CellPress

Opinion

Gaze Control as Prediction

John M. Henderson[1,*]

The recent study of overt attention during complex scene viewing has emphasized explaining gaze behavior in terms of image properties and image salience independently of the viewer's intentions and understanding of the scene. In this Opinion article, I outline an alternative approach proposing that gaze control in natural scenes can be characterized as the result of knowledge-driven prediction. This view provides a theoretical context for integrating and unifying many of the disparate phenomena observed in active scene viewing, offers the potential for integrating the behavioral study of gaze with the neurobiological study of eye movements, and provides a theoretical framework for bridging gaze control and other related areas of perception and cognition at both computational and neurobiological levels of analysis.

Trends

Gaze is directed to task- and goal-relevant scene regions.

Gaze control is based on predictions concerning where specific goal- and task-relevant objects are likely to be found.

Predictions for gaze control are based on knowledge gained from past experience with scenes.

Predictions are likely to draw on memory representations of specific scene instances and general scene categories.

Let's Take a Look

Some 85 years ago, Guy Thomas Buswell established that viewers tend to look at the regions of scenes that are likely to contain the information that is most meaningful and relevant [1] (Figure 1, Key Figure). Similarly, in classic textbook demonstrations Alfred Yarbus showed that where we look in a complex scene is strongly influenced by our current goals and viewing task [2]. In the years since those groundbreaking studies, the role of knowledge in the control of gaze and attention in complex, meaningful scenes has been repeatedly shown. Despite this overwhelming evidence, the literature over the past couple of decades has focused almost exclusively on trying to explain where people look in scenes in terms of image properties alone, ignoring the viewer's understanding of the scene. This tendency to focus on the image rather than the viewer's understanding of the meaning of the scene is likely to be due in part to a version of the principle of the drunkard's search, in which the inebriated driver looks for lost car keys under a streetlamp because that is where the light is. In the case of attention in scenes, image properties fall under the light of the streetlamp; it is far easier to build models that account for where we attend based on image properties than it is to build models based on scene meaning and viewer goals. A model based on image properties requires methods to measure and quantify those properties, a condition that is already within grasping distance given advances in computational vision and visual neuroscience. By contrast, a model based on a full understanding of the scene, its meaning, and its relationship to the viewer's current goals and task requires a relatively complete model of human cognition, and that is still a few years off. From this perspective it makes sense that investigators have tended to focus on the tractable and put off the less tractable for another day.

Clearly, however, the image- or saliency-based approach is fundamentally limited when it comes to accounting for how human viewers direct their attention through real scenes [3–7]. The question then becomes how to make theoretical progress in a way that takes viewer knowledge into account. In recent years a new framework for understanding how the brain thinks and perceives has emerged. This framework conceptualizes the brain as a 'prediction machine' [8], a system that uses past experience to generate expectations or predictions concerning what and where items and events are likely to be encountered next [9–11]. When these predictions are supported by incoming evidence from the environment, all is well and nothing new is learned. When a prediction is partially or completely wrong, however, the knowledge that was used to

[1]Center for Mind and Brain and Department of Psychology, University of California, Davis, 267 Cousteau Place, Davis, CA, USA

*Correspondence: johnhenderson@ucdavis.edu (J.M. Henderson).

generate that prediction has to be updated so future predictions can be more accurate. The upshot is that errors of prediction lead to knowledge gain. For some proposals, only information associated with prediction error is coded from the environment, a process called predictive coding [12–14].

In this Opinion article, I outline the proposal that **gaze control** (see Glossary) in natural scenes can be understood as the consequence of spatial prediction. Specifically, the proposal is that where a viewer looks and attends in a complex scene is the result of a prediction about where the most meaningful and task-relevant information is to be found in that scene. Adopting this view provides an overarching theoretical context for unifying many of the disparate phenomena observed in active scene viewing. For example, it offers potential for integrating the behavioral study of gaze control with the neurobiological study of eye movements more generally [15]. It also provides a theoretical framework that can assist in bridging between gaze control and other related areas of perception, cognition, and motor control. It provides a potential framework for understanding how **overt** and **covert attention** are related. Furthermore, taking this perspective provides potential for integrating the study of language–vision interaction, where linguistic input can determine where to look [16,17] and where we look can influence what we say [18,19].

