Human gaze control during real-world scene perception

John M. Henderson

Psychology Research Building, Michigan State University, East Lansing, MI 48824-1117, USA

In human vision, acuity and color sensitivity are best at the point of fixation, and the visual-cognitive system exploits this fact by actively controlling gaze to direct fixation towards important and informative scene regions in real time as needed. How gaze control operates over complex real-world scenes has recently become of central concern in several core cognitive science disciplines including cognitive psychology, visual neuroscience, and machine vision. This article reviews current approaches and empirical findings in human gaze control during real-world scene perception.

During human scene perception, high quality visual information is acquired only from a limited spatial region surrounding the center of gaze (the fovea). Visual quality falls off rapidly and continuously from the center of gaze into a low-resolution visual surround. We move our eyes about three times each second via rapid eye movements (saccades) to reorient the fovea through the scene. Pattern information is only acquired during periods of relative gaze stability (fixations) owing to 'saccadic suppression' during the saccades themselves [1–3]. Gaze control is the process of directing fixation through a scene in real time in the service of ongoing perceptual, cognitive and behavioral activity (Figure 1).

There are at least three reasons why gaze control is an important topic in scene perception. First, vision is an active process in which the viewer seeks out task-relevant visual information. In fact, virtually all animals with developed visual systems actively control their gaze using eye, head, and/or body movements [4]. Active vision ensures that high quality visual information is available when it is needed, and also simplifies a variety of otherwise difficult computational problems [5,6]. A complete theory of vision and visual cognition requires understanding how ongoing visual and cognitive processes control the orientation of the eyes in real time, and in turn how vision and cognition are affected by gaze direction over time.

Second, because attention plays a central role in visual and cognitive processing, and because eye movements are an overt behavioral manifestation of the allocation of attention in a scene, eye movements serve as a window into the operation of the attentional system. Indeed, although behavioral and neurophysiological evidence suggest that internal visual attentional systems (*covert* visual attention) and eye movements (*overt* visual attention) can be dissociated [7], the strong natural relationship between covert and overt attention has led investigators to suggest that studying covert visual attention independently of overt attention is misguided [8].

Third, eye movements provide an unobtrusive, sensitive, real-time behavioral index of ongoing visual and cognitive processing. This fact has been exploited to a significant degree in the study of perceptual and linguistic processes in reading [9–11], and is coming to play a similarly important role in studies of language production and spoken language comprehension [12,13]. Eye movements have been exploited to a lesser extent to understand visual and cognitive processes in scene perception, although after 25 years of relative inactivity, the study of gaze control in scenes has recently experienced a rebirth. Several advances in technology have sparked this renewed interest, including more accurate and robust stationary eyetrackers, new mobile eyetrackers that can be used in the natural environment, progress in computer graphics technology enabling presentation of full color scene images under precisely controlled conditions, and new computational methods for analyzing image properties (Figure 2).

Early studies of gaze control demonstrated that empty, uniform, and uninformative scene regions are often not fixated. Viewers instead concentrate their fixations, including the very first fixation in a scene, on interesting and informative regions (Box 1 and Figure 3). What

Figure 1. Scan pattern of one viewer during visual search. The viewer was counting the number of people in the scene. The circles represent fixations (scaled in size to their durations, which are shown in milliseconds) and the lines represent saccades. The figure illustrates that fixation durations are variable even for a single viewer examining a single scene.

Corresponding author: John M. Henderson (john@eyelab.msu.edu).

http://tics.trends.com 1364-6613/$ - see front matter doi:10.1016/j.tics.2003.09.006

