

# The Role of Spatial Memory in Guiding Attention During Natural Vision

**Ran Carmi (carmi@usc.edu)**

Neuroscience Program, University of Southern California  
3641 Watt Way, Los Angeles, CA 90089-2520, USA

**Laurent Itti (itti@usc.edu)**

Neuroscience Program, University of Southern California  
3641 Watt Way, Los Angeles, CA 90089-2520, USA

## Abstract

Paying attention to the right thing at the right time underlies the ability of humans and other animals to learn, perceive, and interact with their environment. A central unresolved question is the time frame in which spatial memory guides attention, with current estimates ranging from a single fixation to seconds, minutes, or even days. Here we answer this question by revealing the time course of attentional selection during natural vision. We asked human participants to visually explore either continuous or scene-shuffled video clips, and quantified the impact of memory-free influences on overt attentional selections (saccades) based on a computational saliency model. Overall, scene shuffling resulted in no significant differences in the impact of memory-free influences compared to continuous viewing. However, abrupt scene transitions (jump cuts) led to sharp peaks in the impact of memory-free influences, which then declined progressively across 7 fixations for up to 2.5 seconds. These results indicate that visual exploration of dynamic scenes critically depends on spatial memory traces that persist across several fixations for up to a couple of seconds.

## Introduction

Attentional selections are determined by interactions between memory-free (bottom-up) and memory-dependent (top-down) influences (Henderson, 2003; James, 1890). Bottom-up influences are stimulus-centric factors, such as visual onsets (Jonides & Yantis, 1988), which can lead to automatic, or reactive, selection of attention targets. Top-down influences refer to behavioral goals and expectations (Yarbus, 1967), which can guide attention proactively based on prior knowledge. Among potential top-down influences, there is a consensus that sensory snapshots are overwritten within a single fixation (up to a few hundred milliseconds), while semantic information can be accumulated and utilized for guiding attention across many fixations. In contrast, the utilization of perceptual memory, especially involving locations of previously fixated targets, has been debated extensively (Chun & Jiang, 1998; Hayhoe, et al., 2003; Hollingworth & Henderson, 2002; Horowitz & Wolfe, 1998; Levin & Simons, 1997; Maljkovic & Nakayama, 1996; Melcher & Kowler, 2001; Najemnik & Geisler, 2005; Rensink, 2002).

Two competing theories about the accumulation and utilization of perceptual memory have emerged. According to the “world as an outside memory” (WOM) theory (O’Regan, 1992; Rensink, 2000), humans rely on the

continuity of the world to access external information on demand, leading to conscious perceptions that are seemingly rich and continuous, in the absence of perceptual memory that persists across several fixations. More recent studies have shown that perceptual information is accumulated over time during inspection of static scenes (Hollingworth & Henderson, 2002; Melcher & Kowler, 2001). Nevertheless, other results strongly supported the WOM theory in the context of memory utilization (Horowitz & Wolfe, 1998; Melcher & Kowler, 2001; Najemnik & Geisler, 2005).

In contrast, the “implicit memory” (IM) theory (Chun & Nakayama, 2000; Land & Furneaux, 1997) postulates that previously attended targets typically trigger perceptual memory traces that are routinely utilized across several fixations. Such memory traces may often be hidden from conscious introspection, leading to the common failure of human observers to explicitly report changes in scenes (Rensink, 2002). Proponents of the IM theory argue that location-specific memories are critical for facilitating target detection, especially in the context of visually-guided actions that may require motor planning in spatial coordinates (Hayhoe et al., 2003; Land & McLeod, 2000).

This study relied on computational tools to establish the time frame in which perceptual memory guides attention during visual exploration of novel dynamic scenes. The rationale for studying this behavior is that humans, especially during development (Atkinson & Braddick, 2003), spend a lot of time visually exploring new people or environments, without necessarily being engaged in highly practiced goal-oriented behaviors.

## Materials and Methods

### Participants

16 paid participants (6 women and 10 men), 23- to 32-years old, provided written informed consent, and were compensated for their time (\$12/h). All participants were healthy, had normal or corrected-to-normal vision, and were naïve as to the purpose of the experiment.

