

*Chapter 10*

## TRANSSACCADIC MEMORY: BUILDING A STABLE WORLD FROM GLANCE TO GLANCE

DAVID MELCHER

*Oxford Brookes University, UK, and University of Trento, Italy*

CONCETTA MORRONE

*San Raffaele University and Institute of Neuroscience of the National Research Council, Italy*

*Eye Movements: A Window on Mind and Brain*

*Edited by R. P. G. van Gompel, M. H. Fischer, W. S. Murray and R. L. Hill*

*Copyright © 2007 by Elsevier Ltd. All rights reserved.*

**Abstract**

During natural viewing, the eye samples the visual environment using a series of jerking, saccadic eye movements, separated by periods of fixation. This raises the fundamental question of how information from separate fixations is integrated into a single, coherent percept. We discuss two mechanisms that may be involved in generating our stable and continuous perception of the world. First, information about attended objects may be integrated across separate glances. To evaluate this possibility, we present and discuss data showing the transsaccadic temporal integration of motion and form. We also discuss the potential role of the re-mapping of receptive fields around the time of saccades in transsaccadic integration and in the phenomenon of saccadic mislocalization. Second, information about multiple objects in a natural scene is built up across separate glances into a coherent representation of the environment. Experiments with naturalistic stimuli show that scene memory builds up across separate glances in working memory. The combination of saccadic re-mapping, occurring on a timescale of milliseconds, and a medium-term scene memory, operating over a span of several minutes, may underlie the subjective impression of a stable visual world.

## 1. Introduction

While the input to the visual system is a series of short fixations interleaved with rapid and jerky shifts of the eye, the conscious percept is smooth and continuous. The eyes move but the world is normally perceived as stable (unless, for example, one passively moves the eyeball with one's fingers). This sharp contrast between the input from the eye and the naïve visual experience has long puzzled philosophers and, more recently, psychologists and neuroscientists. One of the first to notice this problem was the Persian scholar Alhazen (965–1039 AD): “For if the eye moves in front of visible objects while they are being contemplated, the form of every one of the objects facing the eye . . . will move on the eyes as the latter moves. But sight has become accustomed to the motion of the objects' forms on its surface when the objects are stationary, and therefore does not judge the objects to be in motion” (Alhazen, 1989).

There have been several attempts to solve the problem of visual stability. One suggested mechanism is a transsaccadic memory buffer to support the fusion of separate images (Irwin, Yantis, & Jonides, 1983; Jonides, Irwin, & Yantis, 1982). The basic assumption, which followed the computational tradition of Marr, was that the goal of the visual system was to build a metric internal representation of the world. Perception was defined as the creation of three-dimensional volumes that could be matched to similar representations in memory. It follows from this assumption that our subjective perception of a continuous world, which is perceived to extend beyond the limits of the fovea, might depend on the storage of a virtual three-dimensional space in the brain that is eye-invariant. This proposal has been attacked on various fronts, including the philosophical argument that visual perception does not involve viewing the world in an internal “Cartesian theater” (Dennett, & Kinsbourne, 1992). A critical challenge for an internal, metric representation that integrates shape information across saccades is the fact that objects move and rotate with respect to the viewer. It is impossible, for example, to represent in metric detail multiple viewpoints of a face as it turns towards or away from the viewer, unless one is viewing a painting by Picasso. Objects in an internal theater would have to be continually updated in terms of position and viewpoint, which would be both computationally expensive and practically impossible given the limited acuity of peripheral vision.

The extreme opposite view to the Cartesian theater is the proposal that perceptual stability depends, paradoxically, on the lack of internal representation of the world (O'Regan & Noë, 2001). Proponents of “active vision” have suggested that detailed visual information can be gleaned by making an eye movement “just in time”, when it is needed, rather than storing large amounts of unnecessary information (Ballard, Hayhoe, & Pelz, 1995; Findlay & Gilchrist, 2003). This can be seen as an extension of the “ecological approach” to vision, which focuses on the visual information readily available from the world rather than on internal representations. In practice, however, it is still necessary for the brain to know where to look for the information that it needs, since eye movements are not random and are rarely wasted in natural tasks (Land & Hayhoe, 2001; Land, Mennie, & Rusted, 1999). We do not need to see the whole world in clear detail in any particular glance, but we need to be able (have the potential) to see any particular area of the

world in clear detail in the next glance. That “potential” may be a critical part of naïve perception, but it requires knowledge. Thus, details about the layout of the scene and the position of important objects must be represented, raising the spectre of the internal theater yet again.

The basic hypothesis underlying the experiments reported in this chapter was that in order to yield a subjective impression of a stable and detailed scene, without actually building a detailed representation in the mind, there must be at least two different mechanisms of scene memory operating at different timescales. The first timescale is that of a single saccade, in which local, photographic detail is replaced by new visual input (Irwin et al., 1983; McConkie & Zola, 1979; Tatler, 2001), while more invariant information survives and combines with relevant input from the new fixation (Loftus, 1972; Melcher, 2001; Tatler, Gilchrist, & Rusted, 2003). The second time frame is over a period of minutes or seconds, in which the observer interacts with the immediate visual environment. In order to perceive a stable, immediately available world, we have suggested that it is important to be able to learn about the identity and location of previously fixated objects without resorting to a time-consuming search. A number of studies that have shown that information about the identity and location of objects in a scene persists (Germys, De Graef, & Verfaillie, 2002; Henderson & Hollingworth, 2003; Pollatsek, Rayner, & Collins, 1984) and accumulates (Loftus, 1972; Melcher, 2001, 2006; Melcher & Kowler, 2001; Tatler et al., 2003) across multiple glances. We have suggested that this information is available in a “medium-term” memory store that involves the long-term memory system but does not necessarily require consolidation into permanent memory if the object and/or environment is not sufficiently salient or fails to be repeatedly viewed (Melcher, 2001, 2006).

## **2. Combining basic visual information across saccades**

We have investigated the mechanisms involved in learning about objects in a scene across a saccade. It is clear that integration of two separate views cannot be anything like overlaying two photographs. We thought that a likely place to start looking for memory that survives saccades is at the next step of visual processing, in which consistent and predictive information about an object is extracted over time.