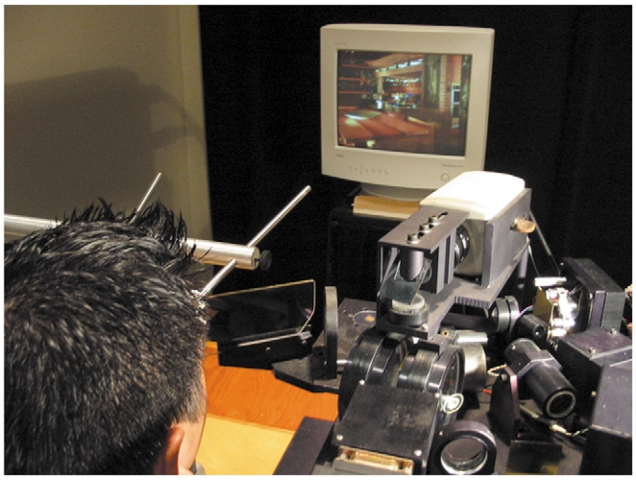

Figure 2. Participant on a dual-Purkinje-image eyetracker. The availability of eyetrackers with high spatial and temporal resolution, along with high quality visual displays, has greatly contributed to recent gains in our understanding of gaze control in scenes. (Photo courtesy of Dan Gajewski.)

constitutes an 'interesting and informative' scene region? Recent work on gaze control has focused on two potential answers: bottom-up stimulus-based information generated from the image, and top-down memory-based knowledge generated from internal visual and cognitive systems.

Stimulus-based gaze control

Three general approaches have been adopted to investigate the image properties that influence where a viewer will fixate in a scene. First, scene patches centered at each fixation position are analyzed to determine whether they differ in some image property from unselected patches. Using this 'scene statistics' approach, investigators have found that high spatial frequency content and edge density are somewhat greater at fixation sites [14,15], and that local contrast (the standard deviation of intensity in a patch) is higher and two-point correlation (intensity of the fixated point and nearby points) is lower for fixated scene patches than unfixated patches [16–18].

Second, properties of early vision are instantiated in a computational model and used to predict fixation positions. One prominent model of this type generates visual saliency based on the known properties of primary visual cortex [19–22]. In this 'saliency-map' approach, the visual properties present in an image give rise to a representation (the saliency map) that explicitly marks regions that are different from their surround on one or more image dimensions such as color, intensity, contrast, edge orientation, and so forth over multiple spatial scales. The maps generated for each image dimension are then combined to create a single saliency map. The intuition behind this approach is that regions that are uniform along some image dimension are uninformative, whereas those that differ from neighboring regions across spatial scales are potentially informative. The salient points in the map serve as a prediction about the spatial distribution of gaze in a scene, and these points can be correlated with observed human fixations [23,24]. The saliency map approach serves an important heuristic function in the study of gaze control because it provides an explicit model that generates precise quantitative predictions about fixation locations and their sequences. Important questions remaining to be answered within this approach include the following:

Box 1. An early study of human gaze control

Buswell was among the first investigators to measure both a viewer's direction of gaze and the duration of each fixation in a scene [38]. Using an ingenious apparatus, Buswell reflected light from the cornea onto photographic film and quantized time by interposing the blades of a fan rotating at 30 Hz into the reflection. Buswell demonstrated that fixations are not randomly placed in a scene; instead, viewers tend to cluster fixations on informative image regions. Buswell also accurately estimated mean fixation durations and saccade amplitudes as well as the variability in these measures. Based on his observations, Buswell concluded that there was an important relationship between eye movements and visual attention: 'Eye movements are unconscious adjustments to the demands of attention during a visual experience.' ([38], p. 9).

How many times is a saliency map computed for a given scene?

It may be that one saliency map is computed across the entire scene during the first fixation on that scene, or that a new saliency map is computed in each fixation. In the former approach, the initial map could be used to generate an ordered set of sites that are fixated in turn. This approach assumes that a single map is retained over multiple fixations, an assumption that is suspect given the evidence that metrically precise sensory information about a scene is not retained across saccades [25,26]. The alternative approach is to compute the saliency map anew following each successive fixation. This approach does

Figure 3. Distribution of fixations over a scene. Representation of all fixations (indicated by red dots) produced by 20 participants viewing a scene in preparation for a later memory test. Note that the fixations are clustered on regions containing objects; relatively homogenous regions of a scene receive few if any fixations.

away with the need to retain the saliency map across fixations, but potentially increases computational load because a new saliency map must be generated every few hundred milliseconds. This approach also requires that 'inhibition of return' (IOR) [27] be retained across fixations and properly assigned to points within the regenerated saliency map to ensure that gaze does not oscillate between highly salient points. Given that IOR appears to be object-based as well as space-based in humans [28], this problem seems tractable.