Here the emphasis is specifically on considering gaze control in terms of spatial prediction, since most research on overt attention in scenes has focused on understanding where people look [3,20–22]. However, a more encompassing theoretical framework would extend this approach to other aspects of gaze control as well, such as explaining the amount of time each scene region is **fixated** as a consequence of predictions about location, identity, and meaning (Box 1).

The Prediction Approach to Gaze Control: Examples

Four brief examples serve to illustrate the prediction approach to gaze control and the nature of the phenomena that support it. These examples are meant to be illustrative rather than exhaustive.

Object Search in Scenes

People can find common objects very quickly in complex real-world scenes. In experiments demonstrating this ability, target objects are placed at either an expected location or an unexpected location in each scene [23–28]. Viewers are asked to find those objects as quickly as possible. A typical experiment presents the name (or picture) of the target object for the current trial followed by the scene, which may or may not contain the object. The key finding is that viewers typically find objects effortlessly, often within one or two eye movements [23,24,29–31]. Indeed, viewers are so good at this that it is often not possible to study learning or repetition effects for real object search because search performance is already at ceiling [32]. Furthermore, viewers can rapidly find objects in scenes from a brief scene glimpse based on extraction of the 'gist' of the scene and can quickly find objects based on that gist even when the scene is no longer visible [33–35]. Viewers can also quickly find an object they are searching for when that object is not at all visually salient as long as the location of the searched-for object is constrained by the scene's meaning and structure [5,30]. For example, a coffee cup half-hidden by a box of Wheaties will still be fixated very quickly if it is in its expected location on the kitchen table [4]. How can we account for the strong influence of scene context on object search? The proposal here is that viewers use learned knowledge about where a given object is likely to be found from past experience with a given scene category to predict (given a new instance of that scene category) the location of the target, and this prediction is used to direct gaze.

Scene-Based Contextual Cueing

Because object search in scenes is so efficient, it can be difficult to study the learning processes associated with establishing the contextual relationships that underlie spatial prediction [32]. To

Glossary

Covert attention: selective processing of a location or object at the expense of other locations or objects via internal processes without a corresponding movement of the eyes.

Fixation: stable eye position directed toward a specific scene location or object used for visual information acquisition.

Fovea: the highest-resolution region of the retina.

Gaze control: the process of directing fixation through a scene in real time in the service of ongoing perceptual, cognitive, and behavioral activity.

Overt attention: selective processing of a location or object at the expense of other locations or objects due to a movement of the eyes.

Saccade: a fast, ballistic eye movement that reorients fixation from one location to another within a scene.

Key Figure

Viewers Direct Their Gaze to Meaningful and Informative Scene Regions.

Figure 1. (A) Example viewer's first six fixations (red dots) while searching for coffee cups. Red lines represent saccades and the black dot represents the initial fixation position at scene onset. (B) Example of first fixations for twelve viewers searching for paintings (red dots) and coffee cups (blue dots). In both cases, fixations are placed in the scene based on predictions concerning where the search targets should be found. Figure by Henderson, 2016; available at https://dx.doi.org/10.6084/m9.figshare.4245446.v1 under a CC-BY4.0 license.

overcome this challenge, scene-based contextual cueing provides an experimental method in which viewers learn new scene–object spatial constraints across search trials. For example, in a typical scene-based contextual cueing study viewers are presented with arbitrary object–scene pairs such as letters embedded in photographs [36–40]. The arbitrary nature of the pairings

Box 1. Looking Time

Overt attention in a scene involves both guiding gaze to the appropriate location and holding attention in that location for the appropriate amount of time given current processing demands [73,74]. The amount of time that gaze is directed to each scene region is highly sensitive to the difficulty of understanding the information in the scene [75–78]. A particularly strong effect of processing difficulty on fixation time is produced by the predictability of the currently fixated object in the context of its scene: unpredictable objects are fixated for more time than predictable objects [23,24,79]. In a typical experiment demonstrating this effect, objects are manipulated so that they are consistent or inconsistent with the overall meaning of the scene in some way. For example, inconsistency might be created by placing an object in a contextually inappropriate scene (e.g., placing a live chicken in a bedroom) or by violating physical constraints like support (e.g., placing a chicken so that it is floating in mid-air). Viewers are asked to view these scenes while they perform a task. For example, they might be asked to view scenes to find particular objects, to remember the scenes for later, or simply to decide how much they like each scene. Fixation time is then measured on objects when they are contextually consistent or inconsistent. The main result, replicated across tasks and studies, is that consistent objects are fixated for less time than inconsistent objects.

