melcher, Crespi, Bruno, and Morrone (2004) report for motion.

Finally, our results also have implications for the dwell time of attention (Duncan, Ward, & Shapiro, 1994), the minimum time between the arrival of two stimuli that allows attention-limited processing of the first to be completed without interruption by the second. Our results will demonstrate that the local dwell time for attention at a given location (the time over which attention is sampling from that location) is not fixed at the rather long value found for stationary attention (150 to 300 ms)—it can be extremely brief (50 ms or less). Yet, while it is in motion, attention still has a long, overall dwell period in the same range (150 to 300 ms) but the dwell “window” is spread across space and it is only open briefly at any one

location. This allows moving attention to sample a short duration stimulus that is part of a rapid stream by quickly moving through the location where the stream is presented. Moving attention appears to capture predominantly that one stimulus and sidestep much of the interference that the next stimulus at that location would otherwise create.

## Experiment 1: Color and motion

In our first experiment, we examined the conjoint report of both color and motion features. In this case, the motion is that of the object's surface texture which moves independently of the object. The surface motion is inward or outward while at the same time the attended object either remains in place (Figures 2a and 2b), or, when tracked in the guide ring, moves clockwise or counterclockwise (Figures 2c and 2d). In the display, rectangular patches at each of 10 locations contain either red or green dots that are moving either toward or away from fixation within the rectangular patch (indicated by arrows). The patch itself does not move, but the dots forming the texture within it do. The motions and colors reverse on each alternation so that the patches with green dots moving outward (Figure 2a) become patches with red dots moving inward (Figure 2b) and vice versa. In the stationary attention condition, the observer fixates the central dot and attends to one location (for example, the bottom patch) throughout the trial, reporting the direction of either the red or green dots at that location. In the attentive tracking condition (Figures 2c and 2d) the observer fixates the central point and using the moving white ring as a guide, tracks the encircled patch with attention. This ring moves in synchrony with the alternation so that it always encloses a single pairing, here, red moving inward. While following the guide ring, the observer sees the attended patch and its surrounding ring as one object in apparent motion, sweeping around the circle, and reports its pairing of color and motion: in Figure 2, red dots paired with inward motion.

## Methods

### Participants

The participants included authors PC and AH and 3 naïve observers. Observers ranged in age from 18 to 58 years and had normal or corrected-to-normal visual acuity and normal color vision.

### Apparatus

The experiments were run on Apple Macintosh G4 computers with custom software written in C using the Vision Shell Graphics Libraries (Comtois, 2003). Displays

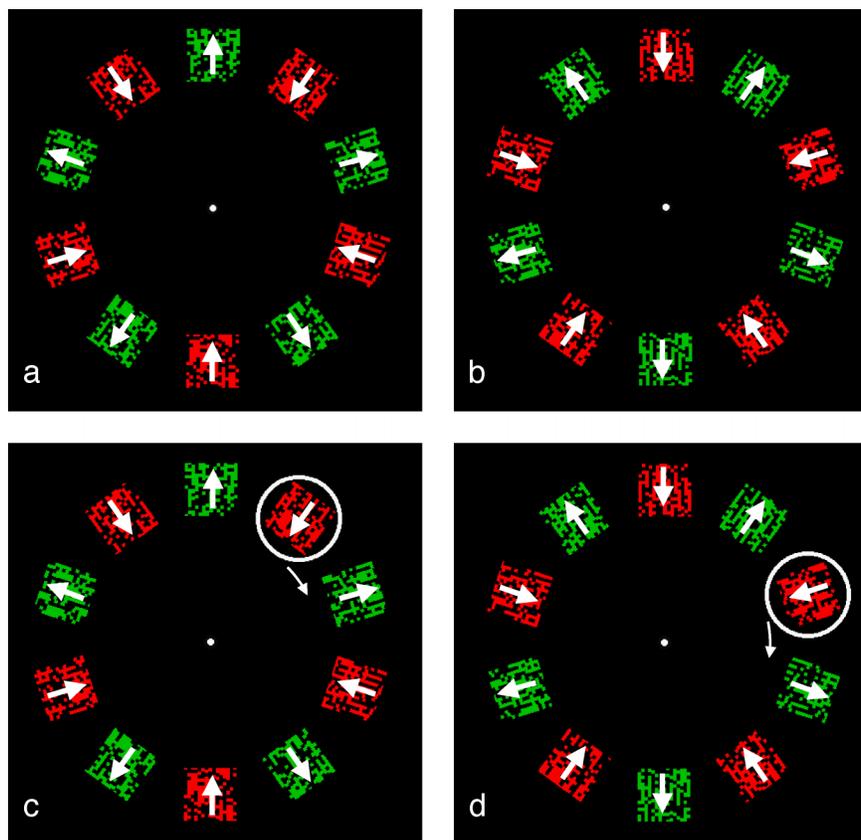


Figure 2. (a, b) Stationary attention baseline. Observer attends to a single location while fixating the central dot. Arrays a and b alternate and the rate is varied to determine the rate at which it is no longer possible to report which color is paired with which direction of dot motion (about 2 or 3 Hz). Click on the links here to see demonstration movies at [2-Hz](#) alternation rate where the pairing of colors and motions is easy to report, or at [3 Hz](#), or [4 Hz](#), where this becomes more difficult (actual display rate may deviate from intended rate). This stationary attention condition evaluates the time requirements of local analyses. (c, d) Moving attention condition. Identical stimulus except there is now a ring that the observer follows with attention as it moves one position per alternation, continuing around the display, potentially enabling the accumulation of local computations. It always arrives at a location when the same value (inward moving red dots in the trial shown here) is present. The direction of dot motion and their color are now much easier to report, with the threshold rate for 75% accuracy more than twice as high as in the stationary attention condition. Click on the links here to see demonstration movies at alternation rates of [2 Hz](#), [3 Hz](#), or [4 Hz](#), where reporting the pairing of color and motion within the moving guide should be easier than for the stationary case.

were presented on CRT monitors with 85-Hz refresh rate and resolution of  $1280 \times 1040$  pixels. Responses were collected on USB keyboards. A chin rest was used to maintain viewing distance of 57 cm.

### Stimuli

Ten rectangular patches of dots were spaced uniformly around a circular array. The array had a diameter of 13 degrees of visual angle and each test patch was a square of colored, moving dots,  $2.4 \times 2.4$  degrees visual angle. Individual dots were square, 0.13 degrees in size. The dots (either red or green, with a density of 50%) were randomly positioned on a black background. Each test patch was rotated to be aligned with the radius from the fixation point through the patch's center (Figure 2). At each location, the dot motion was either inward toward fixation or outward away from fixation at 12 degree per

second. The two colors, red and green, were produced with solely the red and green phosphor, respectively. The red phosphor was set at maximum luminance ( $30 \text{ cd/m}^2$ ) and the green reduced to 60% maximum ( $40 \text{ cd/m}^2$ ) which gave approximate equiluminance set using a minimum flicker match at 10 Hz. There were two types of test array sequences. In a red-inward sequence, red dots always moved inward and green dots outward. In the red-outward sequence, red dots moved outward and green, inward. As shown in Figure 2, the colors and directions of each patch were opposite to those of its adjacent neighbors. The color-motion pairs also alternated in time: for the red-inward test, a location that was red-inward in one array changed to green-outward in the next, and vice versa. The color-motion pairs alternated at each location with ISI of 0 ms (other than the interframe interval of the monitor refresh).

Preceding and following the sequences of test arrays were masking arrays that were identical to the tests,

alternating between red and green, except that the dot motion was replaced with noise having no net motion in either direction (randomly positioned on each frame, updated every 12 ms). The timing of the preceding and following masks was the same as for the test arrays and the masks were present at all 10 locations.

In the guide conditions, a white ring 4.0 degrees of visual angle in diameter with width of 0.2 degrees was placed around one of the locations in each array. It remained there until the next array was presented whereupon it stepped to the adjacent location, either clockwise or counterclockwise (depending on the guide direction of the trial). The guide was present and in motion, stepping from location to location, throughout the trial.

### Procedure

A trial began with only a central fixation point on the screen and the presentation of a short tone. In all cases, observers fixated throughout the trial. Following a delay of 500 ms the sequence of masking and test arrays began. There was first a sequence of 3 mask arrays of contrast increasing linearly to test levels followed by 6 test arrays and then ending with 3 mask arrays dropping linearly in contrast. The pre- and post-masking provided by the first and last three mask arrays avoided a sudden onset or offset that could allow successful identification even at very high alternation rates (Beaudot, 2002; Dakin & Bex, 2002). Observers reported whether red was paired with inward or outward motion by pressing the 'z' key with the left index finger or the '/' key with the right index finger, respectively, on a keyboard in front of them. The observers had unlimited time for response but the display terminated with the key press if the response occurred before the end of the sequence of test arrays. Following the response, there was a delay of 500 ms to the beginning of the next trial.

Eight different rates were tested in random order, 8 times each, in each session. The 8 rates were linearly spaced over a range that bracketed the observer's threshold as estimated in a practice session. Each observer repeated the 64-trial sessions at least 4 times for the guided condition and 4 times for the no-guide condition. On half the trials, red was paired with inward and on the other half, red was paired with outward motion. The observer needed to report only one of the pairings and so could completely describe the stimulus by pressing 'z' or '/' on the keyboard in front of them. In the stationary attention, no-guide condition, observers attended to one location, with the bottom location suggested although observers were not discouraged from noticing pairings at any other location. They were asked not to track any motion they might see moving around the display. Most did not perceive any, as tracking attentively when the guide was not present seemed possible only for practiced observers, and even then, only at slow to moderate rates

and with significant effort. All reported that they had not engaged in any tracking during the no-guide, stationary attention conditions. On guided sessions, observers were asked to attend to the stepping ring and report the pairing they saw within the ring. The guide ring stepped clockwise around the display on half the trials and counterclockwise on the other half. The ring appeared and began stepping 500 ms before the onset of the stimulus arrays providing time to acquire and track the guide before the arrays appeared. The starting location of the guide was set so that it arrived at the bottommost location on the 3rd of the 6 test arrays. On half the trials, the guide encircled a red patch and on the other half of trials it encircled a green patch. In both cases the observer reported the red direction (if green was seen, either in the guided or stationary attention cases, the observer reversed the direction to generate the appropriate response for red).

### Results

The percent correct was plotted as a function of rate for each observer and the best-fitting Probit curves were used to determine the 75% thresholds (Figure 3).

During the stationary attention trials, observers were attending to an alternating stimulus and, based on earlier results (Moradi & Shimojo, 2004), we expect accurate reporting of the pairing only at slow rates. Indeed, the average threshold rate for reporting the pairing with 75% accuracy was about 2.2 Hz (range 1.8 to 2.5 Hz), similar to the value reported by Moradi and Shimojo (2004) for a single test patch of alternating color and motion, viewed in the fovea.