What image properties should be included in the saliency map?

One prominent model assumes that the saliency map can be derived from a weighted linear combination of spatial orientation, intensity, and color [19], but there is as yet no strong evidence that these specific features have a unique or even central role in determining fixation placement in scenes. Predictions of fixation positions based on these features correlate with observed human fixations, with the correlations decreasing as a visual pattern becomes more meaningful [24]. Evidence from the scene statistics method suggests that additional image properties might need to be implemented in saliency-map models to account for gaze control completely [14–18].

How should stimulus-based and knowledge-based information be combined?

The fact that gaze control draws on stored knowledge implies that image properties about potential fixation targets must somehow be combined with top-down constraints. How is this accomplished? One approach is to construct the initial stimulus-based saliency map taking relevant knowledge (e.g. visual properties of a search target) into account from the outset [29]. Another approach is to compute a stimulus-based saliency map independently of other knowledge-based maps. For example, Oliva *et al.* [23] filtered an image-based saliency map using a separate knowledge-based map highlighting regions likely to contain a specific target. Other methods are certainly possible. Which approach best accounts for human gaze control, and which will best support artificial active foveated vision systems, is an important current topic of investigation.

Where in the brain is the saliency map computed and represented?

A final issue concerns the neural implementation of the saliency map. Is there a single neural map, perhaps computed directly over image properties in V1 [30]? Or might there be multiple maps computed over multiple brain areas combining input from a variety of bottom-up and top-down sources, as has been suggested in the spatial attention literature [31]. This issue is likely to receive increased scrutiny in the coming years [30,32,33].

Correlation or causation?

A shortcoming of both the scene statistics and saliency map approaches to human gaze control is that they are correlational techniques, so they do not allow a causal link to be established between image properties and fixation site selection. A third method establishes causality by directly manipulating the information present in an image. For example, foveal and extra-foveal visual information have been manipulated independently using the 'moving-window technique' [34]. Results from these studies indicate that high spatial frequency information (edges) is preferentially used over low spatial frequency information to direct gaze to peripheral scene regions. More studies of this type will be needed to directly test hypotheses about the influence of scene properties on gaze control.

Knowledge-driven gaze control

Human eye movement control is 'smart' in the sense that it draws not only on currently available visual input, but also on several cognitive systems, including short-term memory for previously attended information in the current scene, stored long-term visual, spatial and semantic information about other similar scenes, and the goals and plans of the viewer. In fact, fixation sites are less strongly tied to visual saliency when meaningful scenes are viewed during active tasks [23,35,36,37]. The modulation or replacement of visual saliency by knowledge-driven control can increase over time within a scene-viewing episode as more knowledge is acquired about the identities and meanings of previously fixated objects and their relationships to each other and to the scene [35]. But even the very first saccade in a scene can take the eyes in the likely direction of a search target, whether or not the target is present, presumably because the global scene gist and spatial layout acquired from the first fixation provide important information about where a particular object is likely to be found [23,35].

Episodic scene knowledge

Henderson and Ferreira [13] provided a typology of the knowledge available to the human gaze control system. This knowledge includes information about a specific scene that can be learned over the short term in the current perceptual encounter (short-term episodic scene knowledge) and over the longer term across multiple encounters (long-term episodic scene knowledge). An example of short-term episodic knowledge is the memory that the latest issue of *Trends in Cognitive Sciences* is on my computer table. Short-term knowledge supports a viewer's propensity to refixate areas of the current scene that are semantically interesting or informative [35,38,39,40], and ensures that objects are fixated when needed during motor interaction with the environment [36]. Long-term episodic knowledge involves information about a particular scene acquired and retained over time, such as knowing that my office clock resides on my filing cabinet. Recent

Box 2. What is gaze control for?