I propose that this finding can again be seen as a consequence of prediction on gaze control. Viewers use learned knowledge about the objects that are found in a given scene category to predict which objects will be encountered. These predictions then influence gaze, with prediction violations leading to an increase in the amount of time overt attention is directed to the current object. This increased time allows knowledge representations to be updated to take into account the unexpected information.

supports the investigation of questions concerning how spatial constraints are learned, the time course of learning, and the nature of the information that is learned. Concretely, in a typical scene-based contextual cueing experiment, viewers search for a target letter that has been embedded in each scene. When they find a target, they identify it with a button press. A trial sequence is divided into blocks. Learning blocks are followed by test blocks. Two types of trial are presented in each block. A novel trial shows a scene that the participant has not previously seen in the experiment. A repeated trial presents a scene that the participant has seen before. Critically, the target's location in each repeated scene is fixed so that the viewer can learn the spatial constraints for each scene–target pairing over repetitions. The target's identity is randomly selected with each repetition so that the viewer cannot simply learn pairings of scenes and target identities but instead must learn the locations of the targets. (The experimenter can also check that the target has been found by analyzing the viewer's fixation location at the time of response.) Each block randomly intermixes repeated scenes with novel scenes. Typically, viewers are not told about the block structure or the repetition of scenes and are left to pick up the spatial regularities implicitly, although in scene-based contextual cueing they are usually aware of the relationships [38,40]. In the test block, search performance is compared in repeated and novel scenes. (Unrepeated scenes in other learning blocks also provide a control for assessing baseline search time.) The critical finding is that, under these conditions, viewers learn the target locations very quickly, with five to seven repetitions typically sufficient to reach asymptotic search performance. Eye-tracking data show that the contextual cueing effect reflects a change in gaze behavior. After learning, fixation is directed very quickly and efficiently to the learned target location, often in one or two eye movements. The brain uses learned spatial regularities in the environment to predict where the target will be found and eye movements are based on this prediction.

Dynamic, Real-World Events

So far we have considered static images. What about moving images that depict ongoing dynamic events that unfold over extended time? It turns out that when watching dynamic images depicting real events, viewers are just as efficient at directing their gaze to relevant locations as they are when they view static images despite dynamic images being continuously updated. For example, in a set of classic studies using an early portable eye tracker, cricket batters were found to direct their attention to the location on the ground where the ball will bounce on the way to the batsman [41]. Interestingly, batters direct their gaze to this bounce point even before the ball reaches that point [41,42]. The nature of the bounce is important in determining the trajectory that the ball will take to the batter and attending to this point is therefore critical for good performance. So, to take best advantage, batters shift their gaze to the bounce point based on a prediction about where the ball will bounce even before it arrives so that they can predict where the ball will be when it reaches them. This type of prediction is not restricted to cricket [43,44]. Figure 2 shows an example from my laboratory, this time from tennis. In this study, part of the Dynamic Images and Eye Movements project [45], viewers watched video of a tennis match while their eye movements were recorded. We then analyzed the probability that viewers fixated the bounce point of the serve before the ball arrived. As can be seen in Figure 2, there was a strong tendency for viewers to predict the bounce point and to use this prediction to drive attention and gaze. Notice that this predictive act cannot be accounted for by image features or visual salience, because the bounce point was not defined by differences in image features or salience and the exact bounce point changed from serve to serve. Interestingly, as with cricket, viewers often looked at the bounce point tens of milliseconds before the ball arrived. Analysis of this behavior suggests that viewers initially watched the motion of the server and then shifted gaze via a **saccade** to the bounce point as the ball left the racket. Attention was guided by a prediction about the location of the bounce point generated from the timing and motion of the serve.