At faster alternation rates where the pairing of features in the stimuli could not be clearly distinguished locally, the tracking enabled by the stepping guide ring allowed the pairings to be easily seen and accurately reported. While tracking, the threshold rate at which the color and motion pairing was reported with 75% accuracy ranged from 3.6 to 6.2 Hz (4.9 Hz averaged across the 5 observers). The difference between the two conditions across observers was significant ( $p < 0.01$ , paired  $t$ -test) and the threshold rate with moving attention was, on average, 2.2 times that found when observers attended to just a single location.

The next experiment presents several control conditions to determine if the advantage for moving attention is simply due to the presence of the guide ring (no), is limited by the presence of any perceptual asynchrony in registering color and motion (no), depends on the perception of motion (yes), is due to the reduction in interruption from the competing color and motion which are no longer part of the attended stream (partially), and increases with the number of locations visited (yes). The results from these controls allow us to attribute the

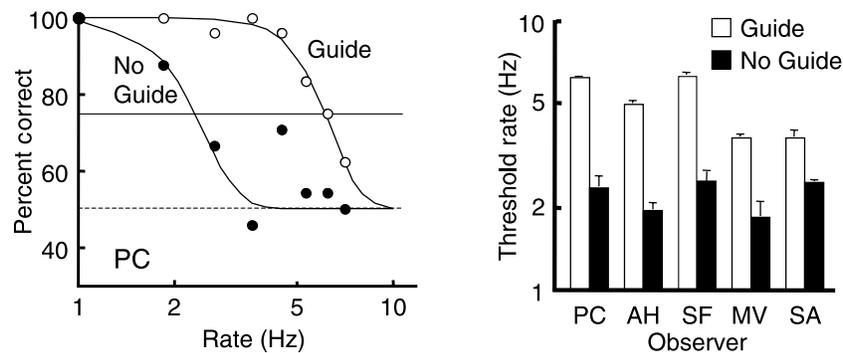


Figure 3. Accuracy of color–motion report. Left panel: Accuracy for guide and no-guide sessions is shown for one observer, PC, as a function of alternation rate (plotted on a log scale). Dashed horizontal line is the chance level. Solid horizontal line is the 75% threshold criterion. Right panel: Threshold rates are shown (on log scale) for 5 observers with white bars for the thresholds with moving attention and black bars for the thresholds with stationary attention (no guide). Vertical lines above the bars show +1 SE. Threshold rate for reporting correct pairing of color and motion is 3.6 to 6.2 Hz when following the moving target (white bars). Threshold is only 1.8 to 2.5 Hz when attending to a stationary location (black bars).

facilitation of the guide to a mobile analysis, an accumulation of partial results in an object-based, non-retinotopic representation.

## Experiment 2: Color and motion controls

In [Experiment 1](#), performance with moving attention was significantly improved compared to performance with stationary attention. This might be due to an effect of the guide itself, acting as a transient cue and here we investigate alternative ways of presenting the guide ring to distinguish any local effect of the guide from effects due to the movement of attention that the guide directs.

To address the possible local cuing contributions of the guide ring we used 1) a ring that flashes around a single

pairing (say, red-inward, [Figure 4](#)) each time it occurs at a single location, 2) multiple rings that encircle all the pairings of one kind (all red-inward, [Figure 5](#)) in each array, 3) one ring on each array that always encircles the same pairing (say, red-inward) but at a random location for each successive array (not shown), and 4) we also tested the effect of doubling the number of successive test arrays (from 6 to 12) between the leading and trailing masks so that the number of arrays with the same stimulus values (say, red-inward) at any one location will match the number of arrays with those stimulus values encircled in the moving ring condition: 6 in both cases.

There is another factor, color–motion asynchrony (Moutoussis & Zeki, 1997), that we need to consider in interpreting the threshold rates for reporting color and motion in our stimuli. Using a large central patch of moving dots that alternated in color and direction, Moutoussis and Zeki (1997) reported that the motion reversal had to occur about 100 ms before the color

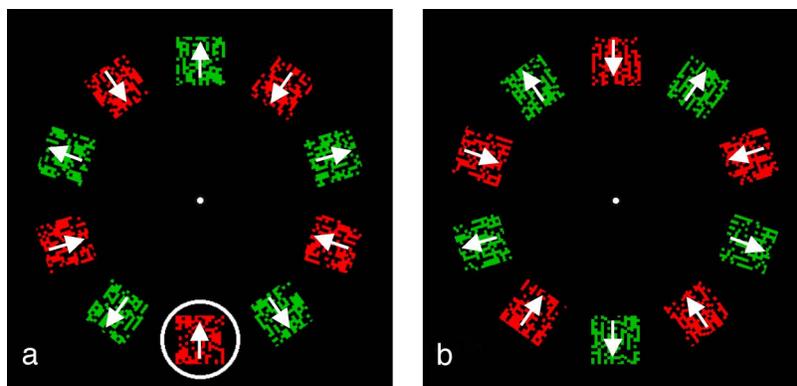


Figure 4. Single ring control. Arrays a and b alternate as in [Experiment 1](#) but now a white ring cues the target pairing (red-in here) at a fixed location every time it occurs there. The observer fixates the central dot and attends to the bottom location to report the color and motion pairing within the ring. Click on the link [here](#) to see a demonstration movie at an alternation rate of 4 Hz where reporting the pairing of color and motion within the guide may be difficult.

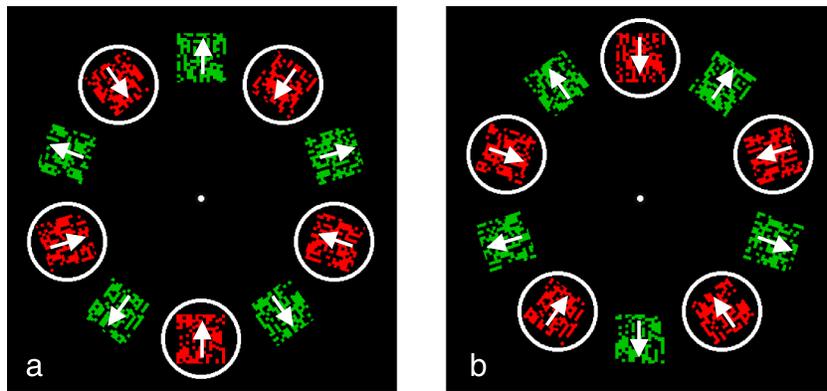


Figure 5. Multiple ring control. Arrays a and b alternate and all target pairings (red-in here) are encircled on each frame. The observer fixates the central dot and can attend anywhere but must avoid tracking any motion of the rings. Click on the link [here](#) to see a demonstration movie at 4 Hz where reporting the pairing of color and motion within any of the guides may be difficult.

reversal for the two changes to appear optimally aligned. This is important because any perceptual asynchrony between the color and motion will reduce the proportion of time that the percept has the same pairing as the physical stimulus. This change in the perceived pairing can by itself limit the threshold rate for accurate reporting (Arnold, 2005). For example, at an alternation rate of 2.5 Hz, a 100 ms asynchrony will shift the perceived alignment by  $\frac{1}{4}$  cycle. With a  $\frac{1}{4}$  cycle shift, both colors will be paired for an equal duration with both motions and the observer can do no better than 50% correct. Clearly, the values of asynchrony of about 100 ms seen in the literature could not be present in our condition with moving attention since the 75% threshold was about 5 Hz, well above the rate of 2.5 Hz where the asynchrony would force performance to drop to chance. However, in our condition with stationary attention, the rate threshold was a little above 2 Hz and this could be a consequence of color–motion asynchrony. To evaluate the effect of asynchrony, we measured the perceived asynchrony in our stimulus with stationary attention and with moving attention and tested the accuracy of report again after compensating for the asynchrony.

Finally, we examine whether the advantage of the guide is due to accumulation across locations or just to the absence of interruption (interruption masking, Breitmeyer, 1984; Scheerer, 1973; Smith & Wolfgang, 2004; Spencer & Shuntich, 1970). For the stationary attention condition, the observer must process the sequence of alternating pairs, “red-in” then “green-out”, for example, and even though the “red-in” information may be available from the attended location in the first array, its processing is apparently interrupted by the subsequent “green-out” at the same location that arrives before the processing of the “red-in” combination is completed. With moving attention, identical pairings may be selected in successive arrays and so the processing can potentially continue uninterrupted. If interruption of processing were the only

cause of poor performance with attention at a single location, then a guide that passed over just one location (with all other locations in the array blank, see Figure 8) ought to retrieve sufficient information for accurate identification. Critically, processing could continue to completion uninterrupted by subsequent feature pairs as attention would now move on to only blank locations.

However, an alternate explanation of the poor performance for stationary attention is that the processing of the information from each array is not only interrupted before completion but it is also insufficient to support identification of the color–motion pair even if it could run to completion. The advantage of moving attention may be not only the avoidance of interruption but also the opportunity for accumulating partial results from successive locations. This would require an accumulation in an object-specific representation that is independent of location.

To test whether one sample is enough or whether post-selection mechanisms can accumulate information across locations, we first determine the threshold rate when moving attention samples only a single location (avoiding interruption masking but not providing multiple samples). We then present from 1 to 5 adjacent test locations in sequence (Figure 8) to determine whether additional samples improve performance. We also compared moving attention to stationary attention to contrast the rates of accumulation in the two cases.

## Methods

### Participants

The participants included one author, PC, and 2 observers, MV and SA, who had participated in Experiment 1. The observers for the asynchrony controls were the three authors and one observer from Experiment 1, MV. Observers ranged in age from 28 to 59 years and had

normal or corrected-to-normal visual acuity and normal color vision.

### Apparatus

Unchanged from [Experiment 1](#).

### Stimuli

The same stimuli were presented as in [Experiment 1](#) with the following exceptions. In the stationary guide condition, the guide appeared on every second array but always at the same location, so that it always encircled the same pairing, for example, red-in. The guide was not presented when the complementary pair, green-out ([Figure 4](#)) was at that location. In the multiple guide condition, every second location had a guide ring and these locations alternated in each array so that all the guide rings encircled the same pairing at each location and in every array ([Figure 5](#)). In the long duration version of the stationary guide condition 12 test arrays were presented instead of the 6 presented in all other conditions.

In the asynchrony control, the timing of the motion reversal relative to the color reversal was shifted over 8 possible values in order to identify the relative timing that provided optimal performance, all at a fixed rate (ranging from 1.9 Hz, MV, to 2.45 Hz, AH) in one condition with stationary attention and no guide ring and in a second condition with moving attention. The relative timing for the optimal response identifies the temporal shift that maximally aligns the two features in the percept, providing the best chance for accurate report. Once this perceptual asynchrony was established in both cases, the stimulus, otherwise identical to that of [Experiment 1](#), was corrected by this amount and a further session was run to determine the threshold rate with asynchrony compensated.

In the single sample condition (test of reduced interruption masking), only the bottom location was presented and only one target was displayed (preceded and followed by the non-target pair and the masking arrays). In the moving attention version, the starting location of the ring was set so that it arrived at this location when the target was there.