### ***2.1. Transsaccadic integration of motion***

To look for evidence of transsaccadic integration, we chose to study visual motion since its long integration times can exceed typical fixation durations (Burr & Santoro, 2001; Neri, Morrone, & Burr, 1998). The output of individual motion detectors is combined over both space and time, allowing even a weak signal to be seen given a sufficiently long view. The visual system is extremely sensitive to regularities and coincidences hidden in noise, making even a few dots moving coherently in a noise “snowstorm” detectable after several hundred milliseconds of viewing. In the case of complex motion patterns and

biological motion, the time period of integration can exceed one second (Burr & Santoro, 2001; Neri et al., 1998). It is important to note that the improvement in motion sensitivity as a function of stimulus duration cannot be explained by information summation, but involves true integration of motion information over time (Burr & Santoro, 2001).

At the same time, it has been argued that the visual “buffer” is “refreshed” with each new fixation (McConkie & Zola, 1979; Tatler, 2001), implying that long motion integration times must either continue across saccades or must be a laboratory phenomenon limited to artificially stable viewing conditions. We tested temporal integration of simple translational motion to see if it would continue even when a saccadic eye movement was made during the stimulus presentation (Melcher & Morrone, 2003). The stimulus was a cluster of black and white random noise dots (6° diameter) presented on a gray background. During each trial there were two periods of coherent horizontal motion (150 ms each) embedded in 10 s of random noise (dots re-plotted randomly on every frame), allowing for the presence of a saccadic eye movement in between the two motion signals (Figure 1, top). Motion coherence sensitivity was measured using a staircase procedure (for details, see Melcher & Morrone, 2003).

The delay between the two motion signals was varied across blocks of trials. At brief delays between the two motion signals, coherence sensitivity was not affected, showing full temporal integration. For larger delays, sensitivity decreased as a function of delay duration, up to the point at which the delay reached the temporal limit of integration. Beyond a certain delay duration, performance was equivalent to that found with a single period of motion signal, as expected from previous studies showing temporal limits on motion integration (Burr & Santoro, 2001).

On some blocks of trials, an eye movement (12°) was made from above to below the stimulus during the interval between the motion, while in the other blocks of trials the observer maintained fixation either above or below the stimulus. Temporal integration of coherent motion signals was not reduced by saccades. Rather, the two motion signals were combined across the saccade (Figure 1, bottom). It is important to note that the motion signals that were integrated across saccades were often, by themselves, below the level of conscious discrimination threshold. Thus, the methodology used excluded cognitive strategies or verbal recoding since the observer could not detect each motion signal – only by combining the two subthreshold signals could motion be correctly discriminated.

Interestingly, integration durations were actually higher for trials with saccades compared to maintained fixation. A comparable effect has been shown for the McCollough aftereffect, in which trials with eye movements produced longer aftereffect durations than those with fixation (Ross & Ma-Wyatt, 2004). These findings suggest that transsaccadic integration may involve a specific memory mechanism that is not active (or less active) in the case of maintained fixation. One possible explanation for this phenomenon is described in a later section (5.3).

Why might information be integrated across separate glances? Ironically, the ability to perceive relevant change (in location) may depend on the capacity to ignore irrelevant change (noise) while keeping track of what is consistent and unchanging (coherent motion

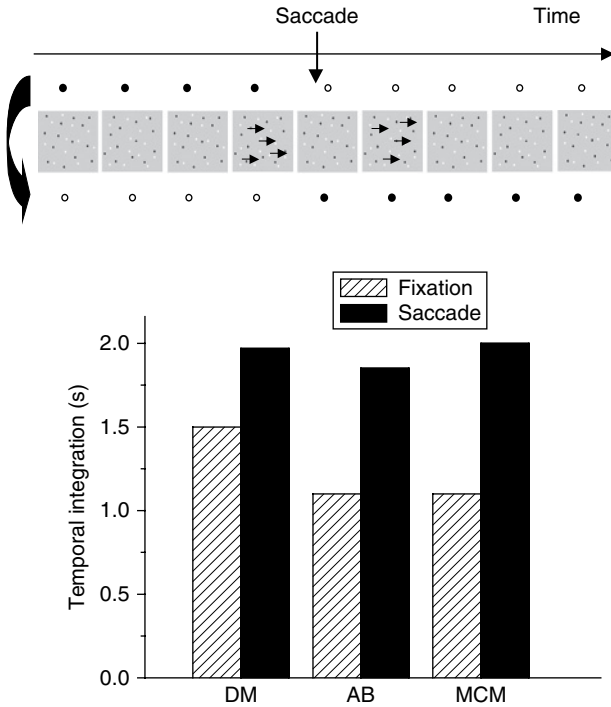


Figure 1. Transsaccadic integration of motion. *Top*: Schematic illustration of the stimuli comprising two short intervals of coherent translation motion embedded in 2 seconds of random noise. The subjects performed motion discrimination task while fixating the top circles or making a saccade from top to bottom circles in the interval between the two coherent motion. *Bottom*: Estimates of temporal integration for motion direction discrimination obtained by measuring coherence thresholds as function of the temporal separation between the two coherent motion signals (see Melcher & Morrone, 2003) for three subjects (DM, AB and MCM).

in one direction). In a complex and ever-changing environment, it is critical that the brain filters out the noise and identifies constancies and patterns in the perceptual input.

### 2.2. Transsaccadic integration of form

The temporal integration of motion found across saccades might be unique to motion processing or, conversely, might reflect a more widespread mechanism in visual perception. Since different features are processed in separate areas of cerebral cortex, the motion stimulus used in the previous study may involve a different processing stream (the dorsal “action/where” system) from perception of form (the “what” system) (Goodale & Milner, 1992; Mishkin, Ungerleider & Macko, 1983). Thus, it would premature to make any generalization about object processing based on our findings with motion, which is usually associated with the “where/action” stream. It is also important to note that we can

usually recognize an object in a single glance, raising the question of whether temporal integration (as found with motion) would be necessary. There are, however, exceptions in which form identification changes over time including slit-viewing (Anstis & Atkinson, 1967; Moore, Findlay & Watt, 1982; Nishida, 2004; Yin, Shimojo, Moore & Engel, 2002), form-from-motion (Domini, Vuong, & Caudek, 2002; Ullman, 1983), object priming (Germys, De Graef, & Verfaillie, 2002; Pollatsek, Rayner, & Collins, 1984), the interpretation of ambiguous figures (for an interesting review see Piccolino & Wade, 2006) and form adaptation (Blakemore & Campbell, 1969; Kohler & Wallach, 1944; Melcher, 2005; Purkinje, 1823).