### Stimuli

50 video clips (30 Hz, 640x80 pixels/frame, 4.5-30 s, mean  $\pm$  s.d.: 21.83  $\pm$  8.41 s, no audio) from 12 heterogeneous sources, including indoor/outdoor, daytime/nighttime scenes shot at various locations in Los Angeles, video games, television newscasts, interviews, commercials, and sporting

events. These continuous clips were cut every 1-3 s ( $2.09 \pm 0.57$  s) into 523 clip snippets (clippets), which were re-assembled into 50 MTV-style clips (see Fig. 1a). The range of clippet lengths was motivated by previous studies showing that inter-observer agreement in attentional selection diverges significantly within this time frame (Mannan, et al., 1997).

To minimize the ability of participants to anticipate the exact timing of jump cuts, the clippet lengths were randomized within the chosen time range. Continuous and MTV-style clips were matched in length, and each MTV-style clip contained at most one clippet from a given continuous clip. This MTV-style manipulation was inspired by the cinematic practice of introducing jump cuts to compress time while preserving semantic continuity (Anderson, 1996; Hochberg, 1986). The critical difference is that our MTV-style clips were deliberately designed to maximize semantic unrelatedness between adjacent scenes depicted in different clippets, and no attempt was made to mask the jump cuts.

### Experimental design

Participants were divided randomly into 2 groups of 8 participants each. One group inspected continuous clips, and the other group inspected MTV-style clips. All participants sat with their chin supported before a 22" color monitor (60 Hz refresh rate) at a viewing distance of 80 cm ( $28^\circ \times 21^\circ$  usable field-of-view). Their task was: "try to follow the main actors and actions, and expect to be asked general questions after the eye-tracking session is over". Participants were told that the questions will not pertain to small details, such as specific small objects, or the content of text messages, but would rather help the experimenters evaluate their general understanding of what they had watched. The purpose of the task was to allow participants to engage in natural visual exploration, while encouraging them to pay close attention to the display throughout the viewing session.

### Data acquisition and processing

Instantaneous position of the right eye was recorded using an infrared-video-based eye tracker (ISCAN RK-464, 240 Hz,  $<1^\circ$  spatial error). Calibration and saccade extraction procedures have been described elsewhere (Itti, 2005). In short, the raw eye-position data was segmented into saccade, blink and fixation/smooth-pursuit periods. 8812 and 10187 saccades were extracted from the raw eye-position data of the continuous and MTV-style groups, respectively.

### Attention-priority maps

Attention-priority maps were computed based on the saliency model (Carmi & Itti, 2006). Multi-scale feature pyramids were generated at the input frame rate (30Hz), and converted through a series of computations into proto-saliency maps. Each proto-saliency map was fed into a two-dimensional layer of leaky integrator neurons that accounted for sensory persistence and provided temporal smoothing at 10kHz.

### Prediction of single saccades

Normalized prediction for all human saccades was calculated by sampling the attention-priority map at the saccade target, and dividing that local value by the global maximal value in the instantaneous map. Concurrently, a baseline saliency sample was taken at a random location based on a spatially uniform distribution of potential targets. Further details about the saliency sampling procedure and its rationale can be found elsewhere (Carmi & Itti, 2006).

### DOH metric

The difference of histograms (DOH) metric outputs a scalar that quantifies the human tendency to initiate saccades towards high priority targets. It measures the rightward shift of the human saccade histogram (saccade frequency as a function of model response) relative to the random saccade histogram:

$$DOH = (1/DOH_I) \times \sum_{i=1}^n W_i \times (H_i - R_i)$$

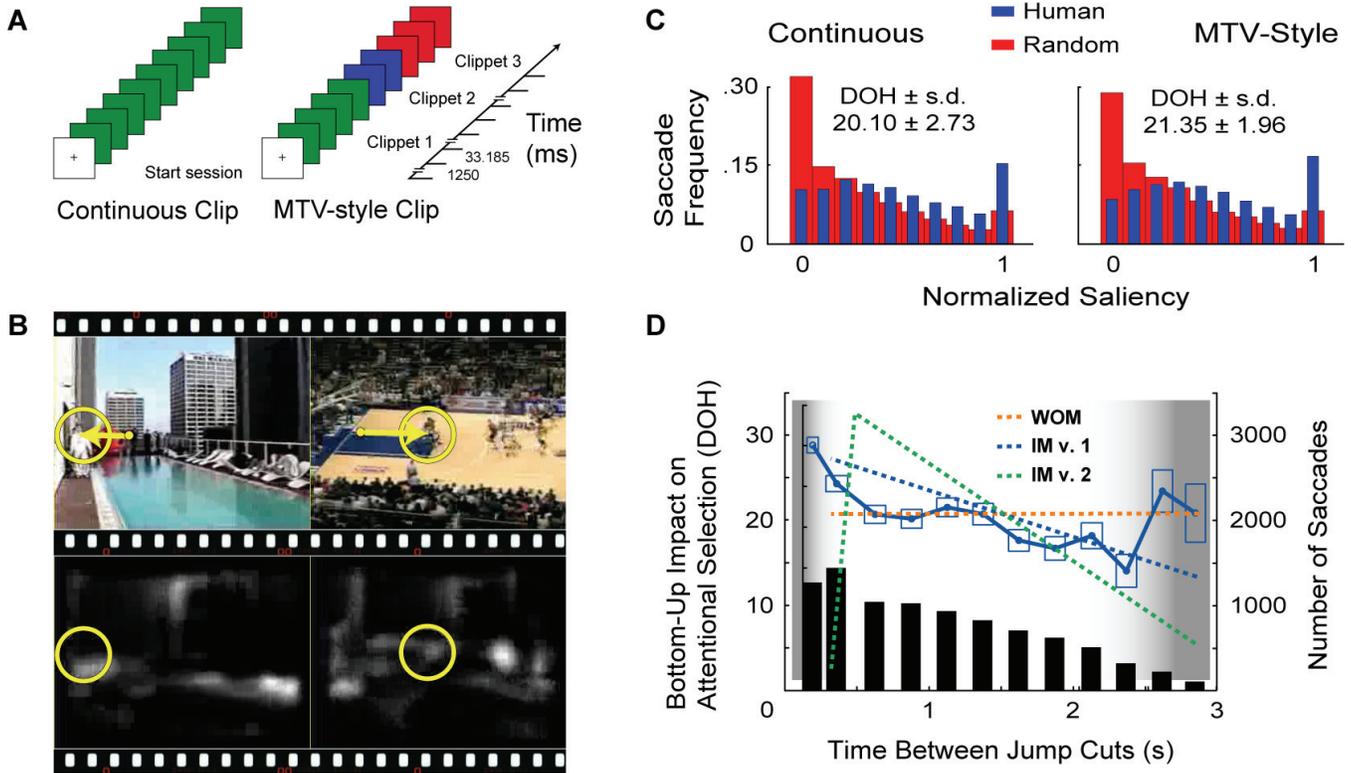
where  $H_i$  and  $R_i$  are the fractions of human and random saccades, respectively, which fall in bin  $i$  with boundaries  $(i-1)/n$ ,  $i/n$ , where  $n=10$  is the number of bins, and  $W_i = (i - 0.5)/n$  is the mid-value of bin  $i$ .

DOH is normalized by the theoretical (ideal) upper bound, thus providing a conservative estimate for the percentage-wise impact of bottom-up versus top-down influences on attentional selection (Carmi & Itti, 2006).

### Results

The oft-competitive interactions between bottom-up and top-down influences imply that the less participants act reactively (based on the instantaneous input), the more likely they are to act proactively (based on memory), and vice versa. Thus, even though computational models of top-down guidance during natural vision are unavailable at present, we can nevertheless infer modulations in top-down guidance by quantifying modulations in the impact of bottom-up influences. To this end, we first established the availability of potential bottom-up influences based on a computational saliency model that has no memory other than sensory persistence (Itti, 2005; Itti & Koch, 2000).

The key design features that distinguish this particular saliency model from the available alternatives are described elsewhere (Carmi & Itti, 2006). In short, they include the detection of dynamic signals, spatial interactions between local detectors, and neural grounding. While the results presented here are not tied to the implementation details of a particular saliency model, they do depend on accurate estimation of bottom-up influences. The detection of dynamic signals is particularly relevant in this context, because previous studies have shown that stimulus changes are the strongest bottom-up influences on attentional selection in dynamic environments (Findlay & Walker, 1999; Jonides & Yantis, 1988).