In the test for accumulation with moving attention, the number of locations with color–motion pairs present was varied from 1 to 5 ([Figure 8](#)) and there were always 6 test arrays (preceded and followed by the mask arrays). When there was only one location presented, it was the bottom-most location and additional locations were added to the left and right in a balanced order. The starting location of the ring was set so that it arrived at the bottom-most location on the 3rd of the 6 test arrays. In the test for accumulation with stationary attention, either 2 or 10 test arrays were presented giving either 1 or 5 arrays with the target encircled with the guide ring, always at the bottom-most location.

### Procedure

The procedure was the same as [Experiment 1](#) with the following exceptions. All trials, except for those of the asynchrony, single sample and accumulation with stationary attention controls, were at one fixed alternation rate that was selected individually for each observer based on their data from [Experiment 1](#) to generate low accuracy responses with stationary attention and high accuracy with moving attention. This frequency ranged from 3 Hz (MV, SA) to 4 Hz (PC). In the single and multiple guide conditions, observers were asked to report the motion that they saw paired with red within the guide rings (or its complement if green was seen) and to avoid tracking the stimuli or the guides in any motion trajectory. In the random and moving guide conditions, observers were to track the guide as well as they could and report the encircled pairing (or its red complement). For the asynchrony evaluation, the performance was tested at 8 different relative timings with 8 trials per timing in random order. Then we created a new stimulus that compensated for the asynchrony and the frequency threshold with this stimulus was estimated, again using the procedure of [Experiment 1](#). For the single-sample control and the accumulation with stationary attention control, presentation rate was varied as in [Experiment 1](#) in order to determine a threshold.

## Results

### Controls for effects of the ring

The results in terms of percent correct at the single presentation rate (3 Hz or 4 Hz, depending on the observer) are given in [Figure 6](#) for the 4 control conditions and the No Guide and Moving Guide comparison conditions. First of all, the No Guide and Moving Guide comparisons replicate the results of [Experiment 1](#) where performance improves significantly with the moving guide and is near chance for stationary attention. None of the 4 control conditions improved performance indicating that the presence of the ring itself was not the factor that caused the improvement. A single guide at the same location on every second array, always encircling the same color–motion pairing, gave no advantage even if the trial was doubled in duration (to match the total number of encircled patches for the Single and the Moving guide, rings presented with target patch 6 times in each case). Placing a ring around every target pairing on every array did not help either. Only the moving guide ring improved performance. Moreover, given that the randomly moving guide did not help, it may be that the ring must move in a predictable manner if it is to facilitate the report of the target features. The predictability of the motion may allow attention to arrive at each location in synchrony with the ring just as smooth pursuit eye movements can become exactly synchronized with predictable motions overcoming

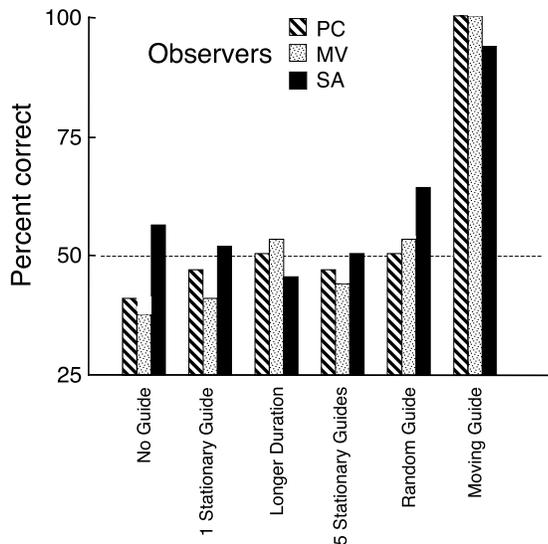


Figure 6. Control experiments. Accuracy data from 3 observers at 3-Hz rate for MV and SA and 4 Hz for PC. The dashed line at 50% represents chance performance. The No Guide and the Moving Guide conditions duplicate the stationary and moving attention conditions of [Experiment 1](#) but at the fixed rate (3 or 4 Hz). The 1 Stationary Guide condition had a ring at the bottom location on alternate frames, always encircling the same pairing ([Figure 4](#)). The Longer Duration condition was the same as the 1 Stationary Guide but with twice the number of test arrays. The 5 Stationary Guides had a ring around all target pairings on each array ([Figure 5](#)). In the Random Guide condition, the guide moved to random positions on each array and observers attempted to attend to it. Click on the link [here](#) to see a demonstration movie at 4 Hz where reporting the pairing of color and motion within the randomly moving guide may be difficult.

the phase lag that the inevitable neural delays on feedback should impose (Bahill & McDonald, 1983).

### Asynchrony controls

Our measurements of perceptual asynchrony revealed only small effects. In previous work with very similar displays (Holcombe & Cavanagh, 2008), we have found that the onset of the ring at the same time as the motion and color change reduces or eliminates any perceptual asynchrony between motion and color. Our measurements here showed that the motion reversal had to lead the color reversal by about 30 ms to maximize performance with stationary attention. This asynchrony is much smaller than that seen in the stimulus of Moutoussis and Zeki (1997). There may be a number of reasons for the reduced asynchrony with our display (for example, motion speed and eccentricity are quite different from the values used by Moutoussis & Zeki, 1997) but an asymmetry this small will have little effect at the limiting rates we found with stationary attention (about 2 Hz). Specifically, a 30-ms

asynchrony cannot explain the chance level performance of our observers at 4 Hz (75% at 2 Hz). This asynchrony will only bring performance to chance when the rate reaches 8 Hz where 30 ms represents about a  $\frac{1}{4}$  cycle shift between color and motion, so that both colors will be paired for an equal duration with both motions and the observer can do no better than 50% correct. With moving attention and a guide ring, the asynchrony we measured was again about 30 ms but observer threshold here was higher, about 6 Hz, so conceivably the asynchrony may have contributed to this limit. However when we corrected for the asynchrony (in 2 observers) we found no systematic improvement in the limiting rate either for stationary attention, or moving attention.

### Single-sample controls

The results of the single-sample control conditions demonstrate that moving attention can evade, to some extent, the interruption masking from the stimuli alternating at each location. When only a single target is sampled, the threshold rate is higher for moving attention than for stationary attention (striped bars versus solid bars in [Figure 7](#)). In both cases, the target patch was preceded and followed by one patch of the complementary pairing at the single test location. With moving attention sampling a single patch, there are no color–motion pairs preceding or following the target pair in the attended stream as there are when attention remains at the test location and this absence of interference may lead to the observed improvement.

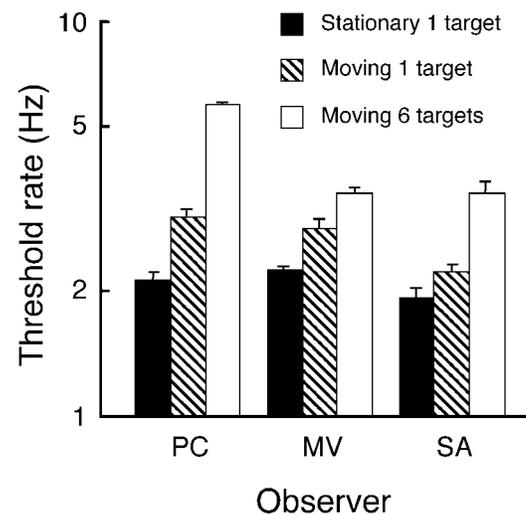


Figure 7. Single-sample control. Solid bars show the thresholds for a single target encircled by a guide ring with attention held at the target location. When attention follows a moving guide and passes through the same single target, threshold rates are higher (striped bars). For comparison, thresholds of these observers from [Experiment 1](#) are shown for moving attention (sampling 6 target locations) with outline bars. Vertical lines above the bars show +1 SE.

Multiple successive samples of the same pair in the attended stream provided a further benefit, as is evident by comparing the threshold rate for moving attention with a single sample to that for the multiple samples of [Experiment 1](#) (data from [Experiment 1](#) for these 3 observers are shown in outline bars in [Figure 7](#)). The first result—that moving attention is able to reduce the interruption masking with a single target—indicates that interruption masking occurs to some degree in the attended stream and not exclusively in retinotopic coordinates.

### Accumulation controls with moving attention

The second result—that the reduction in masking does not account for all the advantage seen in [Experiment 1](#)—is demonstrated directly in the final control experiments. These accumulation controls ([Figure 8](#)) showed that the remaining advantage was due to subthreshold summation across locations. Specifically, at the fixed rates (3 Hz, MV and SA, 4 Hz PC) used for this control, performance is near chance when sampling only a single patch with the stepping guide ring. However, as the number of target samples increases, the performance improves, reaching asymptote at 3 or 4 locations (outline symbols, [Figure 8](#), right panel).

A performance benefit may reflect probability summation, the accumulating probability of a correct response given multiple, independent samples. This statistical improvement does not require any continuing visual analysis. However,

the curves do not show the initial negative acceleration that is a signature of probability summation where the probability correct should be  $1 - 0.5 \cdot (1 - p)^n$  with  $p$  as the probability of completing the stimulus analysis on any one sample, and  $n$  the number of samples. In addition, the performance can only improve if the probability of a correct response on a single sample is above chance and that is not apparent in the data for presentation of one or even two locations in [Figure 8](#). These results suggest, instead, subthreshold summation where a single mechanism accumulates information across space and time at a level prior to the decision stage.

### Accumulation controls with stationary attention

In the stationary attention condition of the accumulation control, we tested 1 and 5 repetitions of the target pair at one location with stationary attention (a guide ring on alternate arrays at the single test location). Rather than improving, performance remained at chance as additional alternations were added (filled symbols in [Figure 8](#)). This chance performance was recorded at the same rates, 3 Hz (MV and SA) or 4 Hz (PC), where accumulation was found with moving attention. The similarity in performance with 1 and 5 samples with stationary attention at these fixed rates was not a floor effect, as at all rates tested, including slower, easier rates, performance remained similar for 1 vs. 5 samples (the thresholds for 1 and 5 samples were both close to 2.2 Hz for all

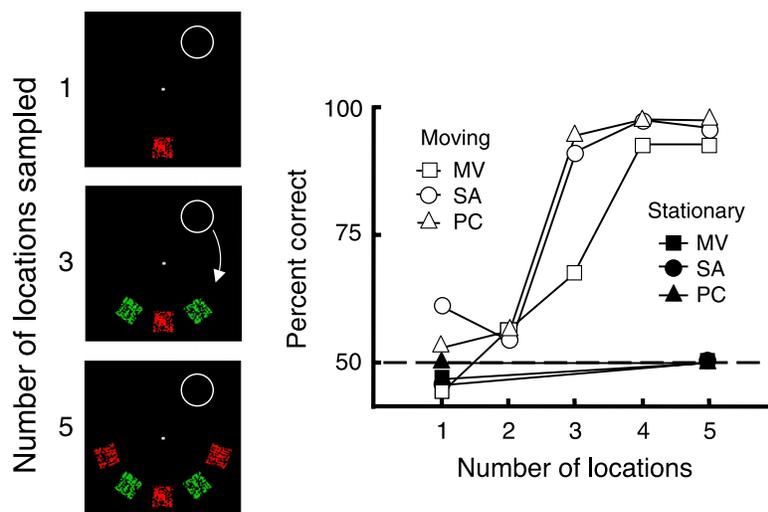


Figure 8. Accumulation control conditions. Moving: One through 5 patches of alternating color and motion were presented and, while fixating the central dot, observers tracked the moving ring. The outline symbols in the right-hand panel shows the percent correct report of the color and motion pairing present in the ring—the ring steps in synchrony with the alternation so the same pairing is present in the ring at each filled location. Performance is near chance for all 3 observers when the ring's trajectory samples only 1 or 2 locations but rises to very high accuracy once 3 or 4 locations are sampled. Stationary: One or 5 target patches are presented at a single location, alternating with the complementary pair. The guide ring encircles only the target pair and always at the same location. Performance does not improve (filled symbols) with additional samples. Click on the links here to see demonstration movies of the guide ring stepping around the display and sampling [one](#), [three](#) or [five](#) locations at a 4-Hz alternation rate where reporting the pairing of color and motion within the moving ring should become easier as the number of locations increases.