We modified the stimulus used in the previous experiment in order to test form coherence (Braddick, O'Brien, Wattam-Bell, Atkinson, & Turner, 2000). The basic logic was to take a complex form (a radial or circular pattern) and to show a small portion of that pattern (embedded in random dots) in each single frame. Over a series of frames, the observer begins to perceive the global shape of the pattern. The observer's task was to indicate whether the pattern was radial or circular (Figure 2, top). Individual dots did not move coherently across separate frames. There were two short periods ("coherent form stimulus") containing the form pattern embedded in a longer period of dynamic random noise with no pattern. As in the first experiment, the stimulus was 6° in diameter and observers were instructed to either make a 12° saccade from the fixation point above the stimulus to the one below the stimulus (or vice-versa) or to maintain fixation during the entire trial (in separate blocks). The trial lasted for 4 s in total, with 100 dark or white dots plotted against a background of mean gray. A particular trial contained either one or two periods of coherent form. As in the first experiment, coherence sensitivity was measured using a staircase procedure for each condition (single or double signal) and delay duration (0–800 ms). The duration of the coherent form signal was 100 ms, compared to 150 ms for the first experiment (motion). The use of a shorter stimulus period was motivated by the need to keep performance well below ceiling.

As expected, form coherence sensitivity decreased as the two motion signals were separated in time beyond the critical limit (Figure 2, bottom). The pattern of form integration was similar for trials with saccades (triangles) and trials with maintained fixation (squares), demonstrating transsaccadic integration of form. Overall, the temporal integration time (around 700 ms) was about half of that found for motion for the same observer (more than 1500 ms). Yet the finding of transsaccadic integration remained the same for the two tasks, implying that transsaccadic integration also occurs for shape information.

Of course, the radial/circular stimulus did include a motion component since the dots were dynamically replotted on each frame, even if the pattern did not move in a consistent direction. Thus, it was important to also test transsaccadic perception without any motion cues whatsoever. Further evidence that form processing is not strictly retinotopic comes from studies of shape adaptation aftereffects (Melcher, 2005). In these experiments, four different adaptation aftereffects were tested: contrast, tilt, complex form and face adaptation. We found that the magnitude of the aftereffect was modulated by whether or not the adapter and test were in the same spatial location. While contrast adaptation

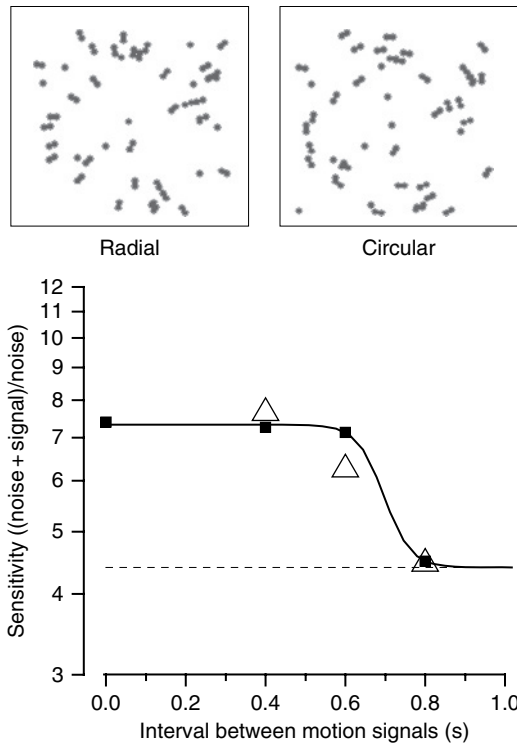


Figure 2. Temporal integration of form. *Top*: Example of the stimuli used for form discrimination. The stimuli sequence was the same as shown in Figure 1 (top), except that the coherent motion intervals were substituted with coherent form intervals. *Bottom*: Sensitivity, the inverse of the proportion of dots displayed along the radial or circular form to the total number of dots at coherence threshold, as function of the separation between the two coherent form interval measured during fixation (filled square) and during saccade (open triangles). The dotted line shows the sensitivity when only one coherent form interval is presented.

showed no transsaccadic aftereffects, the other conditions showed increasing aftereffect size as the stimulus increased in complexity. This suggests that shape processing becomes increasingly eye-independent along the visual processing pathway (see Melcher, 2005 for details).

### 3. Transsaccadic accumulation of memory for natural scenes

One simple way to understand what people learn across separate glances would be to show them a natural scene and allow them to freely scan the picture. Then, the observer would be stopped at a certain point and asked about what they had just viewed. This logic had worked well in an initial set of studies of scene memory (Melcher, 2001; Melcher and

Kowler, 2001). In those experiments, the task was to freely recall the names of objects in computer-generated rooms. Viewing time ranged from 250 ms to 4 s. The number of items recalled increased linearly as a function of viewing time.

One novel aspect of the recall experiments was that some displays were displayed more than once to test for memory savings across separate views. Surprisingly, the number of items recalled in these retest displays also increased at the same rate as a function of total viewing time. In other words, a display shown twice for 2 s each would lead to the same performance as a display shown once for 4 s. This memory accumulation occurred despite the fact that the retest displays were separated by an average of 4 other scenes. Previously, such temporal integration had only been found for brief views (400 ms) of scenes tested by a recognition test (Loftus, 1981). The finding of memory accumulation across repeated presentations of the same display suggested that visual working memory was capable of keeping in mind several scenes (each full of many objects), not just several objects.

A new set of experiments tested transsaccadic memory under more natural viewing conditions with photographs and pictures (Melcher, 2006). Instead of free recall, observers (23) were asked questions about the color, location or identity of specific objects in the display (Examples include the following: (1) What color is the tablecloth? peach, white or blue (2) Where is the teacup? bottom right, bottom left or center (3) What food is in the middle plate of the three-tiered plate stand? cake, sandwiches or sausage rolls?). Stimuli were shown for up to 20 s, in order to look for ceiling effects in memory performance. Critically, some images were not followed by a memory test after the first presentation but only after being shown a second time. This manipulation served to measure whether the memory accumulation across repeats of the same stimuli found previously for object recall extended also to questions about specific object attributes. Re-test trials were shown 4–6 trials after the initial display of that stimulus. Images included drawings of realistic scenes, reproductions of paintings and photographs of both indoor and outdoor natural scenes. The memory test contained a series of written questions on the screen about specific objects in the scene and a list of three choices (see Melcher, 2006 for more detail).