**Figure 1.** Continuous and MTV-style Experiments.

(A) Schematic of continuous and MTV-style clips. Colored squares depict video frames. Different colors reflect semantically unrelated dynamic scenes (clippets).

(B) Saccades that straddle an MTV-style jump cut (mtvclip04, participant JR). The temporal offset between the saccade initiation times was 431.4 ms, during which JR smooth-pursued the fixated person who was moving to the right due to a leftward camera pan (the camera was stationary when the second saccade was initiated). Superimposed markers (yellow) depict eye-position prior to saccade initiation (filled circle), saccade trajectory (arrow), and saccade target (ring). Upper filmstrips show the instantaneous input frames at the time of saccade initiation. Lower filmstrips show the corresponding saliency maps.

(C) Saccade frequency, pooled over all participants and clips from the continuous (left) and MTV-style (right) groups, as a function of normalized saliency at saccade target. Red and blue vertical bars represent the random and human saccade histograms, respectively. Bottom-up impact was measured based on the saliency model using the Difference of Histograms (DOH) metric, which quantifies the human tendency to saccade towards salient locations. Numbers above histograms depict the mean DOH  $\pm$  inter-participant s.d.

(D) Bottom-up impact as a function of average time to saccade from the last jump cut (blue circles). Dashed lines depict theoretical predictions, as described in the text. Solid line depicts actual data. Saccades were pooled across all jump cuts, and binned into consecutive 250 ms intervals. Vertical and horizontal sides of error boxes depict the bootstrap s.d. of DOH values based on 1000 subsamples, and the s.d. of the average time to saccade, respectively. Shaded backgrounds indicate time periods in which participants were increasingly more sensitive to bottom-up influences. Light background indicates the time period in which top-down influences played an increasingly more dominant role in guiding attention. Black bars depict the number of saccades in each temporal bin.

### Average bottom-up impact on attentional selection

By highlighting conspicuous locations in the display, saliency maps reflect the potential availability of bottom-up influences, but not their actual impact on attentional selection. To this end, we developed the DOH metric that quantifies the agreement between human attentional selection and model-generated attention-priority maps.

Fig. 1B shows two saccades that straddle a jump cut. Fig. 1C shows saccade frequencies as a function of normalized saliency in the continuous (left) and MTV-style (right) viewing condition. It demonstrates that approximately 30% of the random saccades targeted the lowest possible saliency (0-10% of the max), while 15% of the human saccades targeted the highest possible saliency (90-100% of the max), with the remaining saccades targeting intermediate saliency values. The

random saccade histogram reflects the actual distribution of saliency values, while the human saccade histogram shows the distribution of saliency values sampled by human saccades. The DOH metric measures the human tendency to visit salient locations by quantifying the rightward shift of the human saccade histogram relative to the random saccade histogram.

Given that the overall frame content is identical across viewing conditions, the MTV-style distribution of saliency values, as reflected by the random saccade histogram, is expected to closely resemble its counterpart from the continuous condition. If so, three scenarios are possible for the human saccades:

(1) The MTV-style human histogram might be shifted to the left compared to its continuous counterpart. This result would indicate that human observers were less likely to select targets based on bottom-up influences in the MTV-style condition, even though the potential availability of valid top-down influences was limited compared to the continuous condition. A likely interpretation would be increased confusion due to the rapid succession of novel scenes in the MTV-style condition, which may lead observers to select targets more randomly than in the continuous condition. This scenario would be uninformative for distinguishing between memory utilization theories, and we made sure during preliminary testing that human observers could successfully follow the main actions and actors in the MTV-style condition.

(2) The MTV-style manipulation might have no effect on the relative shift of the human saccade histogram. This result would indicate that top-down influences during continuous viewing are rarely utilized beyond 2 s (the average length of persistent context in MTV-style clips), leading to the same overall bottom-up impact in both viewing conditions. This scenario is consistent with either the WOM theory or a transient version of the IM theory, in which perceptual memory is utilized across fixations, but rarely beyond 2 s.