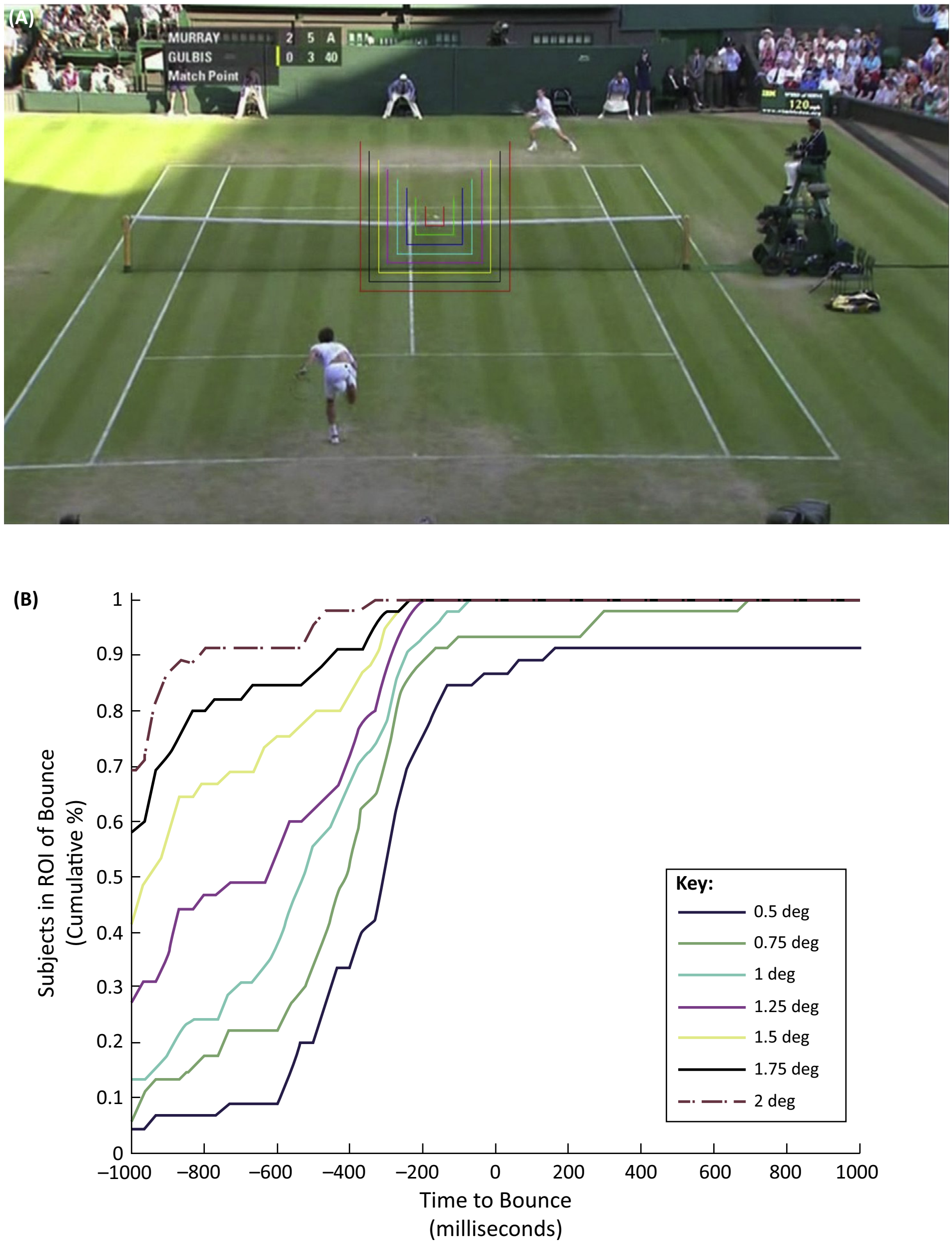

Figure 2. Viewers Direct Their Gaze to Meaningful Scene Regions in Dynamic Events. When watching a tennis match, viewers predict the bounce point and use this prediction to direct fixation. (A) shows a frame from the video with the scoring regions superimposed. (B) shows the cumulative percentage of viewers looking in the region around the bounce point before the ball arrives (negative time points) and after the ball arrives (positive time points) for regions of varying size around the bounce point. Viewers start looking to the bounce point before the ball arrives, often generating a saccade from the server's racket to the bounce point as the ball leaves the racket. Figure by Henderson, 2016; available at https://dx.doi.org/10.6084/m9.figshare.4245440.v1 under a CC-BY4.0 license.