On each trial, the stimulus display was presented for a time period of 5, 10 or 20 s. The 5 s and 10 s trials were run in the same blocks (since they also contained re-test trials), while the 20 s trials (with no re-tests) were run in separate blocks. After each trial, participants were either given a memory test or instructed to continue on to the next trial. Participant responses were given by keyboard press and recorded for later analysis. The order of conditions was randomized across observers (see Melcher, 2006 for more detail).

Figure 3 shows the average percent correct performance for all types of questions as a function of total viewing time. The solid circles show performance after 5, 10 or 20 s on the non-repeat trials. Open squares show re-test trials, in which the memory test was given only after the second time that the picture had been shown. The leftmost square shows performance after seeing a display twice for 5 s each time on repeat trials, for a total viewing time of 10 s. The rightmost square shows percent correct response after seeing a 10-s and a 5-s display for a total viewing time of 15 s (on half of those trials,

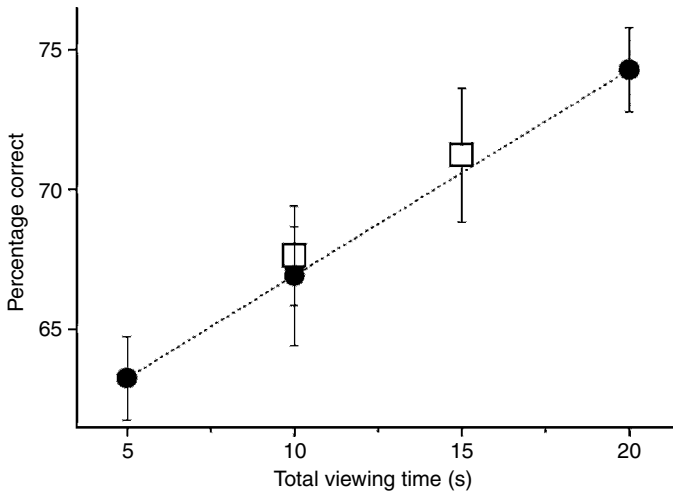


Figure 3. Performance in answering questions about the content of natural scenes as a function of the display duration. Filled circles show performance on continuous presentation trials, while open squares show percentage correct when the stimulus was shown twice before the test (10 s total viewing time = viewed twice for 5 s each time).

the 10-s duration trial preceded the 5-s trial, while the other half of trials contained the opposite pattern).

Overall, performance improved for longer views. Performance improved linearly as a function of total viewing time, with no difference between re-test trials and normal trials of the same total viewing time, consistent with our previous study with free recall. Similar results were found for each type of question (color, location or identity).

This study extends the results with object perception across single saccades and provides further evidence that information about the visual properties of objects persists across saccades (Germys et al., 2002; Henderson & Hollingworth, 2003; Loftus, 1972; Polatsek et al., 1984; Tatler et al., 2003). Moreover, these results show that many scenes, each containing multiple objects, are maintained in memory over a period of minutes. In other words, it is not only objects that persist across saccades, but the scene context as well.

#### 4. The cost of transsaccadic integration

In real life, we are typically surrounded by many objects, many of which are not being directly attended at a given point in time. The results of the first three experiments suggest the existence of an active, perhaps default, integration mechanism that combines information across saccades. Given the fact that the world does change over time, this suggests that we also need a dedicated mechanism to detect these changes in order to avoid incorrectly integrating information from different (or no longer visible) objects.

The visual system appears to depend on the detection of visual transients to notice changes in scenes (for review, see Kanai & Verstraten, 2004). Under laboratory conditions in which changes occur during saccades, or simulated saccades, observers perform poorly at detecting changes that occur without accompanying visual transients (Blackmore, Brelstaff, Nelson, & Troscianko, 1995; Grimes, 1996; Rensink, O'Regan, & Clark, 1995). Perhaps this failure reflects a natural tendency to integrate, rather than compare, images across saccades.

Consider the case of transsaccadic localization. When the saccadic target changes position during the saccade, observers often fail to report shifts in target position. This can be considered a form of change blindness. When the target is blanked during the saccade and reappears only after the saccade has landed, then observers regain their ability to notice the change in location (Deubel, Bridgeman, & Schneider, 1998). Surprisingly, it requires a blank delay of around 200 ms for observers to regain their best performance.

We hypothesized that for change detection a similar temporal delay might be required after the interruption in the display (saccade, blank screen or blink) in order to accurately detect the change on all trials. We tested the ability of observers to detect the number of changes (either one or two) that occurred to a display of 1–10 items. Items could change color (red, green, blue) or shape (circle, square, rectangle, ellipse) and all changes were easy to detect when they occurred without a blank intervening interval. There were five separate experiments, with different instructions. The first experiment was a simple change detection task with 1–10 items. The second experiment, with 3–6 items, required observers to discriminate the number of changes (one or two) that had occurred during the blank delay. If two changes occurred, they always happened to different items, so that the same item did not change twice. The remaining experiments all used four item displays. In the third experiment, a temporal delay was introduced between the first and second change, such that the second change occurred without a blank screen preceding it. Thus, the second change was visually obvious, even if it was not always detected. Observers in the fourth and fifth experiment were instructed to attend either to the first change only (immediately after the blank delay) or to wait and report the presence or absence of a change to an item that occurred after the stimulus display had reappeared.