(3) The MTV-style human histogram might be shifted to the right compared to its continuous counterpart. This result would indicate that top-down influences that guided attention beyond 2 s during continuous viewing were replaced by bottom-up influences during MTV-style viewing. This scenario is inconsistent with the WOM theory, and instead supports an even more persistent version of the IM theory.

Fig. 1C (right) shows the actual saccade histograms in the MTV-style condition. The random saccade histogram mirrors its continuous counterpart, while the human saccade histogram is slightly shifted to the right compared to the continuous condition. However, these small differences across viewing conditions are not statistically significant ( $t[14]=1.05$ ,  $p>0.20$ , based on the inter-participant s.d.). These results indicate that top-down influences were not utilized beyond 2 s in either of the viewing conditions. As such, they are consistent with either the WOM theory (O'Regan, 1992; Rensink, 2000), which claims that perceptual memory is not utilized across fixations at all, or a transient version of the IM

theory (Chun & Nakayama, 2000; Land & Furneaux, 1997), in which perceptual memory is utilized across fixations, but rarely beyond 2 s. To discriminate between these competing interpretations, we analyzed the bottom-up impact on attentional selection at a finer time scale than 2 s, as described in the next section.

### Time course of bottom-up impact on attentional selection

Attentional selections after jump cuts are particularly informative for examining memory utilization, because these events abruptly invalidate any attention-guiding internal representations, followed by the potential re-establishment of such representations. Competing theories of memory utilization predict differential patterns of memory utilization after jump cuts, as described below:

(1) Scenario #1 (Fig. 1D, orange dashed line): The WOM theory (O'Regan, 1992; Rensink, 2000) holds that perceptual memory is constantly being overwritten on a fixation-by-fixation basis even during continuous viewing. Hence, jump cuts are expected to have little to no effect on the balance between bottom-up and top-down influences.

(2) Scenario #2 (Fig. 1D, blue dashed line): the attention system relies heavily on perceptual memory, but this reliance is contingent on the availability of persistent context. This version of the IM theory (Chun & Nakayama, 2000; Land & Furneaux, 1997) predicts that internal representations elicited by cliplet X (before the jump cut) will be swiftly replaced by novel internal representations triggered by visual inputs from cliplet X+1 (after the jump cut). In this case, the impact of bottom-up influences is expected to peak early on, because they are faster acting than top-down influences (Henderson, 2003). After the scene gist and layout are recognized, top-down influences are expected to kick in and gradually replace bottom-up influences.

(3) Scenario #3 (Fig. 1D, green dashed line): internal representations based on cliplet X interfere with the accumulation and utilization of novel representations based on cliplet X+1. This version of the IM theory predicts that bottom-up impact will hit a nadir immediately after jump cuts, because at that point in time observers will be selecting attention targets based on irrelevant attention-priority maps. The same temporal pattern to the one described in scenario #2 is expected to emerge after an initial confusion.

Fig. 1D (solid blue plot) shows the human/model agreement as a function of time between MTV-style jump cuts. Bottom-up impact on attentional selection peaked during the initial 250 ms after jump cuts, as depicted by the first data point ( $\text{DOH} = 28.81 \pm 0.93$ ). This result demonstrates that the visual system adapts rapidly to changing conditions, which is inconsistent with the notion of memory utilization across jump cuts (scenario #3). However, both the WOM theory (scenario #1) and the context-contingent version of the IM theory (scenario #2) are consistent with this result. The following monotonic decreases in bottom-up impact for up to 2.5 s after jump cuts indicate corresponding increases in the impact of

competing top-down influences. This result rules out scenario #1, leaving scenario #2 as the only scenario that is consistent with the results. To verify that these results are not artifacts of the DOH metric or the temporal binning procedure, we performed additional analyses of bottom-up impact following jump cuts based on an alternative metric (average saliency at human saccade targets) and/or saccade index (not shown). Either way, a consistent decrease in bottom-up impact following jump cuts was evident across 7 consecutive saccades for up to 2.5 s. Interestingly, Fig. 1D also shows delayed increases in bottom-up impact occurring 2.5 s after jump cuts. This result cannot be explained by either of the memory utilization theories addressed here, and is discussed below in the section: "Attention and scene understanding".