Self-Generated Real-World Events

Another type of dynamic image relates to actions in the context of self-generated events. Similar predictive gaze behavior is observed in these dynamic events. For example, when making a peanut-butter-and-jelly sandwich, viewers shift their gaze to objects in the local environment just as those objects are needed for the task. That is, viewers look at the loaf of bread before picking it up to retrieve two slices, look at the top of the peanut butter jar before unscrewing the lid, look at the knife before picking it up, look at the jelly jar opening before scooping up a knife-full, and so on [46,47]. In each case, the looking action is based on current task needs and a prediction about where the required item is to be found [47–50]. Here, however, the predictions are generated from an action schema for making a sandwich. Interestingly, it appears that viewers initially spend some time scanning the environment to find the locations of all of the needed items before they start their task [47]. This behavior appears to be designed to establish a representation of the scene that the action schema can operate over in generating spatial predictions. In a sense, the quick look-ahead scan sets up an episodic representation of the current scene. This episodic scene representation can then provide a foundation for precise spatial prediction in the same way that scene gist can be used as the basis for probabilistic spatial prediction given a conceptual class of scenes [4,31,33]. In other words, episodic representations of specific scenes support predictions about specific instances (e.g., in this particular kitchen a coffee cup can be found exactly at this location) and conceptual representations of scene classes support predictions about those classes (in kitchens generally, coffee cups can typically be found at these locations), with both types of representation supporting predictions that can drive where viewers look.

What Do We Gain by Considering Gaze Control as Prediction?

Thinking about gaze control as prediction offers several advantages over other conceptualizations. First, it allows integration with other areas of cognitive science and neuroscience that consider the fundamental function of brains and their cognitive systems to be the generation of predictions [8,10,51]. This perspective has recently been adopted in many areas of cognitive science and cognitive neuroscience, including learning [52], memory [53–55], reward in motor planning [56,57], object and event perception [58,59], language processing [60–64], high-level scene perception [65], and reading [66,67] (Box 2). Taking a prediction approach to attentional guidance in scenes therefore provides potential leverage for identifying general principles of brain and cognition.

Second, thinking about gaze control as prediction may allow a more integrative approach to understanding the cognitive and neural systems supporting both overt (i.e., gaze) and covert (i.e., internal) spatial attention. For example, to the extent that covert attention plays a similar functional role in vision and draws on similar cognitive and neural systems for control as overt attention, it is likely that much of the description concerning the deployment of covert attention can be embraced by the prediction approach [68]. Reframing covert attention as prediction is illustrated by considering the classic Posner cueing task. In this task a cue directs covert attention to one or the other side of a display area. A target then appears on the cued or uncued

Box 2. Reading

It is well known that gaze moves through text in a directed manner during reading [80,81]. Gaze primarily travels in one direction (e.g., left to right for reading English). Most words are fixated, but some words are skipped and others are fixated multiple times. Sometimes gaze moves backward so that previously read text can be re-read. Can eye movements during reading be understood as a consequence of prediction concerning where useful information is to be found given the current level of comprehension? From the prediction perspective, forward movements are due to the expectation that each new word contains the next incrementally most important syntactic and semantic information in the text [82,83]. Occasionally, however, the next word is constrained enough that there is little to be gained from fixating it, and in that case the word can be skipped. Gaze is then sent to the location (e.g., two words downstream) predicted to contain the most useful information. Similarly, when comprehension is very difficult or fails altogether, attention is sent back into an earlier region of the text. The region that is fixated is the region that is predicted to contain the most useful information for unraveling the current comprehension difficulty [81,84]. In this sense, sending the eyes to a particular region of text (the next word, two words ahead, or five words back) can be thought of as a consequence of prediction concerning where the most useful information for understanding the text is to be found at each point in time.

side with some probability (e.g., 80% likely on the cued side, 20% likely on the uncued side). The participant's task is simply to press a button to indicate that the target has appeared. The definitive finding is that response time is faster when the target appears at the cued than at the uncued location, suggesting that attention has been directed to that location before the target's appearance. One simple extension of the current proposal is that covert visual attention is the internal manifestation of spatial prediction in the same way that overt attention (gaze) is the external manifestation of prediction [69]. That is, we might say that the brain has predicted the appearance of the target at a particular location, and we refer to this predictive brain act as covert attention.