Normally, change detection is fairly good for small displays of less than about four items (Figure 4a). For discriminating number of changes, however, observers were surprisingly poor even with a display of only three items (Figure 4b). It is often assumed that change detection involves placing the pre-change display items in visual short-term memory (VSTM) for comparison with the post-change display. The VSTM limit of about 4 items is inferred from change detection performance, such that VSTM as a concept (and debates about its properties) has become largely defined by how people perform in change detection tasks (for example, Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). There are two possible ways to reconcile the failure to detect number of changes in a three-item display with previous studies of VSTM. The first option is to posit an additional step or mechanism involved in comparing post-stimulus to pre-stimulus displays when more than one change needs to be counted. The second, more problematic, possibility is that performance on change detection task is not based on comparing post-change items to

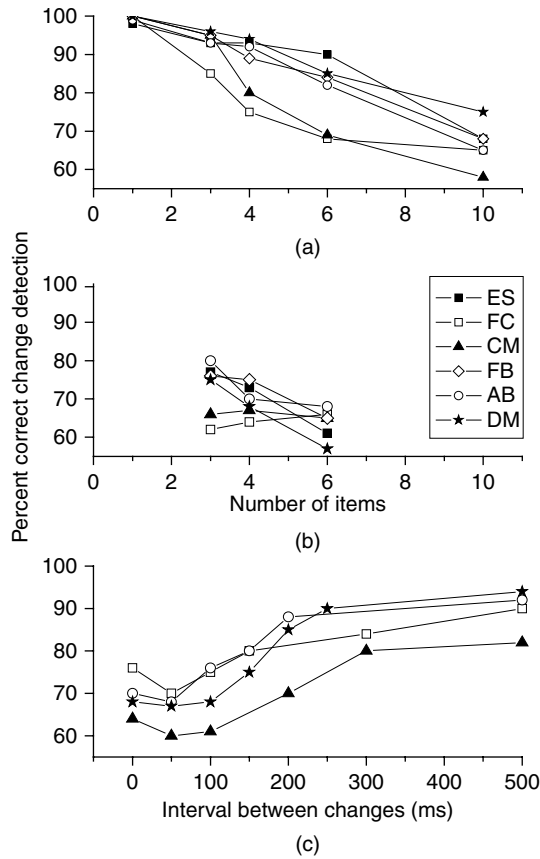


Figure 4. Performance in detecting a single change (a) or double change (b) to the objects presented in a complex display as function of the number of the item present in the display. Different symbols show performance for different subjects. (c): Performance in detecting number of changes as a function of the temporal separation between the two changes.

VSTM on an object-by-object basis at all. Change detection might involve comparing the entire pre-change display to the whole post-change display as a pattern rather than as a collection of independent objects. In the latter case, change detection would not be a valid methodology to determine the number of items stored in VSTM but rather the complexity of information stored in VSTM.

It might have been expected, based on the principle of chunking, that change detection would be easier when the two changing items were near to each other. On the contrary, the average distance between the two changing items was similar for correct and incorrect trials (AB: 5.3° vs 5.3°; CM: 5.7° vs 5.8°; DM: 6.3° vs 6.2°; FB: 5.3° vs 5.8°). Thus, the ability to discriminate number of changes did not depend only on the spatial distance between the two changing targets.

As expected, performance improved as a function of the duration of the temporal delay added between the first and the second change. Figure 4C shows that the time to reach optimal performance, equal to that found for detection of one vs no changes, ranged from 150 to 300 ms for a four-item display. When the two changes were separated in time, subjects were better able to detect the second change. Again, it is important to note that the visual transient created by the second change was not masked by a blank screen, blink or saccade, making the inability to correctly detect both changes surprising.

The remaining experiments tested the influence of instruction on the ability to detect one of the two changes. On four-item trials in which subjects were told to attend only to the second change, detection for that second change was nearly perfect, even when the second change was delayed by only 50 ms (ranging from 90 to 100% correct), demonstrating that low-level visual masking was not responsible for these effects. Thus, it was possible to ignore the change that occurred across the blank delay and focus only on whether or not a change occurred in the post-blank display. It was more difficult, however, to attend only to the first change and ignore the second. When subjects were instructed to ignore a second change following 50 ms later, detection for the first change was poorer than when there was no second change at all (AB: 80% vs 90%; DM: 76% vs 94%; FB: 76% vs 89%). This suggests that the second change, even when ignored, influenced the ability to consciously report the existence of the first change.

The current findings complement other studies showing failures of awareness when attention must be quickly shifted between two or more different items, time periods or locations, such as the attentional blink (Raymond, Shapiro, & Arnell, 1992). Our results are in agreement with the idea that change detection requires focused attention to only one object or cluster of items at a time (Rensink, 2000). When the results of this experiment are viewed in the context of transsaccadic integration, it suggests that the change detection task may be difficult because it interferes with the default strategy of memory accumulation. Change detection involves an attempt to compare two representations rather than combine them (Becker, Pashler, & Anstis, 2000). In the real world, such as walking through a crowded city sidewalk, the brain is faced with a myriad of changes across fixations. Rather than concentrating on the continual flux, the visual system searches for constancies and patterns. Thus, “blindness” to changes in parts of the scene that are not currently relevant, along with the build-up of information about aspects of the scene that are consistent or important to current goals, may be the natural strategy for the visual system in everyday life.

## **5. Discussion**

### ***5.1. Principles of transsaccadic memory***

Human memory is often described in terms of limits: a limited capacity in terms of items or a limit in duration. In the case of transsaccadic memory, it is certainly true that not everything is remembered (such as in the case of change blindness), and it is equally true

that not everything is forgotten (visual details are retained in memory). The question, then, is why some information is forgotten while other information is remembered.

The experiments reported here, for both a single saccade and across minutes, emphasize the balanced cooperation of learning and forgetting to separate the wheat from the chaff. An efficient system should retain information that is predictive and invariant across views but not remember transient details (Melcher, 2005). Trans-saccadic memory is not just about putting more and more information into storage; it is equally critical to forget information from previous fixations when that information is no longer relevant. We believe that this principle holds for both of the time periods studied here, a single saccade or multiple saccades within the same scene. In the case of motion and form integration, the information that is integrated over time and combined across the saccade is not a particular visual detail or set of dots, but a trend in the overall pattern of stimulation. The results described in this chapter emphasize the way in which what is seen in each new fixation depends on what has been seen in previous fixations – perception does not start over from scratch in each glance. The studies of motion integration (Melcher & Morrone, 2003), form aftereffects (Melcher, 2005) and scene memory (Melcher, 2001, 2006) all suggest that transsaccadic memory is not merely a static storage of abstract information, but rather an active factor that modifies processing of subsequent visual information.