## Discussion

Our results support and elaborate the IM theory (Chun & Nakayama, 2000; Land & Furneaux, 1997) by establishing quantitatively the time frame in which perceptual memory guides attention during natural vision. Specifically, we showed that perceptual memory is heavily utilized across up to 7 consecutive saccades and 2.5 s, even in the absence of visually-guided actions. Furthermore, we demonstrated that memory utilization is strongly contingent on the availability of semantically persistent context. These results refute predictions made by the WOM theory (O'Regan, 1992; Rensink, 2000), other than in extreme circumstances, such as shortly after jump cuts, when it would be maladaptive to rely on perceptual memory for guiding attention.

## Memory and attention

The existing controversy about the time frame in which perceptual memory guides attention may be attributable to the difficulty of generalizing conclusions from studies performed in artificial laboratory conditions. Alternatively, it has been suggested that the plethora of memory utilization estimates may simply reflect the flexibility of the attention system, which can automatically make pragmatic choices between relying on vision versus memory, depending on which source of information is more likely to improve performance (Oliva, et al., 2004).

We propose that this intuitively appealing hypothesis suffers from two major shortcomings:

(1) Weak predictive power. If every situation potentially leads to a different balance between vision and memory, then any result could simply be explained away as reflecting an unknown pragmatic optimization.

(2) Vision and memory are not interchangeable sources of information. For example: while watching players taking jump shots in a basketball game, our participants sometimes made saccades towards the hoop, even before the ball left the player's hands (i.e., before the ball's trajectory could have been analyzed based on its visual motion). It appears that such attentional selections depended on simultaneous integration of several bottom-up and top-down influences, including the movement of

the player, prior knowledge of what typically happens to balls when players take jump shots, and the exact location of the hoop. The notion that vision and memory can be used interchangeably may apply exclusively to laboratory stimuli that are designed to minimize the involvement of prior world knowledge.

Future work could build on our contribution in several ways. One promising direction would involve functional neuroimaging using fMRI (Hasson, et al., 2004) or EEG (Michel, et al., 2004). The MTV-style manipulation provides a controlled stimulus-based technique for repeatedly inducing amnesia, followed by increased reliance on top-down influences. As such, it could potentially be used for mapping the spatiotemporal connectivity between brain areas that control attentional selection. Recurring surges in activation after jump cuts would identify brain areas that are involved with bottom-up processing, while increased activation over time between jump cuts would indicate top-down processing.

## Attention and scene understanding

It is sometimes argued that attention plays a minor role in scene understanding because people can recognize the gist of static scenes very rapidly without making any attention shifts (Henderson, 2003). However, accurate perception of dynamic scenes, which pose more complex perceptual challenges compared to static scenes, may require well-coordinated attentional selections.

The surprising increases in bottom-up impact that occurred 2.5 s following jump cuts (Fig. 1D) lead us to propose novel testable hypotheses about potential interactions between attention and scene understanding. One possible explanation is that participants learned to anticipate the occurrence of jump cuts over the course of the experiment. Such anticipations may have prompted observers to shift to "bottom-up mode" when they perceived a high likelihood of an impending jump cut, thus minimizing the frequency of anomalous memory-driven selections following jump cuts. Alternatively, the delayed increase in bottom-up impact may reflect late novelty effects arising from the typical rate of change in natural stimuli. According to this hypothesis, the shift to "bottom-up mode" was triggered by novel bottom-up influences, and its perceptual function is to keep observers appraised of new events or pertinent changes to previously attended targets. To test the relative impact of novelty versus anticipatory effects, future studies could manipulate the rate of stimulus changes by manipulating the frame rate and/or the lengths of cliplets.

## Natural versus artificial approaches to studying vision

This study has important implications for the raging debate about using natural versus artificial stimuli for studying vision (Felsen & Dan, 2005; Rust & Movshon, 2005). Proponents of the artificial approach argue that it is the best way to perform hypothesis-driven experiments, and contend that it is sufficient for characterizing neural mechanisms employed in everyday life. Furthermore,

they criticize the natural approach as being too difficult and also unhelpful for generating novel hypotheses. In contrast, proponents of the natural approach point to the disappointingly slow progress in understanding neural computation at both the network and single neuron levels, and claim that the overwhelming reliance on overly simplified artificial stimuli has greatly impeded progress. This important methodological debate has so far focused on low-level vision and the response properties of single neurons. We propose that the disadvantages of using highly simplified stimuli are particularly strong in the context of high-level vision. Future related studies could avoid such pitfalls by using hybrid natural-artificial stimuli, as was done here, or by employing video game engines to generate complex artificial stimuli.