Scenes can be identified and their gist apprehended very rapidly, well within the duration of a single fixation [44,47,49,50]. This rapid apprehension may require little attention [51] and can be based on global image statistics that are predictive of the scene's identity and semantic gist [52,53]. Given that scenes are identified and their gist understood within a fixation, what function does gaze control serve?

Studies of both change detection and object identification demonstrate that close or direct fixation is typically needed to identify objects in scenes and to perceive their visual details [54,55,56,57]. Fixation is also tightly tied to memory encoding, both for short-term memory [42,55,58,] and long-term memory [42,59]. In addition, fixation provides a deictic pointer to entities in the world. This pointer can serve as the origin for coordinate system transformations in vision, cognition, and motor control [5]. The deictic function of gaze also facilitates language production, comprehension, and acquisition by providing information about the producer's focus of attention [43,60,61].

Box 3. Does scene viewing produce systematic scan patterns?

The study of eye movement sequences (scan patterns) has received less attention among scene researchers than perhaps it should. One reason for this lack of interest is that the general concept of scan patterns became entangled with a specific theory that did not survive close empirical scrutiny. In 'scan-path theory', the scan pattern produced during complex image viewing was assumed to be an integral part of the memory for that image [62,63]. Learning a new image was taken to involve encoding both its visual features and the gaze sequence used to acquire them, with the motor pattern becoming part of an integrated memory representation. Recognition was assumed to require recapitulating the gaze sequence. Similarity in scan patterns across learning and recognition was taken to support the theory. However, two results were inconsistent with scan path theory. First, it is not necessary for a viewer to make eye movements to recognize scenes that were learned using eye movements, as shown by studies of scene identification within a single fixation [44,47,49,50]. Second, although different viewers tend to fixate similar regions of a given scene (so fixations cluster on informative regions; see Figure 3 in main text), the sequence of fixations over those regions is highly variable. Furthermore, a given viewer shows very little consistency in scan pattern across repetitions of a given image [64,65]. Lack of empirical support for scan path theory led to a general reduction in interest in the issue of scan patterns.

The nature of scan patterns in scenes is a relatively understudied issue, and future work may produce evidence for some consistency over scenes. Such an effect need not be a consequence of the storage and recapitulation of scan patterns, as proposed by scan path theory. Instead, similarity in scan patterns could be the result of stimulus-based or knowledge-based control. For example, the saliency map approach assumes that the relative ordering of saliency in a scene should predict the sequence of fixations taken through that scene. To the extent that saliency remains unchanged, scan patterns should be similar over scene repetition. Knowledge-based approaches that order potential saccade targets by task relevance similarly predict some consistency in scan patterns within and across viewers. To facilitate future investigation, investigators might want to adopt the theory-neutral term 'scan pattern' to refer to an observed sequence of fixations in a scene.

evidence suggests that good memory for the visual detail of a viewed scene is preserved over relatively long periods of time [41,42]. An interesting example of the influence of episodic memory on gaze control is a viewer's tendency to look at an empty scene region when that region previously contained a task-relevant object [43].

Scene-schema knowledge

A second source of information that can guide gaze is 'scene-schema knowledge', generic semantic and spatial knowledge about a particular type of scene [44–46]. Schema knowledge includes information about the objects likely to be found in a specific category of scene (e.g. kitchens typically contain stoves), and spatial regularities associated with a scene category (e.g. staplers are typically found on desks), as well as generic world knowledge about scenes (e.g. staplers do not float in the air). Scene identity can be apprehended and a scene schema retrieved very rapidly [47], and schema knowledge can then be used to limit initial fixations to scene areas likely to contain an object relevant to the current task [23,35]. An interesting issue concerns the function that fixation serves given the rapidity with which scenes are understood (Box 2).