For complete scenes with multiple objects, we have argued for a central role of forgetting as part of an efficient memory (Melcher, 2001; Melcher, 2006; Melcher & Kowler, 2001). The traditional “cognitive psychology” model is to view memory acquisition as a one-way process of moving information through bottlenecks from sensory to short-term to long-term memory. We have argued, on the contrary, that it would be grossly inefficient to have a single short-term memory/long-term memory (STM/LTM) distinction, since it would mean that every attended item, after a few additional fixations, would have to either be forgotten entirely or transferred permanently into LTM. Instead, we have argued, there is a “medium-term” period of proto-LTM that involves keeping in mind information about a particular environment for a particular task over a period of minutes, without all of that information necessarily being consolidated into LTM (Melcher, 2001; Melcher, 2006; Melcher & Kowler, 2001). While the exact nature of visual memory remains a matter of debate (Germys et al., 2002; Hollingworth, 2004; Tatler, Baddeley, & Land, 2005), the medium-term memory theory can account for a number of aspects of memory that are difficult to explain with a dichotomous STM/LTM theory (Melcher, 2001, 2006; Pierrot-Deseilligny, Muri, Rivaud-Pechoux, Gaymard, & Ploner, 2001).

A second theoretical consideration is that the environment, and our interaction with it, is structured in certain regular ways (Gibson, 1979; O’Regan & Noë, 2001). Thus, we might expect our transsaccadic memory to take advantage of the inherent structure of the environment in the efficient encoding and retrieval of scene information. There is now considerable evidence that the “object” is an organizing principle for perception and cognition (Blaser, Pylyshyn, & Holcombe, 2000; Duncan, 1984; Melcher, Pappathomas, & Vidnyanszky, 2005). One possibility is that the “scene”, like the object, serves as a structure to organize information in memory (Biederman, 1972; Gould, 1976; Melcher, 2001; Tatler et al., 2003). Perhaps one reason that the gist and layout of a scene are

gleaned in a single glance (Potter, 1976) is that such information is used to organize the data from subsequent fixations. One piece of evidence for this hypothesis comes from the finding that it is more difficult to learn objects in a scene when the background context has previously been associated with a different set of objects (Melcher, 2001; Melcher & Kowler, 2001). Thus, an object and its background context may naturally and automatically linked, as has been suggested since the time of ancient Greece (Cicero, 1942/55 B.C.).

### ***5.2. Remapping of receptive fields across a single saccade***

Can the transsaccadic integration effects reported here be simulated using simple and basic operations? If so, are these operations biologically plausible? Based on the known receptive field properties of the neurons in the visual areas involved in motion and form processing, we hypothesize that integration across a single saccade may involve two mechanisms working together: (1) peri-saccadic changes in retinotopic visual receptive fields, and (2) neurons with eye-independent receptive fields.

Based on electrophysiological recordings, it is now well established that only in a few visual areas (probably only V1 and V2) do receptive fields (RF) fail to show peri-saccadic changes (Duhamel, Colby, & Goldberg, 1992). Many neurons start to respond to stimuli positioned at locations different from their direct retinal afference before the eye movement begins. These effects are particularly evident in lateral intraparietal area (LIP) (Duhamel et al., 1992), but also very common in V3 and V3A and V4 (Nakamura & Colby, 2002). There are several ways that receptive fields can change around the time of the saccade. For some of these neurons the size of the RF can shrink, such as cells in V4 (Tolias et al., 2001), or enlarge by extending in the opposite direction of the saccade (Kubischik, 2002). Retinotopic RF can undergo re-mapping before the saccade, becoming de facto craniotopic from the moment of the initial remapping up to the next intention to move the eye compensating for the retinal shift induced by the eye movement.

During the re-mapping of the RF, which can be intuitively visualized as a moving RF in cortical space, it is plausible to suggest that the interested neuronal network is still able to carry out the visual analysis that it normally performs during fixation (Figure 5). The analysis could be form, contrast, motion or location, depending on the specificity of the area. For example a V3A neuron that participates in the analysis of motion would continue to participate to the motion analysis, but would now respond to a signal at a different eccentricity of the retinal input that correspond to the same external spatial location. This means that for a limited time period, this neuron is able to integrate visual information about form, colour and motion independently of eye position. This integration would, presumably, be performed just as during normal fixation, without particular loss of acuity or changes in RF size.

The existence of changes in RF around the time of saccades cannot, by itself, account for our data because the remapping is fast and transient: a complete predictive shift of the RF take place in less than 100 ms for a range of saccadic size from 10 to 20°. (Kusunoki & Goldberg, 2003; Nakamura & Colby, 2002). Our results for form and

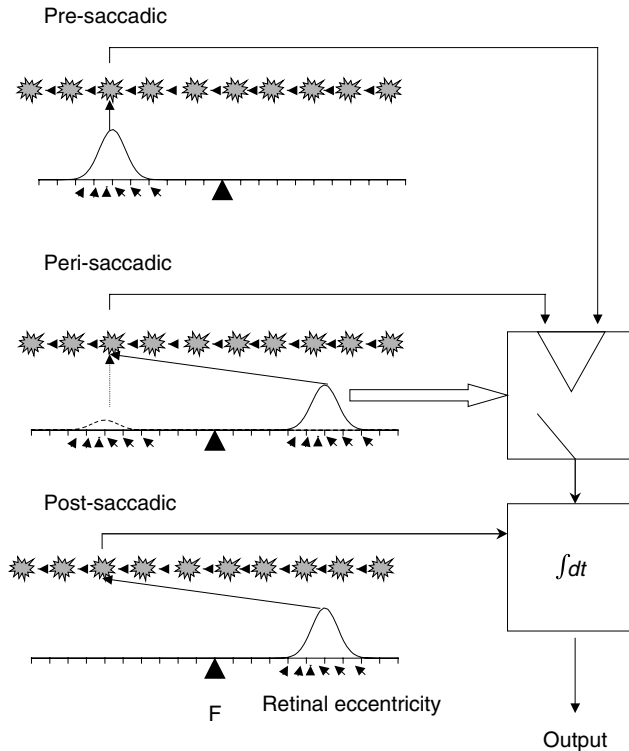


Figure 5. Schematic diagram of the proposed model of transsaccadic integration. Three separate instants are represented: top, time during fixation, middle the time around saccadic onset, bottom, the time period after the saccade when fixation is again stabilized. In each panel, the bottom graph represent the retinotopic activation map of an area of motion-sensitive cortex, with the big arrow indicating the foveal representation and the little arrows the motion selectivity of the neurons. The top layers represent an assembly of neurons that construct a receptive field anchored to the head position or to external space. Each individual neuron will necessarily select different retinotopic input from the lower layer depending on the gaze position (the input is shown by the arrows from the activity function to the selected cell). During fixation, the output of the same craniotopic neuron is integrated. Across saccades, the integration of the pre- and post-saccadic fixation is contingent (white arrow to the switch) on appropriate post-saccadic activity by the motion-selective neuron.

motion show integration between events that are separated from saccadic onset as far as 800 ms, a separation too large for the re-mapping process. To explain the transsaccadic integration we need a second stage of craniotopic RF (citations) that remap their retinal input in preparation for the saccade (see Figure 5).