3 observers). This result suggests that the improved detection of one of the two alternating stimuli only occurs over multiple presentations when it is the sole attended item across the successive samples, as is the case in the moving attention conditions. Not only does moving attention reduce the interruption masking from the alternation, it also permits accumulation across locations.

### **Spatial and temporal extent of accumulation**

These results also allow a rough estimate of the temporal interval over which stimulus features can be integrated by moving attention. Specifically, the performance increases until about 3 or 4 locations have been sampled, corresponding to an interval of about 300 to 500 ms. It is possible that the integration period is longer than 300 ms and that spatial factors limit the range over which accumulation occurs (perhaps the limited size of large receptive fields performing the analysis). Nevertheless, we can say that the accumulation spans at least 300 ms.

In a related finding, Nishida et al. (2007) recently showed integration of color along a motion path. Individual bars were presented in an apparent motion sequence with the color of each bar alternating between red and green as it moved to each new location. The color seen for the moving bar was yellow, the sum of the two colors implying a mobile integration process. Rather than stepping across a gap as our guide did, successive bars in successive frames in their case were adjacent without a gap. The step size was varied from 3 to 12 minutes of visual angle with best color fusion seen for steps smaller than 10 minutes so the mixing of the colors of successive bars might have been mediated by fairly small local, directionally selective receptive fields. In contrast, the present experiments show accumulation of information from samples separated by 4 degrees, larger than receptive field at this eccentricity in V1 (Smith, Singh, Williams, & Greenlee, 2001). We can also estimate the total spatial span over which accumulation is occurring. Performance in our data reached asymptote once 3 or 4 locations are sampled, corresponding to a spatial span of accumulation of 8° to 12° of visual angle at the 6.5° eccentricity of the stimulus patches.

This 8° to 12° span is larger than even MT or V4 receptive fields (V4 receptive fields span about 6 degrees at 6.5 degrees eccentricity, Gattass, Sousa, & Gross, 1988; MT cells span 6 to 7 degrees at 6.5 degrees, Gattass & Gross, 1981) but well within the large receptive field size for MST cells (often 50° diameter or more, Duffy & Wurtz, 1991a; Tanaka & Saito, 1989). MST is therefore one possible site for integrating motion information across locations. The radial motion we use is among the flow field patterns that activate many MST cells (Duffy & Wurtz, 1991a; Tanaka & Saito, 1989) and our presentation differs from localized stimuli sometimes used to probe

subregions of MST units (e.g., Duffy & Wurtz, 1991b) only in that our stimuli are presented at different times and locations—and presented along with opposing motion at other locations which would normally cancel any cell's response to its preferred organization. For MST to respond to and integrate the single, tracked stimulus as it jumps from location to location, we must rely on strong gating of MST activity by attention (Treue & Maunsell, 1999). Controls run with translating as opposed to radial motion in each patch showed similar results (data not shown), so there appears to be nothing special about the radial motion itself. Although motion integration might be mediated within MST, we have no specific suggestion for color integration other than to suggest that it lies at or beyond area V4/V8 and might be mediated by IT units with large receptive fields that are selective for color (Komatsu, Ideura, Kaji, & Yamane, 1992).

In sum, we have evidence for two factors contributing to the improved performance with moving attention. First, interruption masking appears to set the performance limit when reporting a target that is preceded and followed by a competing stimulus and this interruption masking does not appear to be set in retinotopic coordinates; it can be avoided by sampling the stimulus array in a trajectory that skips over the competing stimulus at each location so that the attended stream samples only the target and not the competing nontarget. Second, the accumulation results establish that the target information can be accumulated across positions even when the target signal is subthreshold at each location.

## **Experiment 3: Reporting individual color and motion features**

In the previous experiments, observers reported which color was paired with which motion and the results show a clear improvement in reporting the pairing when attention followed a moving guide, sampling successive locations containing the same pairing. Here we examine whether the advantage also occurs for reporting individual features when indexed by the guide ring. In this experiment, when the observers have to report color, red and green alternate at each location but the motion is the same everywhere, either inward or outward. When the observers have to report the motion, the color is the same everywhere, either red or green. Since there is no consistent pairing of color and motion to indicate which color or which motion to report, the ring must always be used to cue the to-be-reported feature, either flashing on and off at a fixed location to always encircle, say, red, or moving around the locations to always encircle the target color or motion. If

the threshold rate for reporting the color–motion pairing within a ring in the previous experiments was limited by the additional step of binding the two features, then reporting individual features may be possible at higher rates than reporting the pairing (as found by Bartels & Zeki, 2006 and Bodelón, Fallah, & Reynolds, 2007, with different stimuli). We also test for accumulation of single-feature information across locations using a moving guide that traversed from 1 to 5 adjacent locations (like Figure 8) but observers now reported only color or only motion values seen within the ring.

## Methods

### Participants

The 3 participants including one author, WL, had not participated in Experiment 1 or 2. Observers ranged in age from 26 to 28 years and had normal or corrected-to-normal visual acuity and normal color vision.

### Apparatus

Unchanged from Experiment 1.

### Stimuli

The same stimuli were presented as in Experiment 1 with the following exceptions. In the color-only condition, adjacent locations alternated between red and green and the colors also reversed from one array to the next. The motion was the same at all locations in each array, either all in or all out throughout the trial and this direction was set randomly on each trial. In the motion-only condition, adjacent locations alternated between inward and outward and the directions at each location reversed from one array to the next. The color was the same at all locations in each array (either all red or all green) throughout each trial and this color was set randomly on each trial. In the stationary guide condition, the guide appeared on every second array but always at the same location, so it always encircled the same stimulus, for example, red in the color-only condition. The guide was not presented in the next array when the complementary feature, green, was at that location. In the moving guide condition, the guide ring jumped to the adjacent location in each successive array so that the ring always encircled the same feature. In the variable locations condition (test for accumulation), the alternating stimuli were present at either 1, 3, or 5 adjacent locations of the 10 locations visited by the guide in its trajectory around the display (either clockwise or counterclockwise). When there was only one location presented, it was the bottommost location and additional locations were added to the left and right in a balanced order. The starting location of the ring was set so that it arrived at the bottommost location on the 3rd of the 6 test arrays.

## Procedure

The procedure was the same as Experiment 1 with the following exceptions. In the color-only conditions, observers reported the target color seen in the guide ring by pressing the “/” key for red and the “z” key for green. In the motion-only conditions, observers reported the target motion seen in the guide ring by pressing “/” for inward and “z” for outward. In the accumulation conditions, all trials were at one fixed alternation rate that was selected individually for each observer based on their data from main results for color-only and motion-only conditions. The rate was chosen to generate low accuracy responses with stationary attention and high accuracy with moving attention. These frequencies were 4 Hz (CH), 4 Hz (SD), and 3 Hz (WL) in the motion-only condition and 5 Hz (CH), 5.5 Hz (WL), and 9 Hz (SD) in the color-only condition.

## Results

For the main results in the color-only and motion-only conditions, the percent correct was plotted as a function of rate for each observer and the best-fitting Probit curves were used to determine the 75% thresholds (Figure 9).

During the stationary attention trials, the average rate at which observers could report the motion within the ring at 75% accuracy was about 2.3 Hz (range 1.9 to 3.1 Hz), whereas the average rate for 75% accuracy for color was 3.5 Hz (range 2.5 to 4.3 Hz). Moving attention increased the 75% threshold rate for both features, to 4.2 Hz for motion (range, 4.1 to 4.3) and to a remarkable 9.3 Hz for color (range, 7.3 to 10.6). Individual results are shown in Figure 9. The difference between the tracking and stationary results was significant ( $p < 0.025$ , paired  $t$ -tests) in both conditions and the threshold rate with moving attention was, for motion, 1.8 times the maximum rate found when observers attended to just a single location and for color, 2.7 times.

Clearly moving attention facilitated access to these brief targets even when only a single feature was reported. In addition, as was the case for reporting color and motion together, the accumulation tests (Figure 10) showed an increase in accuracy of report, often from chance levels when only one target was sampled, up to high accuracy when 5 targets were sampled. At the highest alternation rates of the color experiment, the individual colors themselves appeared to observers to be lost, leaving only an impression of a flickering yellow (the individual target color was only revealed while tracking). Our experiment does not formally test whether the individual colors, red and green, are completely lost at high rates—this would require a second set of two colors that added up to the same yellow and a discrimination task where the observers must report which color pair was presented. Nevertheless, the subjective reports of