Task-related knowledge

A third type of knowledge used in human gaze control is 'task-related knowledge'. Task-related knowledge can involve a general 'gaze-control policy' or strategy relevant to a given task, such as periodically fixating the reflection in the rear-view mirror while driving, and moment-to-moment control decisions based on ongoing perceptual and cognitive needs. Task-related knowledge might also produce specific sequences of fixations, although the evidence for such scan patterns is currently weak (Box 3). The distribution of fixations over a given scene changes depending on whether a viewer is searching for an object or trying to memorize that scene [35]. Gaze control differs during complex and well-learned activities such as reading [10], tea and sandwich making [36], and driving [48]. Gaze control is also strongly influenced by moment-to-moment cognitive processes related to spoken language comprehension and production [12,43]. In most cases, the durations as well as positions of individual fixations are influenced by top-down factors, and an important challenge for any complete model of gaze control will be to account for this variation in fixation durations (Box 4).

In summary, humans use knowledge about the world to guide gaze intelligently through a scene. Cognitive systems interact with each other and with the scene image to determine where the eyes fixate and how long they remain fixated at a particular location. Evolution has clearly favored active vision, and understanding why is of central concern in the study of natural scene perception.

Acknowledgements

Preparation of this article was supported by grants from the National Science Foundation (BCS-0094433) and the Army Research Office (DAAD19–00–1-0519; the opinions expressed in this article are those of the authors and do not necessarily represent the views of the Department of the Army or any other governmental organization). I thank Monica Castelhano, Daniel Gajewski, Aaron Pearson, Aude Oliva, and Fernanda Ferreira for their contributions to the ideas presented here and for comments on the article. I also thank Aude Oliva her help with the development of the tools used to produce the Figure in Box 4, Ben Hsieh and Mike Mack for their help with figure graphics, and Dan Gajewski for Figure 2.

Box 4. Fixation duration

Two important aspects of gaze control during scene perception are where fixations tend to be directed (fixation position), and how long they typically remain there (fixation duration). The influence of visual and cognitive factors on fixation duration is widely acknowledged in the human gaze control literature and has been explicitly incorporated in computational models of reading [66], but has generally been overlooked in the computational literature on gaze control in scenes. Conclusions about the distribution of attention over a scene can differ markedly when fixation position is weighted by fixation duration because the distribution of processing time across a scene is a function of both the spatial distribution of fixations and the durations of those fixations (Figure I).

Average fixation duration during scene viewing is ~330 ms, although there is a good deal of variability around this mean both within an individual and across individuals [10,67] (see Figure 1 in main text). Much of this variability is controlled by visual and cognitive factors associated with the currently fixated scene region. For example, individual fixation duration (the duration of each discrete fixation) is affected by scene luminance [68] and contrast [69]. Mean individual fixation duration is also longer for full color photographs than black-and-white line drawings, although the distributions are very similar [67]. Van Diepen and colleagues used a moving mask tied to fixation position and updated in real time to manipulate the visual quality of the scene available at fixation independently of the quality available extra-foveally [34,70]. Individual fixation durations in a scene were elevated when the image at fixation was reduced by contrast or partially obscured by a noise mask, suggesting that fixation duration is influenced by the acquisition of visual information from the currently fixated region. Individual fixation durations are also influenced by viewing task, with longer fixation durations during scene memorization than search [35].

Stimulus and task effects can affect molar measures of fixation time even when individual fixation durations are unchanged. For example, first-pass gaze duration (the sum of all fixations in a region from first entry to exit) on an object in a scene is increased by a visual change to that object, even when the viewer does not notice the change [41,42,57,71,72]. First-pass gaze durations are also influenced by object and scene semantics, with longer gaze durations on semantically informative (i.e. less consistent) than uninformative (i.e. more consistent) objects [35,39,45,57,73].