To explain how this two-stage model might function in transsaccadic perception, it is worth examining motion integration as a specific test case. It has been proposed that LIP neurons support the long integration times observed psychophysically by temporally integrating the decision biases of ensembles of MT cells which are directionally selective

(Mazurek, Roitman, Ditterich, & Shadlen, 2003). In this view, the craniotopic/spatiotopic neuron acts like a non-specific memory buffer that accumulates information from the earlier retinotopic visual area. But which area is selecting the appropriate input to be integrated? How would it be possible that form is integrated with specific form information and not, for example, with irrelevant motion information? An important role for the selection could be performed by the highly selective, mainly retinotopic areas that undergo transient remapping. Those neurons know in advance of the eye movement what will be in the new receptive field, and thus would be in the position to know whether the information before and after the saccade should be integrated (and sent to the craniotopic RF).

Figure 5 illustrates the proposed mechanism, with three time periods and an integrator with an open/closed switch. In the case of motion, a specific direction-selective neuron in MT or V3A feeds its activity to a set of craniotopic neurons that would integrate this activity over time. If its activation during the remapping is constant, then the information will continue to be passed to the same craniotopic RF that will integrate the information between the two fixations. However, if the low-level neuron terminates its response, the switch will open and the craniotopic RF will no longer integrate the information. If the switch is open, then the craniotopic RF will accumulate new information starting afresh after the saccade, with the pre-saccadic information being lost. This would happen in situations in which the two images cannot be integrated, as has been suggested in studies of transsaccadic perception of ambiguous figures (Ross & Ma-Wyatt, 2004) and change detection across saccades (Becker, Pashler, & Anstis, 2000; Grimes, 1996).

Interestingly, the re-mapping of receptive fields has also been implicated in the mislocalization of object position during saccades (Ross, Morrone, Goldberg & Burr, 2001; Burr & Morrone, 2004). Thus, integration across saccades and mislocalization during the saccade might involve the same mechanisms. This simple idea could explain several how information is combined across separate glances without fusion of local, pixel-like visual detail.

### ***5.3. Physiological underpinnings of scene perception and memory***

In real-world scenes, memory involves information about the location and the identity of specific objects, in addition to gist and layout of the scene (Biederman, 1972; Loftus, 1972; Potter, 1976). What brain areas might underlie the type of scene memory studied here? A distinction is often made between object and location working memory (WM), with the former involving temporal regions and the latter parietal regions (Smith & Jonides, 1997). Natural scenes, however, provide an interesting case in which both object and location information are equally important. This suggests that both systems might work simultaneously to represent “what” and “where”, or, perhaps, they might work under different conditions or in differing time frames.

There is a body of evidence suggesting that the type of scene memory studied here may depend in particular on temporal regions. For example, remembering scene-based location, in addition to recognizing object identity, critically involves temporal regions (Goh et al., 2004; Lee et al., 2005; Rolls, Xiang, & Franco, 2005). Thus, memory for

object location within the scene has been correlated with both temporal and parietal regions, which may work together in scene representation. Another possibility, however, is that the parietal system may be limited to the earliest stages of attentional orienting to scene location rather than a more lasting memory of object location. Studies of visual-spatial WM have suggested that a transition occurs from parietal to temporal regions over time (for review, see Pierrot-Deseilligny et al., 2002). It is possible that the scene memory described in this chapter depends on this maintenance of information in temporal cortex that allows for further consolidation into LTM. Given that objects can change their retinotopic coordinates as a result of a saccade, it is important in future studies to examine whether peri-saccadic changes also occur in object- and scene-processing areas in temporal cortex and to what extent craniotopic and spatiotopic representations are used for complex scenes.

It is interesting to note that the frontal-parietal attentional network, rather than brain areas involved in learning about complex scenes, has been implicated in change detection and “change blindness” (Todd & Marois, 2004; Vogel & Machizawa, 2004). The dissociation between brain areas involved in change detection and in scene and object memory may help to explain the dissociations between change detection and memory performance (Henderson & Hollingworth, 2003). If this theorized dissociation is correct, then it suggests that caution should be taken before using VSTM tasks to make generalizations about the operation of memory in everyday life.

## 6. Conclusions

The experiments described in this chapter provide further evidence that visually detailed, yet not metrically exact, information is retained across saccades and used to influence perception and action. The model described here implies that the brain searches for consistent information across saccades and uses this to allow predictive and invariant information to build up over time. The combination of saccadic re-mapping, occurring on a timescale of milliseconds, and a more durable scene memory, operating over a span of several minutes, may underlie the subjective impression of a stable visual world.

## References

- Alhazen, I. (1083). Book of optics. In A. I. Sabra (Ed.) *The Optics of Ibn al-Haytham*. London: Warburg Institute, 1989.
- Alvarez, G. A. & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15, 106–111.
- Anstis, S., & Atkinson, J. (1967). Distortions in moving figures viewed through a stationary slit. *American Journal of Psychology*, 80, 572–785.
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7, 66–80.
- Becker, M. W., Pashler, H., & Anstis, S. M. (2000). The role of iconic memory in change detection tasks. *Perception*, 29, 273–286.

- Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177, 77–80.
- Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is the richness of our visual world an illusion – Transsaccadic memory for complex scenes. *Perception*, 24, 1075–1081.
- Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, 408, 196–199.
- Braddick, O. J., O'Brien, J. M., Wattam-Bell, J., Atkinson, J., & Turner, R. (2000). Form and motion coherence activate independent, but not dorsal/ventral segregated, networks in the human brain. *Current Biology*, 10, 731–734.
- Burr, D., & Morrone, M. C. (2004). Visual perception during saccades In J. Werner (Ed.) *The visual neurosciences* (pp. 1391–1401). Boston: MIT Press.
- Burr, D. C., & Santoro, L. (2001). Temporal integration of optic flow, measured by contrast thresholds and by coherence thresholds. *Vision Research*, 41, 1891–1899.
- Cicero, M. T. (1942/55 B.C.). *De Oratore*. Loeb Classical Library.
- Dennett, D. C. & Kinsbourne, M. (1992). Time and the observer. *Behavioural and Brain Science*, 15, 183–247.
- Deubel, H., B. Bridgeman, B., & Schneider, W. X. (1998). Immediate post-saccadic information mediates space constancy. *Vision Research*, 38, 3147–3159.
- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.
- Domini, F., Vuong, Q. C., & Caudek, C. (2002). Temporal integration in structure from motion. *Journal of Experimental Psychology: Human Perception & Performance*, 28, 816–838.
- Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255, 90–92.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501–517.
- Germys, F. De Graef, P., & Verfaillie, K. (2002). Transsaccadic perception of saccade target and flanker objects. *Journal of Experimental Psychology: Human Perception & Performance*, 28, 868–883.
- Gibson, J. J. (1979). *The Ecological approach to visual perception*. Boston: Houghton-Mifflin.
- Goh, J. O. S., Siong, S. C., Park, D., Gutchess, A., Hebrank, A., & Chee, M. W. L. (2004). Cortical areas involved in object, background and object-background processing revealed with functional magnetic resonance adaptation. *Journal of Neuroscience*, 24, 10223–10228.
- Goodale, M. A., & Milner, A. D. (1992). Separate pathways for perception and action. *Trends in Neuroscience*, 15, 20–25.
- Gould, J. D. (1976). Looking at pictures. In R. A. Monty, & J. W. Senders (Eds.) *Eye movements and psychological processes*. Hillsdale, NJ: Lawrence Erlbaum, pp. 323–346.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Atkins (Ed.), *Perception: Vancouver studies in cognitive science* (Vol. 2, pp. 89–110). New York: Oxford University Press.
- Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics*, 65, 58–71.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 519–537.
- Irwin, D. E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics*, 34, 49–57.
- Jonides, J., Irwin, D. E., & Yantis, S. (1982). Integrating visual information from successive fixations. *Science*, 215, 192–194.
- Kanai, R. & Verstraten, F. A. J. (2004). Visual transients without feature changes are sufficient for the percept of a change. *Vision Research* 44, 2233–2240.
- Kubischik, M. (2002). Dynamic spatial representations during saccades in the macaque parietal cortex. Bochum: Ruhr-Universitaet Bochum.
- Kusunoki, M., & Goldberg, M. E. (2003). The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *Journal of Neurophysiology*, 89, 1519–1527.
- Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28, 1311–1328.

- Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41, 3559–3565.
- Lee, A. C. H., Buckley, M. J., Pegman, S. J., Spiers, H., Scabill, V. L., Gaffan, D. et al. (2005). Specialization of the medial temporal lobe for processing objects and scenes. *Hippocampus*, 15, 782–797.
- Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, 3, 525–551.
- Loftus, G. R. (1981). Tachistoscopic simulations of eye fixations on pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 369–376.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.
- Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, 13, 1257–1269.
- McConkie, G. W. and D. Zola (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, 25, 221–224.
- Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, 412, 401.
- Melcher, D., & Kowler, E. (2001). Scene memory and the guidance of saccadic eye movements. *Vision Research*, 41, 3597–3611.
- Melcher, D., & Morrone, M. C. (2003). Spatiotopic integration of visual motion across saccadic eye movements. *Nature Neuroscience*, 6, 877–881.
- Melcher, D. (2005). Spatiotopic transfer of visual form adaptation across saccadic eye movements. *Current Biology*, 15, 1745–1748.
- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes, *Journal of Vision*, 6, 8–17.
- Melcher, D., Papathomas, T. V., & Vidnyanszky, Z. (2005). Implicit attentional selection of bound visual features. *Neuron*, 46, 723–729.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6, 414–417.
- Morgan, M. J., Findlay, J. M., & Watt, R. J. (1982). Aperture viewing: A review and a synthesis. *Quarterly Journal of Experimental Psychology A*, 34, 211–233.
- Nakamura, K., & Colby, C. L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedures of the National Academy of Sciences USA*, 99, 4026–4031.
- Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, 394, 894–896.
- Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Current Biology*, 14, 830–839.
- O'Regan, J. K. , & A. Noë (2001). A sensorimotor account of vision and visual consciousness. *Behavioral Brain Science*, 24, 939–973.
- Piccolino, M., & Wade, N. J. (2006). Flagging early examples of ambiguity. *Perception*, 35, 861–864.
- Pierrot-Deseilligny, C., R. Muri, M., Rivaud-Pechoux, S., Gaymard, B., & Ploner, C. J. (2002). Cortical control of spatial memory in humans: the visuoculomotor model. *Annals of Neurology*, 52, 10–19.
- Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General*, 113, 426–442.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509–522.
- Purkinje, J. (1823). Beobachtungen und Versuche zur Physiologie der Sinne. Beiträge zur Kenntniss des Sehens in subjectiver Hinsicht. Prague: Calve.
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception & Performance*, 18, 849–860.
- Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, 40, 1469–1487.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1995). Image Flicker is as good as saccades in making large scene changes invisible. *Perception*, 24(suppl.), 26–27.
- Ross, J. & A. Ma-Wyatt (2004). Saccades actively maintain perceptual continuity. *Nature Neuroscience*, 7, 65–69.

- Rolls, E. T., Xiang, J., & Franco, L. (2005). Object, space and object-space representations in primate hippocampus. *Journal of Neurophysiology*, *94*, 833–844.
- Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. (2001). Changes in visual perception at the time of saccades. *Trends in Neuroscience*, *24*, 111–121.
- Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*, 5–42.
- Tatler, B. W. (2001). Characterising the visual buffer: Real-world evidence for overwriting early in each fixation. *Perception*, *30*, 993–1006.
- Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, *58*, 931–960.
- Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception*, *32*, 579–592.
- Todd, J. J. & Marois, R. (2004). The capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–753.
- Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, *29*, 757–767.
- Ullman, S. (1983). Computational studies in the interpretation of structure and motion: summary and extension. In J. Beck, B. Hope, and A. Rosenfeld (Eds.) *Human and Machine Vision*. New York: Academic Press.
- Vogel, E. K. & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.
- Yin, C., Shimojo S., Moore, C., & Engel, S. A. (2002). Dynamic shape integration in extrastriate cortex. *Current Biology*, *12*, 1379–1385.