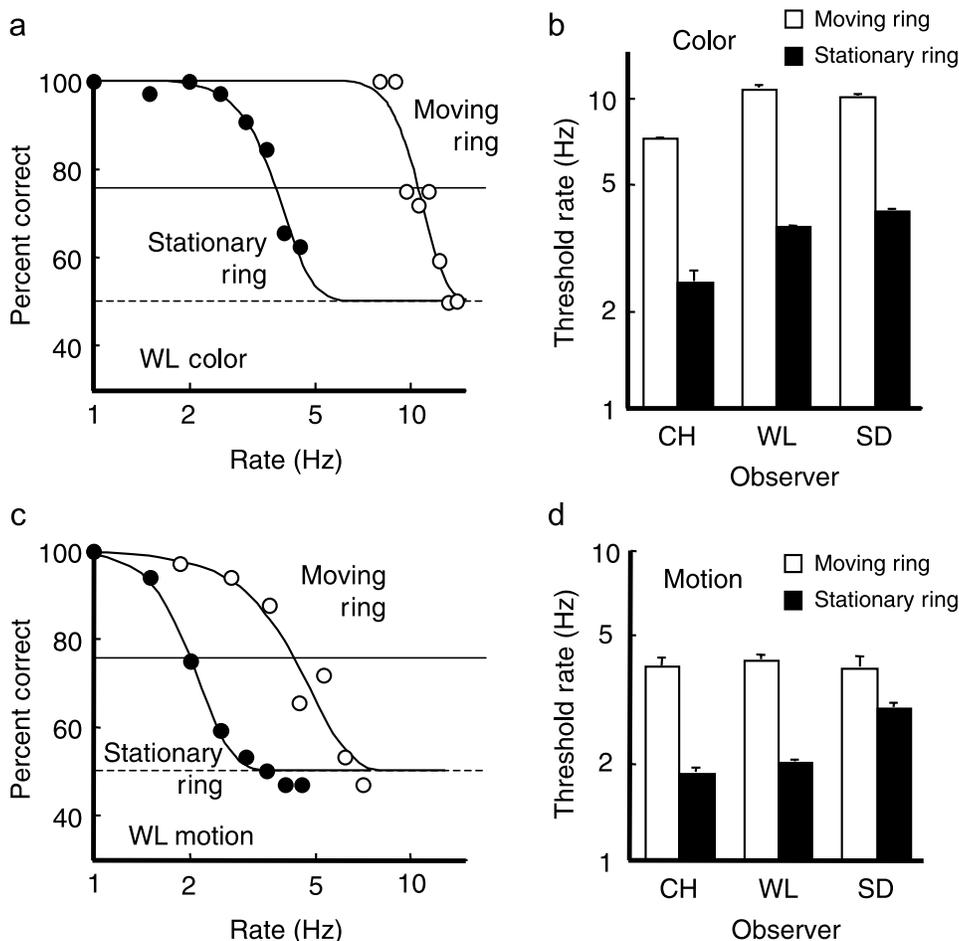


Figure 9. Accuracy of individual color and motion reports. (a) Accuracy for stationary ring and moving ring conditions is shown for color stimuli for one observer, WL, as a function of alternation rate (plotted on a log scale). Dashed horizontal line is the chance level. Solid horizontal line is the 75% threshold criterion. (b) Threshold rates for color stimuli are shown (on log scale) for 3 observers with white bars for the thresholds with moving ring and black bars for the thresholds with stationary rings. Vertical lines above the bars show +1 SE. (c) Accuracy for motion stimuli, observer WL. (d) Threshold rates for motion stimuli.

color summation that can be undone by motion are similar to the recent finding by Watanabe and Nishida (2007). In their study, however, the moving patterns

had small offsets between successive positions (3 to 12 minutes of arc), and thus might involve more local mechanisms.

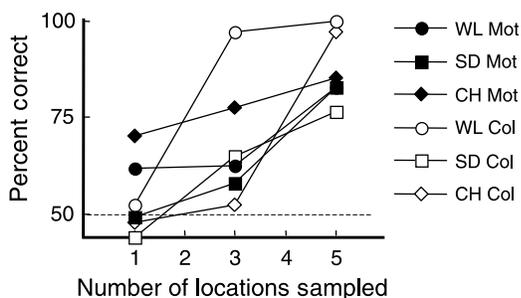


Figure 10. Accumulation for individual features. The percent correct report of the color (outline symbols) or motion (filled symbols) is shown as a function of the number of locations sampled for 3 observers (WL, circles; SD, squares; CH, diamonds). Accumulation is seen in all cases but more so for color.

Overall, motion required slower rates than did color and the rate for 75% correct motion responses was a close match to the threshold rate for the combined color and motion reports of Experiment 1 (although for different observers). It would appear that the rate for the combined report was determined by the slowest of the two individual features. For stationary attention, the threshold rate was  $2.3 \pm 0.36$  Hz when reporting only motion and  $2.2 \pm 0.14$  Hz when reporting the pairing of color and motion (Experiment 1). For moving attention, the threshold rate was  $4.2 \pm 0.06$  Hz when reporting only motion and  $4.9 \pm 0.56$  Hz when reporting both color and motion (Experiment 1). We are comparing different observers in the two experiments so we can draw no firm conclusions, but the results do suggest a small temporal cost at most for combining or binding the two features beyond that required to encode each separately.

Conjunction costs have been reported in previous studies. Bodelón et al. (2007), for example, documented a small but non-zero cost for binding color and orientation; however they used a display that allowed feature binding at much higher rates (Holcombe & Cavanagh, 2001). Bartels and Zeki (2006) showed that the threshold rate for reporting individual color or motion features was higher than the threshold rate for reporting the pairing of the two values, but they tested observers with spatially separate color and motion patches. Moreover, success on their task required perception only of the axis of motion (e.g., left–right vs. up–down), meaning that observers may have perceived only the motion streaks rather than the motions themselves.

## Experiment 4: Alternating target and mask

Does the moving attention paradigm improve all tasks that are limited by exposure time? Our assumption is that moving attention avoids local temporal limits by allowing access to an unchanging stimulus in a moving window (sidestepping the interruption masking that would occur at a stationary location). Moreover, as we showed in the accumulation results (Figures 8 and 10), the information picked up in the moving window appears to be partial at each location and accumulated across locations. This accumulation could call on high-level visual areas with large receptive fields to mediate the continued analysis of the stimulus across time and space. In this experiment we

use a moving attention window with a forward- and backward-masked stimulus to see whether moving attention can reduce the effects of a noise mask. In the display, a high-contrast mask and low-contrast target (a letter or a digit) alternate at each location (Figure 11). Because the alternations are 180° out of phase at each successive location, the moving ring can be set to always encircle the target. Thus, in the moving frame of the attention window, the masks and the target are never superimposed (except by virtue of any imprecisions in attentional tracking, Hogendoorn, Carlson, & Verstraten, 2007); whereas in retinotopic coordinates, the masks and targets are always superimposed. If the mask acts through a low-level process, such as integration masking, which is retinotopic (Breitmeyer, 1984; Schiller, 1966), moving attention should not be able to avoid this local stimulus loss. However, if the masking occurs at a higher level, like interruption masking (Breitmeyer, 1984) or object substitution masking (Lleras & Moore, 2003), then moving attention may dodge the effects of the mask by keeping the target separate from the mask in the attended stream. In any case, even if moving attention cannot evade the masking, it may still allow accumulation of partial information across locations. Finally, a performance benefit, if it occurs, may also reflect probability summation. We already noted that the shape of the accumulation curves in Figure 8 (and to some extent Figure 10) do not show the initial negative acceleration predicted by probability summation. Nevertheless, we will evaluate this possibility again.

In contrast to our experiments with color and motion, we find that moving attention provides no improvement in performance with masked letters and digits. Based on the absence of an advantage for moving attention with

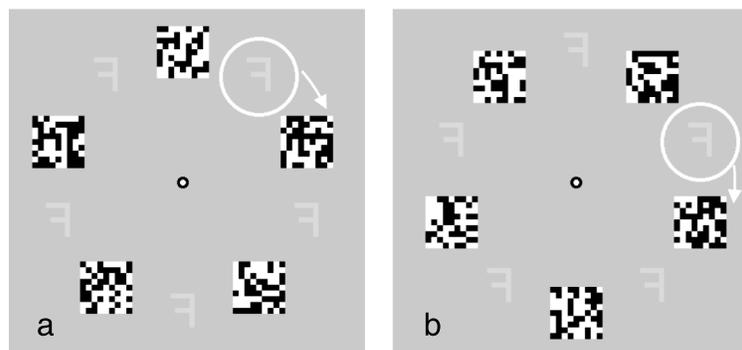


Figure 11. Letter/digit masking. Arrays a and b. In the stationary attention condition, the mask and letter or digit target alternate at each location, but no guide ring is present. The observer attends to one location while fixating the central dot and reports whether the letter or digit is a normal or left–right reversed version (reversed here). In the moving attention condition, the observer also fixates the center but now tracks the guide ring with attention as it continues around the display (only two positions shown here). In each array it encircles the target and never the mask. The observer again reports whether the target is normally oriented or left–right reversed. Click on the links here to see demonstration movies of the stationary attention display at a 2.5-Hz or 5-Hz alternation rate, or the moving attention condition at a 2.5-Hz or 5-Hz alternation rate. Reporting the target orientation should be easier at 2.5 Hz with or without the guide, and difficult at 5 Hz, with or without the guide.

masked targets we will suggest that that the masking effect here is retinotopic (as in integration masking, Breitmeyer, 1984) and also that the target signal that remains after masking is not in a form that can be integrated across locations.

## Methods

### Participants

The participants included authors PC and AH and 2 naïve observers. Observers ranged in age from 18 to 58 years and had normal or corrected-to-normal visual acuity and normal color vision.

### Apparatus

Unchanged from [Experiment 1](#).

### Stimuli

Stimuli were placed at each of 10 locations equally spaced around a circular array of 13 degrees in diameter centered at fixation. At each location, a letter or digit in one array alternated with a dot mask in the next. The phase of the target and mask alternation was opposite for adjacent locations. The target for each trial was a single character, randomly chosen from among the set of 16 horizontally asymmetric characters (B, D, E, F, G, K, N, P, R, S, Z, 3, 4, 5, 6, or 9) in Profont typeface displayed at 2.0 degrees height. The target was in normal or left–right mirror reversed version on each trial with the same version presented at all locations and arrays. The background was gray (38 cd/m<sup>2</sup>) and the target was lighter than the background with a contrast of 20%. The mask was a random pattern (2.2 degrees square) of 100% contrast black and white checks in a 9 by 9 array completely covering the target at that location.

In the guide conditions, a white ring 3.9 degrees of visual angle in diameter with width of 0.2 degrees was placed around one of the locations in each array. It remained there until the next array was presented whereupon it stepped to the adjacent location, either clockwise or counterclockwise (depending on the guide direction of the trial). The guide was present and in motion, stepping from location to location, throughout the trial.

The 6 arrays with both targets and masks were preceded and followed by 3 lead-in and lead-out arrays. The contrast of the target characters increased linearly from 0% to final test values during the 3 lead-in arrays and then dropped from test values to 0% during the 3 lead-out arrays. The gradual contrast ramps avoided a sudden onset or offset that could allow perception even at very high alternation rates (Beaudot, 2002; Dakin & Bex, 2002). The masks were present in all arrays from the beginning to the end of each trial, always at 100% contrast.

The presentation duration of the target and mask array was always 2 refresh frames of the monitor (24 ms). The alternation rate was varied by changing the ISI between the arrays, keeping both the target and mask energy constant. During the ISI, a blank field was shown (blank except for the fixation bull's-eye) that had the same luminance as the background of the target and mask arrays. The guide ring, when present around a target location, appeared halfway through the ISI that precede the target and stayed on until halfway through the following ISI, bracketing the two frames of target in the middle of its appearance.

### Procedure

The procedure was the same as [Experiment 1](#) with the following exception. Observers reported whether the target was in mirror ('z') or normal ('/') version.

## Results

The results ([Figure 12](#)) show that even though the mask was never within the moving guide, masking remained as strong with attention following the guide as it was when attention was directed to one fixed location. The average threshold rates were similar for moving attention (4.16 Hz) and stationary attention (4.02 Hz,  $t = 0.27$ , ns). This implies that the mask operates retinotopically, prior to the arrival of attention, because otherwise, moving attention could separate the test and the mask. This is in strong contrast to the object-level masking apparent in [Experiment 1](#) that was specific to the attended stream of items. In that experiment, in the stationary attention condition, the alternation of the color–motion pairs interrupted the processing of each. In the moving attention condition of [Experiment 1](#), however, the color–motion stimulus was the same in each new location and the analysis of the pairing could be continued across locations without interruption. Why did this same continued access to just one of the two stimuli here, the target, not help separate the alternating target and mask? Possibly because of persistence of the preceding mask at that location, the mask degrades the target an early level where the degree of damage is not influenced by whether or not the mask has been attended.