### Acknowledgments

Supported by grants from NSF, NEI, and NIMA. Computation for this study was supported by the University of Southern California Center for High Performance Computing and Communications ([www.usc.edu/hpcc](http://www.usc.edu/hpcc)).

### References

- Anderson, J.D. (1996). *The reality of illusion: an ecological approach to cognitive film theory*. Carbondale: Southern Illinois University Press.
- Atkinson, J., & Braddick, O. (2003). *Neurobiological Models of Normal and Abnormal Visual Development*. Hove, East Sussex ; New York: Psychology Press.
- Carmi, R., & Itti, L. (2006). Causal Saliency Effects During Natural Vision. *ETRA* (pp. 11-18, 173). San Diego, CA: ACM.
- Chun, M.M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognit. Psychol.*, 36 (1), 28-71.
- Chun, M.M., & Nakayama, K. (2000). On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Vis. Cogn.*, 7 (1-3), 65-81.
- Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nat. Neurosci.*, 8 (12), 1643-1646.
- Findlay, J.M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behav. Brain Sci.*, 22 (4), 661-674.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303 (5664), 1634-1640.
- Hayhoe, M.M., Shrivastava, A., Mruczek, R., & Pelz, J.B. (2003). Visual memory and motor planning in a natural task. *J. Vis.*, 3 (1), 49-63.
- Henderson, J.M. (2003). Human gaze control during real-world scene perception. *Trends. Cogn. Sci.*, 7 (11), 498-504.
- Hochberg, J.E. (1986). Representation of motion and space in video and cinematic displays. In: K.R. Boff, R. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance*: Vol. 1. Sensory processes and perception, pp. 22-21 to 22-64). New York: Wiley.
- Hollingworth, A., & Henderson, J.M. (2002). Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol. Hum. Percept. Perform.*, 28 (1), 113-136.
- Horowitz, T.S., & Wolfe, J.M. (1998). Visual search has no memory. *Nature*, 394 (6693), 575-577.
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Vis. Cogn.*, 12 (6), 1093-1123.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.*, 40 (10-12), 1489-1506.
- James, W. (1890). *Principles of Psychology*. Oxford, England: Henry Holt.
- Jonides, J., & Yantis, S. (1988). Uniqueness of Abrupt Visual Onset in Capturing Attention. *Percept. Psychophys.*, 43 (4), 346-354.
- Land, M.F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 352 (1358), 1231-1239.
- Land, M.F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nat. Neurosci.*, 3 (12), 1340-1345.
- Levin, D.T., & Simons, D.J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychon. B. Rev.*, 4 (4), 501-506.
- Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Percept. Psychophys.*, 58 (7), 977-991.
- Mannan, S.K., Ruddock, K.H., & Wooding, D.S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception*, 26, 1059-1072.
- Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Res.*, 41 (25-26), 3597-3611.
- Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clin. Neurophysiol.*, 115 (10), 2195-2222.
- Najemnik, J., & Geisler, W.S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434 (7031), 387-391.
- Oliva, A., Wolfe, J.M., & Arsenio, H.C. (2004). Panoramic search: the interaction of memory and vision in search through a familiar scene. *J. Exp. Psychol. Hum. Percept. Perform.*, 30 (6), 1132-1146.
- O'Regan, J.K. (1992). Solving the "real" mysteries of visual perception: the world as an outside memory. *Can. J. Psychol.*, 46 (3), 461-488.
- Rensink, R.A. (2000). Seeing, sensing, and scrutinizing. *Vision Res.*, 40 (10-12), 1469-1487.
- Rensink, R.A. (2002). Change detection. *Annu. Rev. Psychol.*, 53, 245-277.
- Rust, N.C., & Movshon, J.A. (2005). In praise of artifice. *Nat. Neurosci.*, 8 (12), 1647-1650.
- Yarbus, A.L. (1967). *Eye movements and vision*. New York: Plenum Press.