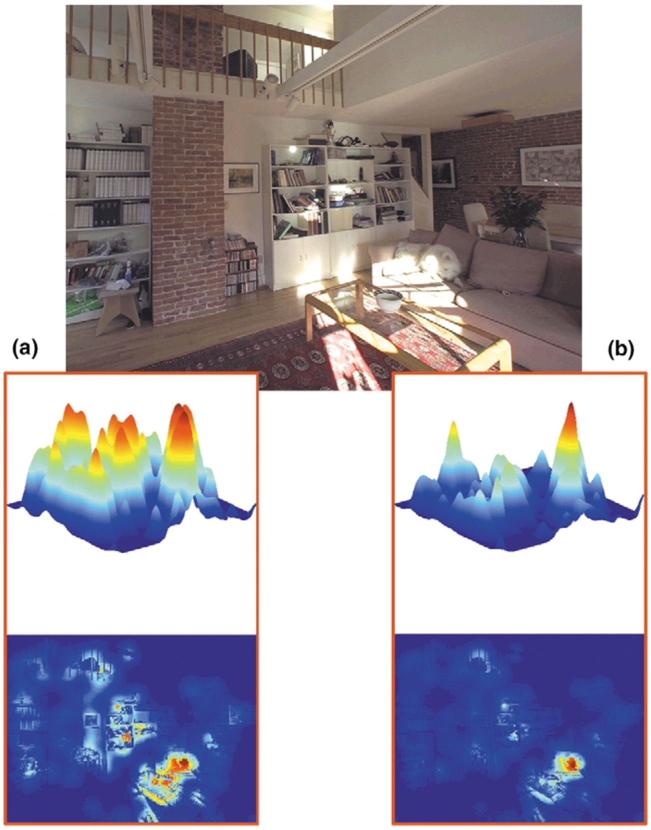

Figure I. Fixation landscape over a scene. Representation of the positions of all fixations by all viewers on a scene **(a)**, and the same fixations weighted by fixation duration **(b)**. These landscapes were created by placing a Gaussian with a diameter of 2 degrees of visual angle (equivalent to the fovea) centered at each fixation point, summing the Gaussians, and normalizing the height of the resulting sums [74] For the duration-weighted landscape, the height of each Gaussian was proportional to the duration of that fixation in milliseconds. Comparison of the unweighted and duration-weighted landscapes illustrates that although fixations are distributed over a good deal of a scene, the majority of fixation time is concentrated on specific objects. The duration-weighted landscape can be interpreted as an 'attentional landscape' associated with scene interpretation. In this scene, fixation time is concentrated on the interesting object: the dog asleep on the couch.