Even though the signal may have been corrupted, the performance at the threshold rate we used, 75%, is well above chance, meaning that useful information was being sampled, at least at some locations. However, that information apparently could not be combined with information from other locations. The absence of any advantage for moving attention here suggests a difference not only in the locus of masking (retinotopic, integration masking vs. nonretinotopic interruption masking) but also

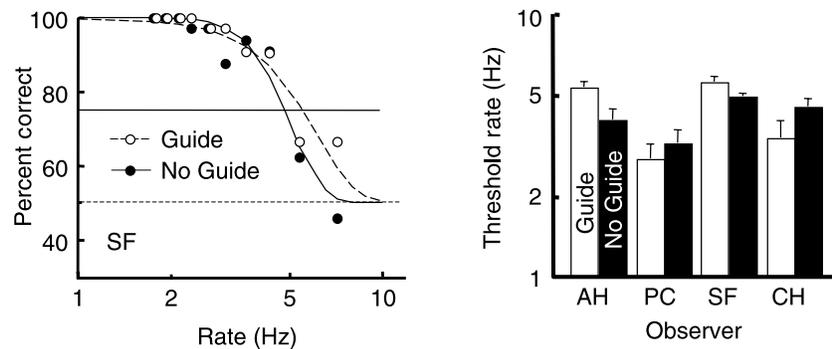


Figure 12. Threshold rate for reporting target orientation (mirror versus normal). Left panel: Accuracy data for one observer as a function of alternation rates shown on log scale. Guide data, open symbols and Guide fit, dashed line; No Guide data, black circles and No Guide fit, solid line. Dashed horizontal line shows chance level; solid horizontal line shows 75% threshold level. Right panel: 75% threshold rates on log scale for 4 observers. White bars for Guide conditions and black bars for No Guide. Thin vertical lines show +1 SE.

a difference in the nature of the information that is acquired at each location. We suggest that the inability to integrate across locations is attributable to the spatial nature of the letter and digit forms. Integration of partial spatial forms would require the appropriate spatial registration of samples from different locations. How local samples are combined by cells with large receptive fields is not well understood (experiments typically present multiple samples in competition, e.g., Motter, 1993) but clearly, if partial information is spatially incomplete (in the extreme, for example, only an individual horizontal, vertical or oblique stroke remains) then the information must be combined with the correct spatial alignment. This need for some object-based coordinates to guide the integration across locations within a receptive field is not required for the color and motion stimuli of the previous experiments. The color or motion information does not depend on any spatial structure that would need to be aligned across samples to reveal its identity and the information may summate usefully as long as the samples fall within a common receptive field. Of course, the improvements in Experiment 1, 2, and 3 with moving attention required more than just global summation within large receptive fields as the receptive fields, to perform any integration, must span several locations that have both targets and distractors. To avoid the distractors, the summation has to be gated by the location and timing of the target object within the receptive field, suggesting an object-based integration window.

The absence of improvement with moving attention also provides further argument against any contribution of probability summation to the accumulating performance seen in Experiments 2 and 3. In those accumulation tests (Figures 8 and 10), performance improved with increasing numbers of filled locations visited by the guide ring. The pattern of the accumulation already ruled out probability summation in those experiments, but here we have more direct evidence against probability summation because

now moving attention provides no improvement at all. Specifically, the performance at the threshold rate with stationary attention is well above chance (75% correct) indicating the presence of a signal. If that signal were available as a decision variable that could be accumulated in any way, it would have led to improved performance with moving attention. But that did not happen.

Note that the absence of improvement with moving attention cannot be attributed solely to the inability to integrate partial letter and digit information across locations. In Experiment 2, the single-sample control with moving attention (Figure 7) showed that there was an improvement in performance for moving attention (with color and motion) even if only one location was sampled. So if moving attention were able to sidestep the masking, it could produce a performance increase for the masked targets here even without any cross-location integration. The absence of an advantage for moving attention with masked targets therefore suggests both that the masking effect here is local (as in integration masking, Breitmeyer, 1984) and also that, even at the 75% performance level, there is neither a partial perceptual signal nor a partial decision variable that can be picked up at each location and usefully accumulated to improve performance.

## General discussion

We have found clear evidence of integration across locations for color and motion but not shape. Our technique allows precise control of local analysis and so provides a test for any further processing that can accumulate information across locations when the local information is verifiably subthreshold. This transpositional integration occurring when the object moves but the eyes do not may be linked to trans-saccadic

integration that occurs when the eyes move but the object does not. In both cases, processes that require the completion of local analyses can be separated from those that can integrate information across locations. The ability to integrate over large distances, during either eye or object motion, may rely on large receptive fields of higher level visual cortices, together with dynamic gating of information (Abbott, 2001).

In our first experiments, the pairing of color and motion could not be reported when the features alternated rapidly and attention was fixed at a stationary location. However, when attention moved to sample the same combination at successive locations, the pairing could be reported with high accuracy. Several control experiments demonstrated that this advantage of moving attention was not a consequence of the ring used to guide attention. Moreover, since a randomly moving guide did not improve performance, the predictability of the motion seems a prerequisite for any benefit of moving attention. The predictability may allow attention to arrive at each location in near synchrony with the ring (Hogendoorn et al., 2007), offering optimal target sampling without the inevitable delay of 80 ms or more required for attention to get to a randomly cued location (Nakayama & Mackeben, 1989). The control conditions also showed that the perceptual asynchrony between color and motion (Moutoussis & Zeki, 1997) was small in our experimental displays and did not measurably limit performance.

Some of the benefit of moving attention in our color and motion tasks was a result of avoiding interruption masking. This was demonstrated in the single-sample control where moving attention improved performance even for a single target location. The interruption of processing from the subsequent nontarget at the same retinal location could be avoided if attention moved on to an empty location. The interruption masking from the competing stimuli therefore appears to be confined to the sequence of attended items, whatever their location.

The second benefit of moving attention for color and motion features came from the accumulation of target analysis across space, as demonstrated in the accumulation controls. Performance improved when moving attention sampled additional locations, rising from chance levels, 50% accuracy, when sampling a single location to between 90 and 100% accuracy when sampling 5 locations. Further experiments with single features (color or motion) showed a similar advantage for moving attention indicating that the accumulation of processing can occur as well for individual features.

Finally, pre- and post-masking of targets by dot fields demonstrated that this type of masking is retinotopic and its effects cannot be avoided by attending to a path that samples only the targets and skips the masks. This unavoidable masking effect did leave partial information but it could not be integrated across locations. We suggest that the spatial detail that defines the letter or digit shapes cannot be easily accumulated from successive

locations due to the lack of adequate spatial registration across samples. In contrast, the color and motion features do not depend on spatial structure for their identity and can be accumulated over locations even with no registration.

### **Superposition masking is retinotopic; interruption masking is not**

The results of [Experiment 4](#) with superimposed targets and random dot masks showed that the advantage of moving attention in our paradigm does not hold for all stimuli. The masked threshold rate for identifying the orientation of the target letter or digit was around 4 to 5 Hz (target to mask SOA of 100 to 125 ms) for both stationary and moving attention. The absence of any improvement implies that the masking was retinotopic: it could not be avoided by attending to the target and not to the mask. If the mask was never at the attended location, but the target was nevertheless degraded, it implies that the mask has effects that persist at the masked location. This retinotopic masking suggests a local interference between the mask and the target in retinotopic coordinates. Integration masking has been shown to be retinotopic (Breitmeyer, 1984; Schiller, 1966) so moving attention would pick up a degraded target representation, even though the mask itself was never attended (or at least relatively unattended, given the inevitable imprecision of attentional selection, Hogendoorn et al., 2007).

For [Experiment 2](#) with color and motion, the single-sample control showed that performance improved when the moving ring directed attention to only the target, avoiding the competing features. The competing features still preceded and followed the target in retinotopic coordinates so the result suggests that the masking from the preceding and following stimuli is specific to the attended stream. This is evidence that interruption masking (Breitmeyer, 1984) occurs in the attended processing stream, and not in retinotopic coordinates.

### **Some object properties accumulate non-retinotopically, some do not**

The evidence for non-retinotopic accumulation came from the control experiments with attention moving across 1 to 5 locations. If the advantage of moving attention were solely to remove the interruption of processing due to the alternating features in the attended stream, performance should have been high when the moving guide passed over a single stimulus. In this case, a single feature pair is sampled at the only filled location as the guide passes over it. There are no preceding or following features that are either the same (as in the standard moving guide case) or the opposite (as in the stationary attention case). If the

interruption from the competing features were all that prevented accurate responses, performance should be high when sampling once from a single location. This was not the case, however, as the single-sample control showed (Figure 7). At the fixed rate chosen for the multiple-sample control (Figure 8), responses were at chance when only a single location was sampled but rose to very high levels once 3 or 4 locations were sampled (Figure 8). This is evidence for subthreshold summation, demonstrating the continuation of processing based on additional samples across space prior to a decision stage that would impose a threshold.

It is also evidence against probability summation. If the improvement with additional samples were due to the statistical advantage of multiple decisions, the accuracy with a single sample would have to be above chance and ought to show an initial negative acceleration in performance with additional samples. This was not true for the moving attention accumulation functions in Figure 8 and, to a lesser extent, in Figure 10.

Our results strongly support cross-location accumulation of color and motion information, but not of letter or digit information (Experiment 4). When letter and digit patterns alternated with noise masks, moving attention did not improve performance, indicating that the information acquired at each location could not be combined with information from subsequent locations. We speculate that separate samples cannot be easily placed in a common spatial register necessary for accumulation across the large distances we have used in our displays.

This result raises the question of how information is integrated from different locations within a large receptive field. The simplest possibility is that features are summed irrespective of location: a bit of red anywhere within the receptive field would contribute to the overall response of a cell tuned to red. This analysis is nonretinotopic in the sense that the local patches of red do not have to align in retinotopic coordinates for their signal strength to sum. Recognition of surface features like color or motion is a minimal type of “holistic” processing: “redness”, for example, has no part structure so any red anywhere can contribute to the perception of red. Other more complex stimuli like faces that may also be recognized holistically should be able to benefit from this position-independent accumulation of evidence. Each presentation of a particular face at a different location, even if below recognition threshold, may contribute to identification as long as each presentation falls within the receptive field of the same cell (tuned to that face) and thus adds to that cell’s activation. Our results with the letter and digit shapes do not show summation across locations, however, so we suggest that letter and digit recognition is not primarily holistic but more part based. In this case, if the stimulus moves during analysis, the processing of the parts can only continue based on coordinates centered on the object. A feature on the left of the object when it first appears needs to be coded, not

in retinotopic coordinates, but as on the left of the object. Only then can it be appropriately combined with other parts picked up when the letter or digit is then presented at a different location. This object-based coordinate scheme would place many requirements on the computation underlying the cell’s processing and we have no evidence for such analysis in our experiments here. In contrast we find evidence for accumulation only for features that do not require any object-based alignment over locations.

Since we have studied only two types of stimuli and in very different procedures, we can only say that cross-location integration occurs for some stimuli and not for others. Physiological studies have not approached this question yet either. Experiments with multiple objects in the receptive field typically look at the interference between these stimuli and modulate the degree of interference by making one or the other task relevant, manipulating attention (Reynolds & Chelazzi, 2004). Sequential presentation at different locations where the successive presentations are taken to be continuations of the same object have not yet been tested nor has the simpler case of moving higher order stimuli (faces, objects, etc.).

Note that in our displays, the summation that we did find could not be categorized as simply non-localized summation within larger receptive fields because it requires, as well, strong attentional selection. Specifically, if we imagine that the integration is occurring over large spatial extents, attention must be strongly gating inputs to the cell from different areas of the field as the target moves. Since it is the location of the moving target that determines the attended locations, this defines an object-based integration process even if the information within each sample is simply summed rather than being combined in register from sample to sample in object-centered coordinates.

## Dwell time and the spatiotemporal sampling window for attention

Dwell time for attention refers to the minimum temporal interval that can be attended independently of preceding and following stimulation (Duncan et al., 1994). The dwell time for selecting and combining the colors and motions is fairly long, 300 ms or more (Arnold, 2005; Moradi & Shimojo, 2004), as it is for other attention-dependent tasks (e.g. Duncan et al., 1994; Holcombe & Cavanagh, 2001). At a rate of 4 Hz or more in Experiment 1, both of the alternating colors, and both of the alternating motions fall within the dwell time (Figure 13) so that processing of the initial combination is not completed before the next pair arrives and interrupts.

Note that our description of interruption masking here equates it functionally to the attentional dwell time of Duncan et al. (1994). We are assuming that attentional

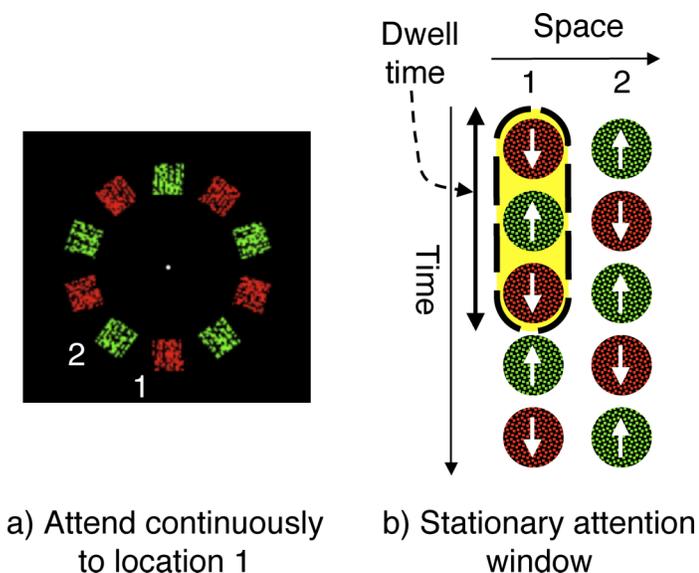


Figure 13. Attention sampling from a fixed location. A channel opens for transfer of information from location 1, but it has a minimum duration or dwell time before it can close again, typically no less than 200 ms (Duncan et al., 1994; Theeuwes, Godin, & Pratt, 2004). During that time, the alternating stimulus cycles between the two values and the resulting interruptions degrade the ability to identify and combine the colors and motions appropriately.

resources are required to identify and link the target’s color and motion features. If a second stimulus arrives within the attentional dwell time, the attentional demands of the second stimulus interrupt the feature analysis of the first. These two, interruption masking and attentional dwell time, are in turn alternate descriptions of the temporal resolution of attention (Verstraten, Cavanagh, & Labianca, 2000), the minimum temporal interval that allows two events to be individuated.

Figure 14 shows the alternating displays and the sampling window for attention as it moves across locations. Our accumulation results (Figures 8 and 10) indicate that when attention moves over the display, its  $\approx 300$  ms dwell time is spread over space, covering  $8^\circ$  to  $12^\circ$  of visual angle (3 to 4 locations) at an eccentricity of  $6.5^\circ$ . Depending on the speed with which it moves, the access by attention to any one location (local dwell time) may last as little as the 50 ms that we observed for the color stimuli of Experiment 4. The implication is that once attention arrives at a location and opens a channel to sample information from that position, it does not stay open once attention moves on.

The integration window of about 300 ms from the accumulation graphs matches reasonably well with the threshold durations for stationary attention conditions. The threshold alternation rates range from 2.2 Hz for color and motion together and for motion alone, to 3.5 Hz for color alone and 4.0 Hz for masked letters and digits. The

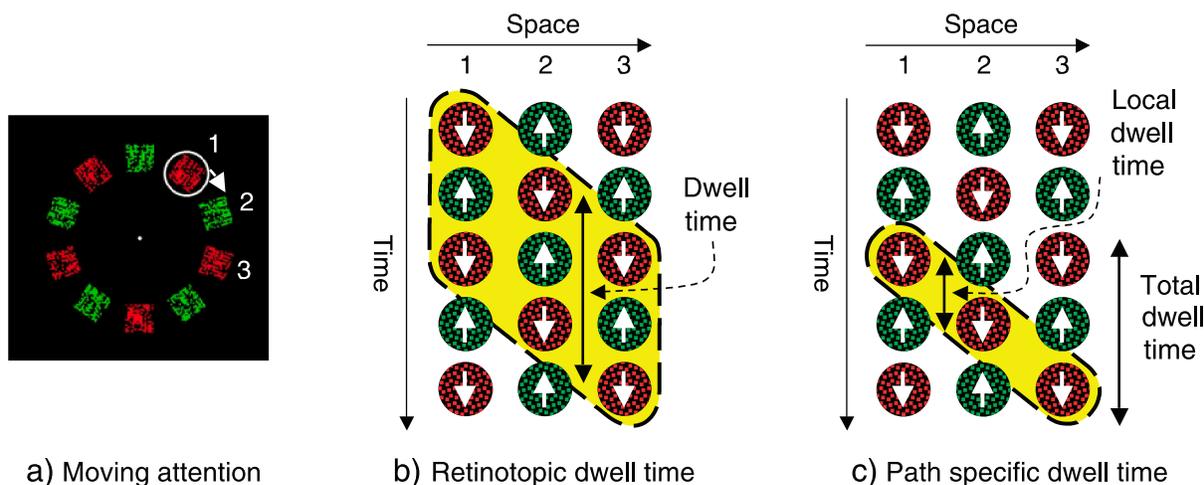


Figure 14. (a) Attention sampling from a moving location. (b) When attention is moving, it might open a channel at each location that must remain open for the minimum dwell time before closing again. With this retinotopic dwell time, alternating stimuli at each location would be integrated to the same extent with moving or stationary attention. (c) However, our data demonstrate that moving attention can offer substantial performance advantages, suggesting that the channel, once open, does not stay open in retinotopic coordinates, but stays open along the trajectory of motion. It may remain open only for a split second at each location (local dwell time as little as 50 ms in Experiment 3), allowing attention to sample very brief instants from a rapidly changing stream whose elements would be otherwise inaccessible to attention. Our data from accumulation experiments suggest that the total dwell time (the time during which an attention-dependent process is susceptible to interruption from a subsequent stimulus) is still long, about 300 ms, but the stimuli that can interrupt are only those falling on the attention trajectory.

combined duration of a target and its complement or its mask in these cases is therefore 450 ms at the slowest to 250 ms at the fastest. These differences may originate in signal strength and/or the nature of the interference between subsequent stimuli. The rates also correspond roughly to the range of temporal processing limitations found in fMRI studies for higher cortical areas (e.g., 4.5 Hz, V4d, 2 Hz, FFA, McKeeff, Remus, & Tong, 2007) and the differences in threshold rates might be linked to the temporal limitations of the particular cortical area required for integration. For the motion and color stimuli, the alternation with competing stimuli disrupts the labeling of both and the threshold rate ought to reflect the minimum attention sampling time (over which integration or interruption will occur): if more than one value is sampled, performance will begin to decrease. With the masked letters and digits, however, the mask is not a competing category. If the consequence of the persistent local effect of the mask is some mixture of a mask and a letter, the letter might still be recovered from the combination independently of the duration of the mask and letter alternation. The threshold timing for the masking experiment may reflect more directly the timing parameters of low-level masking rather than attentional dwell time.

### Spatiotopic integration versus object-based integration

Integration across successive positions of an object in motion is like integration across locations on the retina when the eyes move but the object does not. How do our results showing integration for some properties of a moving object but not others relate to results from integration across saccades?

Some have suggested that the visual system perfectly fuses pre- and post-saccadic images in world coordinates (e.g., Breitmeyer, 1984), while others claim that no visual information at all is preserved across a saccade and that only semantically coded information can be retained and combined (O'Regan & Levy-Schoen, 1983). In between these two extremes, Irwin (1996) has argued that some limited information is retained across a saccade, but that it is only weakly coded for spatial position. Hayhoe et al. (1991) claimed that the representation built up across saccades does have some visual qualities, similar in nature to a map that allows the geometrical shape of sequentially presented elements to be accumulated in world coordinates. However, Hayhoe et al. presented only three dots, one per fixation, each clearly visible, and asked observers to classify the angle at the top of the triangle formed by the three dots in world coordinates. This summation of simple patterns may be the limit for trans-saccadic integration. The only evidence for trans-saccadic subthreshold summation appears to be Melcher

and Morrone's (2003) report of spatiotopic motion integration. This is similar to our report here of subthreshold summation for motion across locations when tracking a single stimulus with attention. Our failure to find integration of letter and digit forms is also consistent with the difficulties seen for trans-saccadic integration of all but the simplest, suprathreshold patterns (De Graef & Verfaillie, 2002; Hayhoe et al., 1991; Irwin, Brown, & Sun, 1988; Rayner, 1998). In sum, trans-saccadic integration may have the same properties we find here for trans-location integration; specifically, it may occur for features that are homogeneous (not requiring exact spatial alignment across samples) and may be dependent on attention to link the successive samples.

The trans-location integration we found here was limited to attended objects but this was a requirement of our procedure so we cannot test whether attention is required for integration itself. Assuming that the integration is carried out by large receptive fields, then attention would only be required for target selection when distractors are present within the integrating receptive field. If there are no distractors within the receptive field mediating integration, there is nothing to rule out the useful integration of object properties in the absence of attention.