References

1 Matin, E. (1974) Saccadic suppression: a review and an analysis. *Psychol. Bull.* 81, 899–917
2 Thiele, A. *et al.* (2002) Neural mechanisms of saccadic suppression. *Science* 295, 2460–2462
3 Volkmann, F.C. (1986) Human visual suppression. *Vision Res.* 26, 1401–1416
4 Land, M.F. (1999) Motion and vision: why animals move their eyes. *J. Comp. Physiol. Ser. A* 185, 341–352
5 Ballard, D.H. *et al.* (1997) Deictic codes for the embodiment of cognition. *Behav. Brain Sci.* 20, 723–767
6 Churchland, P.S. *et al.* (1994) A critique of pure vision. In *Large Scale Neuronal Theories of the Brain* (Koch, C. and Davis, S., eds), pp. 23–60, MIT Press
7 Luck, S.J. and Vecera, S.P. (2002) Attention. In *Steven's Handbook of Experimental Psychology*, 3rd edn, (Pashler, H. and Yantis, S., eds), pp. 235–286
8 Findlay, J.M. (2004) Eye scanning and visual search. In *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (Henderson, J.M. and Ferreira, F., eds), Psychology Press
9 Liversedge, S.P. and Findlay, J.M. (2000) Saccadic eye movements and cognition. *Trends Cogn. Sci.* 4, 6–14
10 Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422
11 Sereno, S. and Rayner, K. (2003) Measuring word recognition in reading: eye movements and event-related potentials. *Trends Cogn. Sci.* 7, 489–493
12 Tanenhaus, M.K. *et al.* (1995) Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634
13 Henderson, J.M. and Ferreira, F. (2004) Scene perception for psycholinguists. In *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (Henderson, J.M. and Ferreira, F., eds), Psychology Press
14 Mannan, S.K. *et al.* (1996) The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spat. Vis.* 10, 165–188
15 Mannan, S.K. *et al.* (1997) Fixation patterns made during brief examination of two-dimensional images. *Perception* 26, 1059–1072
16 Krieger, G. *et al.* (2000) Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spat. Vis.* 13, 201–214
17 Parkhurst, D.J. and Niebur, E. (2003) Scene content selected by active vision. *Spat. Vis.* 16, 125–154
18 Reinagel, P. and Zador, A.M. (1999) Natural scene statistics at the centre of gaze. *Network* 10, 341–350
19 Itti, L. and Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506
20 Itti, L. and Koch, C. (2001) Computational modeling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203
21 Koch, C. and Ullman, S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227
22 Torralba, A. (2003) Modeling global scene factors in attention. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1407–1418
23 Oliva, A. *et al.* (2003) Top-down control of visual attention in object detection. *IEEE Proceedings of the International Conference on Image Processing* (Vol. I), IEEE, pp. 253–256
24 Parkhurst, D. *et al.* (2002) Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* 42, 107–123
25 Henderson, J.M. and Hollingworth, A. (2003) Global transsaccadic change blindness during scene perception. *Psychol. Sci.* 14, 493–497
26 Irwin, D.E. (1992) Visual memory within and across fixations. In *Eye Movements and Visual Cognition: Scene Perception and Reading* (Rayner, K., ed.), pp. 146–165, Springer-Verlag
27 Posner, M.I. *et al.* (1985) Inhibition of return: neural basis and function. *Cogn. Neuropsychol.* 2, 211–228
28 Leek, E.C. *et al.* (2003) Inhibition of return for objects and locations in static displays. *Percept. Psychophys.* 65, 388–395
29 Rao, R.P.N. *et al.* (2002) Eye movements in iconic visual search. *Vision Res.* 42, 1447–1463
30 Li, Z. (2002) A saliency map in primary visual cortex. *Trends Cogn. Sci.* 6, 9–16
31 Corbetta, M. and Shulman, G.L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215
32 Findlay, J.M. and Walker, R. (1999) A model of saccade generation based on parallel processing and competitive inhibition. *Behav. Brain Sci.* 22, 661–721
33 Gottlieb, J.P. *et al.* (1998) The representation of salience in monkey parietal cortex. *Nature* 391, 481–484
34 van Diepen, P.M.J. *et al.* (1998) Functional division of the visual field: moving masks and moving windows. In *Eye Guidance in Reading and Scene Perception* (Underwood, G., ed.), Elsevier
35 Henderson, J.M. *et al.* (1999) Effects of semantic consistency on eye movements during scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 210–228
36 Land, M.F. and Hayhoe, M. (2001) In what ways do eye movements contribute to everyday activities? *Vision Res.* 41, 3559–3565
37 Turano, K.A. *et al.* (2003) Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Res.* 43, 333–346
38 Buswell, G.T. (1935) *How People Look at Pictures*, University of Chicago Press
39 Loftus, G.R. and Mackworth, N.H. (1978) Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 565–572
40 Yarbus, A.L. (1967) *Eye Movements and Vision*, Plenum Press
41 Henderson, J.M. and Hollingworth, A. (2003) Eye movements, visual memory, and scene representation. In *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes* (Peterson, M. and Rhodes, G., eds), Oxford University Press
42 Hollingworth, A. and Henderson, J.M. (2002) Accurate visual memory for previously attended objects in natural scenes. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 113–136
43 Henderson, J.M., Ferreira, F. eds, (2004) *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, Psychology Press
44 Biederman, I. *et al.* (1982) Scene perception: detecting and judging objects undergoing relational violations. *Cognit. Psychol.* 14, 143–177
45 Friedman, A. (1979) Framing pictures: the role of knowledge in automatized encoding and memory for gist. *J. Exp. Psychol. Gen.* 108, 316–355
46 Mandler, J.M. and Johnson, N.S. (1976) Some of the thousand words a picture is worth. *J. Exp. Psychol. [Hum Learn]* 2, 529–540
47 Schyns, P. and Oliva, A. (1994) From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychol. Sci.* 5, 195–200
48 Land, M.F. and Lee, D.N. (1994) Where we look when we steer. *Nature* 369, 742–744
49 Potter, M.C. (1999) Understanding sentences and scenes: the role of conceptual short-term memory. In *Fleeting Memories* (Coltheart, V., ed.), pp. 13–46, MIT Press
50 Thorpe, S.J. *et al.* (1996) Speed of processing in the human visual system. *Nature* 381, 520–522
51 Li, F-F. *et al.* (2002) Natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9596–9601
52 Oliva, A. and Torralba, A. (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42, 145–175
53 Torralba, A. and Oliva, A. (2003) Statistics of natural image categories. *Network* 14, 391–412
54 Henderson, J.M. *et al.* (2003) Eye movements and picture processing during recognition. *Percept. Psychophys.* 65, 725–734
55 Henderson, J.M. and Hollingworth, A. (1999) The role of fixation position in detecting scene changes across saccades. *Psychol. Sci.* 10, 438–443
56 Hollingworth, A. *et al.* (2001) Change detection in the flicker paradigm: the role of fixation position within the scene. *Mem. Cognit.* 29, 296–304
57 Hollingworth, A. *et al.* (2001) To see and remember: visually specific information is retained in memory from previously attended objects in natural scenes. *Psychon. Bull. Rev.* 8, 761–768