### Mobile computation

Our results from mobile computation experiments reveal that object features like motion and color can be accumulated on the fly. Nishida et al. (2007) and Watanabe and Nishida (2007) recently showed integration of color information with apparent motion at a fine scale, as opposed to our display where we demonstrated integration across gaps of 4° of visual angle. In an apparent motion study more closely related to ours, Shimozaki et al. (1999) reported object-based integration of luminance information over displacements of about 1.5°. These examples of accumulation over large distances suggest the involvement of cortical centers with large receptive fields like those involved in object-based analyses (e.g., Lateral Occipital Complex in humans, see Kanwisher, 2003, IT in monkeys, Ito et al., 1995). These imaging and single cell studies did not present the stimulus in motion within the receptive field to ask whether the stimulus preference is maintained while the stimulus is in motion, but this seems a plausible site for cross-location accumulation of analysis. Kahneman et al. (1992) have proposed temporary memory structures, "object files", to maintain a link between a moving object and its properties but we demonstrated not storage of the results of completed analyses but rather the accumulation of partial analyses across large regions of space. Cortical areas like V4 and MT, because of their restricted receptive field sizes, would not support mobile accumulation over

the large range that we find here (e.g., 8 to 12 degree span at 6.5 degrees eccentricity).

In conclusion, with static stimulus presentations, it is hard to dissociate local from nonretinotopic processing because the object remains available for local analysis. With moving objects, we can isolate nonretinotopic processes and identify which cortical areas are capable of supporting them. Using apparent motion, our tests revealed nonretinotopic accumulation of some object properties (color and motion) but not others (letter and digit identity). Our tests also demonstrated that integration masking occurs locally, in retinotopic coordinates, whereas interruption masking appears to be specific to the attended stream independently of its path in retinal coordinates.

## Acknowledgments

This research was supported by the National Institutes of Health Grant EY-09258 (PC), by a Chaire d'Excellence grant (PC), by a Discovery Project grant from the Australian Research Council (AOH and PC) and by the Graduate Students Study Abroad Program, National Science Council, Taiwan (WLC). We thank Pascal Mamassian and two anonymous reviewers for helpful comments.

Commercial relationships: none.

Corresponding author: Patrick Cavanagh.

Email: patrick.cavanagh@univ-paris5.fr.

Address: 45 rue des Saints Pères, Paris 75006, France.

## References

- Abbott, L. F. (2001). Where are the switches on this thing? In J. L. van Hemmen & T. J. Sejnowski (Eds.), *23 problems in systems neuroscience*. Oxford: Oxford University Press.
- Afraz, S. R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Research*, *48*, 42–54. [[PubMed](#)]
- Arnold, D. H. (2005). Perceptual pairing of colour and motion. *Vision Research*, *45*, 3015–3026. [[PubMed](#)]
- Avidan, G., Hasson, U., Malach, R., & Behrmann, M. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 2. Functional neuroimaging findings. *Journal of Cognitive Neuroscience*, *17*, 1150–1167. [[PubMed](#)]
- Bahill, A. T., & McDonald, J. D. (1983). Smooth pursuit eye movements in response to predictable target motions. *Vision Research*, *23*, 1573–1583. [[PubMed](#)]
- Bartels, A., & Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Research*, *46*, 2280–2286. [[PubMed](#)]
- Beaudot, W. H. (2002). Role of onset asynchrony in contour integration. *Vision Research*, *42*, 1–9. [[PubMed](#)]
- Bex, P. J., Dakin, S. C., & Simmers, A. J. (2003). The shape and size of crowding for moving targets. *Vision Research*, *43*, 2895–2904. [[PubMed](#)]
- Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, *408*, 196–199. [[PubMed](#)]
- Bodelón, C., Fallah, M., & Reynolds, J. H. (2007). Temporal resolution for the perception of features and conjunctions. *Journal of Neuroscience*, *27*, 725–730. [[PubMed](#)] [[Article](#)]
- Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.
- Brown, B. (1972). Resolution thresholds for moving targets at the fovea and in the peripheral retina. *Vision Research*, *12*, 293–304. [[PubMed](#)]
- Burr, D. C. (1981). Temporal summation of moving images by the human visual system. *Proceedings of the Royal Society of London B: Biological Sciences*, *211*, 321–339. [[PubMed](#)]
- Chung, S. T., & Bedell, H. E. (2003). Velocity dependence of Vernier and letter acuity for band-pass filtered moving stimuli. *Vision Research*, *43*, 669–682. [[PubMed](#)]
- Comtois, R. (2003). *VisionShell PPC* [Software libraries]. Cambridge, MA: Author.
- Dakin, S. C., & Bex, P. J. (2002). Role of synchrony in contour binding: Some transient doubts sustained. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *19*, 678–686. [[PubMed](#)]
- d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience*, *10*, 249–55. [[PubMed](#)]
- De Graef, P., & Verfaillie, K. (2002). Transsaccadic memory for visual object detail. *Progress in Brain Research*, *140*, 181–196. [[PubMed](#)]
- DiCarlo, J. J., & Maunsell, J. H. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology*, *89*, 3264–3278. [[PubMed](#)] [[Article](#)]
- Duffy, C. J., & Wurtz, R. H. (1991a). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, *65*, 1329–1345. [[PubMed](#)]

- Duffy, C. J., & Wurtz, R. H. (1991b). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology*, *65*, 1346–1359. [[PubMed](#)]
- Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature*, *369*, 313–315. [[PubMed](#)]
- Gattass, R., & Gross, C. G. (1981). Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*, 621–638. [[PubMed](#)]
- Gattass, R., Sousa, A. P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience*, *8*, 1831–1845. [[PubMed](#)] [[Article](#)]
- Hayhoe, M., Lachter, J., & Feldman, J. (1991). Integration of form across saccadic eye movements. *Perception*, *20*, 393–402. [[PubMed](#)]
- Hogendoorn, H., Carlson, T. A., & Verstraten, F. A. (2007). The time course of attentive tracking. *Journal of Vision*, *7*(14):2, 1–10, <http://journalofvision.org/7/14/2/>, doi:10.1167/7.14.2. [[PubMed](#)] [[Article](#)]
- Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, *4*, 127–128. [[PubMed](#)]
- Holcombe, A. O., & Cavanagh, P. (2008). Independent, synchronous access to color and motion features. *Cognition*, *107*, 552–580. [[PubMed](#)]
- Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science*, *5*, 94–100.
- Irwin, D. E., Brown, J. S., & Sun, J. S. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology: General*, *117*, 276–287. [[PubMed](#)]
- Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, *73*, 218–226. [[PubMed](#)]
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219. [[PubMed](#)]
- Kanwisher, N. (2003). The ventral visual object pathway in humans: Evidence from fMRI. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 1179–1189). Cambridge, MA: MIT Press.
- Kim, D. J., & Tong, F. (2005). Human ventral temporal areas contain flexible position-invariant information about subordinate-level objects [[Abstract](#)]. *Journal of Vision*, *5*(8):855, 855a, <http://journalofvision.org/5/8/855/>, doi:10.1167/5.8.855.
- Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *Journal of Neuroscience*, *12*, 408–424. [[PubMed](#)] [[Article](#)]
- Larsson, J., & Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *Journal of Neuroscience*, *26*, 13128–13142. [[PubMed](#)] [[Article](#)]
- Levi, D. M. (1996). Pattern perception at high velocities. *Current Biology*, *6*, 1020–1024. [[PubMed](#)]
- Lleras, A., & Moore, C. M. (2003). When the target becomes the mask: Using apparent motion to isolate the object-level component of object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 106–120. [[PubMed](#)]
- McKeeff, T. J., Remus, D. A., & Tong, F. (2007). Temporal limitations in object processing across the human ventral visual pathway. *Journal of Neurophysiology*, *98*, 382–393. [[PubMed](#)] [[Article](#)]
- Melcher, D., Crespi, S., Bruno, A., & Morrone, M. C. (2004). The role of attention in central and peripheral motion integration. *Vision Research*, *44*, 1367–1374. [[PubMed](#)]
- Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience*, *6*, 877–881. [[PubMed](#)]
- Moradi, F., & Shimojo, S. (2004). Perceptual-binding and persistent surface segregation. *Vision Research*, *44*, 2885–2899. [[PubMed](#)]
- Morgan, M. J., & Watt, R. J. (1982). Effect of motion sweep duration and number of stations upon interpolation in discontinuous motion. *Vision Research*, *22*, 1277–1284. [[PubMed](#)]
- Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, *70*, 909–919. [[PubMed](#)]
- Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society B: Biological Sciences*, *264*, 393–399. [[PubMed](#)] [[Article](#)]
- Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*, 1631–1647. [[PubMed](#)]
- Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Current Biology*, *14*, 830–839. [[PubMed](#)]
- Nishida, S., Watanabe, J., Kuriki, I., & Tokimoto, T. (2007). Human visual system integrates color signals along a motion trajectory. *Current Biology*, *17*, 366–372. [[PubMed](#)]

- Oğmen, H., Otto, T. U., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research*, *46*, 3234–3242. [PubMed]
- O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, *23*, 765–768. [PubMed]
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. [PubMed]
- Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, *27*, 611–647. [PubMed]
- Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience*, *23*, 339–348. [PubMed] [Article]
- Scharnowski, F., Hermens, F., Kammer, T., Oğmen, H., & Herzog, M. H. (2007). Feature fusion reveals slow and fast visual memories. *Journal of Cognitive Neuroscience*, *19*, 632–641. [PubMed]
- Scheerer, E. (1973). Integration, interruption and processing rate in visual backward masking. I. Review. *Psychologische Forschung*, *36*, 71–93. [PubMed]
- Schiller, P. H. (1966). Forward and backward masking as a function of relative overlap and intensity of test and masking stimuli. *Perception & Psychophysics*, *1*, 161–164.
- Shimozaki, S. S., Eckstein, M., & Thomas, J. P. (1999). The maintenance of apparent luminance of an object. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1433–1453. [PubMed]
- Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, *11*, 1182–1190. [PubMed] [Article]
- Smith, P. L., & Wolfgang, B. J. (2004). The attentional dynamics of masked detection. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 119–136. [PubMed]
- Spencer, T. J., & Shuntich, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology*, *85*, 198–203. [PubMed]
- Tanaka, K., & Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, *62*, 626–641. [PubMed]
- Theeuwes, J., Godijn, R., & Pratt, J. (2004). A new estimation of the duration of attentional dwell time. *Psychonomic Bulletin & Review*, *11*, 60–64. [PubMed]
- Tong, F., & Kim, D. J. (2005). Transformation from position-specific to position-invariant coding of objects across the human visual pathway [Abstract]. *Journal of Vision*, *5*(8):91, 91a, <http://journalofvision.org/5/8/91/>, doi:10.1167/5.8.91.
- Treue, S., & Maunsell, J. H. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *Journal of Neuroscience*, *19*, 7591–7602. [PubMed] [Article]
- Verstraten, F. A., Cavanagh, P., & Labianca, A. T. (2000). Limits of attentive tracking reveal temporal properties of attention. *Vision Research*, *40*, 3651–3664. [PubMed]
- Watanabe, J., & Nishida, S. (2007). Veridical perception of moving colors by trajectory integration of input signals. *Journal of Vision*, *7*(11):3, 1–16, <http://journalofvision.org/7/11/3/>, doi:10.1167/7.11.3. [PubMed] [Article]
- Yin, C., Shimojo, S., Moore, C., & Engel, S. A. (2002). Dynamic shape integration in extrastriate cortex. *Current Biology*, *12*, 1379–1385. [PubMed]