58 Ballard, D.H. *et al.* (1995) Memory representations in natural tasks. *J. Cogn. Neurosci.* 7, 66–80

59 Nelson, W.W. and Loftus, G.R. (1980) The functional visual field during picture viewing. *J. Exp. Psychol. [Hum Learn]* 6, 391–399

60 Brooks, R. and Meltzoff, A.N. (2002) The importance of eyes: how infants interpret adult looking behavior. *Dev. Psychol.* 38, 958–966

61 Macrae, C.N. *et al.* (2002) Are you looking at me? Eye gaze and perception. *Psychol. Sci.* 13, 460–464

62 Noton, D. and Stark, L. (1971) Scan paths in eye movements during pattern perception. *Science* 171, 308–311

63 Noton, D. and Stark, L. (1971) Scan paths in saccadic eye movements while viewing and recognizing patterns. *Vision Res.* 11, 929–944

64 Groner, R. *et al.* (1984) Looking at face: local and global aspects of scanpths. In *Theoretical and Applied Aspects of Eye Movement Research* (Gale, A.G. and Johnson, F., eds), pp. 523–533, Elsevier

65 Mannan, S.K. *et al.* (1997) Fixation sequences made during visual examination of briefly presented 2D images. *Spat. Vis.* 11, 157–178

66 Reichle, E.D. *et al.* (1998) Toward a model of eye movement control in reading. *Psychol. Rev.* 105, 125–157

67 Henderson, J.M. and Hollingworth, A. (1998) Eye movements during scene viewing: an overview. In *Eye Guidance in Reading and Scene Perception* (Underwood, G., ed.), pp. 269–283, Elsevier

68 Loftus, G.R. (1985) Picture perception: effects of luminance on available information and information-extraction rate. *J. Exp. Psychol. Gen.* 114, 342–356

69 Loftus, G.R. *et al.* (1992) Effects of visual degradation on eye-fixation durations, perceptual processing, and long-term visual memory. In *Eye Movements and Visual Cognition: Scene Perception and Reading* (Rayner, K., ed.), pp. 203–226, Springer

70 van Diepen, P.M.J. (1995) Chronometry of foveal information extraction during scene perception. In *Eye Movement Research: Mechanisms, Processes and Applications* (Findlay, J.M. *et al.*, eds), pp. 349–362, Elsevier

71 Henderson, J.M. and Hollingworth, A. (2003) Eye movements and visual memory: detecting changes to saccade targets in scenes. *Percept. Psychophys.* 65, 58–71

72 Hayhoe, M.M. *et al.* (1998) Task constraints in visual working memory. *Vision Res.* 38, 125–137

73 De Graef, P. *et al.* (1990) Perceptual effects of scene context on object identification. *Psychol. Res.* 52, 317–329

74 Pomplun, M. *et al.* (1996) Disambiguating complex visual information: towards communication of personal views of scenes. *Perception* 25, 931–948