At the neurobiological level of analysis, it is well known that many of the systems that are involved in the neural control of eye movements are also involved in the control of covert visual attention [70]. For example, neuroimaging studies in humans have demonstrated a network of brain regions that appear to greatly overlap for overt and covert attention. One previous account of this overlap derives from the premotor theory of attention, which posits that covert attention is due to overt eye movement programming [71,72]. An alternative hypothesis consistent with the prediction proposal is that the overlapping systems for overt and covert attention are associated with prediction, and particularly with spatial prediction. In this view, eye movements and covert visual attention are external and internal manifestations of the same spatial prediction system. Rather than considering covert spatial attention as parasitic on overt attention, as in premotor theory, this perspective leads to the idea that overt and covert attention are two ways for the brain to capitalize on spatial prediction. That is, spatial prediction is the fundamental process, manifest in both changes in perceptual weighting (covert attention) and the behavioral orienting of the eyes (overt attention). Importantly, in this model the same underlying predictions drive both types of attention. In summary, given that both overt and covert attention involve preselecting scene locations for focused preferential processing, it seems reasonable (and is more parsimonious to suppose) that this selection is the result of the same underlying predictive process.

Outstanding Questions

How precise are the spatial predictions that are used to guide gaze?

Are the predictions that guide gaze used strategically or automatically?

Can the prediction approach to gaze control in natural scenes be extended to gaze control in tasks such as face perception, reading (Box 2), and spatial navigation?

Can the prediction approach to gaze control be extended to account for both where fixation is directed and how long fixation is directed there (Box 1)?

How are the memory structures that generate predictions for gaze control encoded, stored, retrieved, and deployed?

Concluding Remarks

An approach to the brain that focuses on prediction has provided new insights into many cognitive domains. A natural extension of this approach is to the study of how we orient our attention over complex natural scenes. Attentional orienting in natural viewing involves gaze control, directing the high-resolution **fovea** and its associated neural machinery to important and informationally rich regions of the visual world. A potentially fruitful way to think about gaze control is in terms of predictions concerning where important and informative information is to be found given the current needs of the cognitive system. Although the focus here has been on gaze control in scenes, the general predictive processing approach to overt attention may profitably be applied to any area of active vision in which attention plays a role, such as reading (Box 2), face perception, driving, and navigation (see Outstanding Questions).

Acknowledgments

The author thanks Andy Clark, Fernanda Ferreira, and Mary Hayhoe for comments on an earlier version of this Opinion article. He thanks Parag Mital for Figure 2 and the related analyses. Development of the ideas outlined here was supported by National Science Foundation grant BCS-1636586.

References

1. Buswell, G.T. (1935) *How People Look at Pictures,* University of Chicago Press
2. Yarbus, A.L. (1967) *Eye Movements and Vision,* Plenum Press
3. Tatler, B.W. *et al.* (2011) Eye guidance in natural vision: reinterpreting salience. *J. Vis.* 11, 5
4. Torralba, A. *et al.* (2006) Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113, 766–786
5. Henderson, J.M. *et al.* (2007) Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye Movements: A Window on Mind and Brain* (Van Gompel, R.P.G. *et al.*, eds), pp. 537–562, Elsevier
6. Rothkopf, C. *et al.* (2007) Task and context determine where you look. *J. Vis.* 7, 16.1–16.20
7. Einhäuser, W. *et al.* (2008) Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *J. Vis.* 8, 2.1–2.19

8. Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204
9. Lupyan, G. and Clark, A. (2015) Words and the world: predictive coding and the language–perception–cognition interface. *Curr. Dir. Psychol. Sci.* 24, 279–284
10. Friston, K. (2010) The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138
11. Bar, M. (2009) The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1235–1243
12. Rao, R.P.N. and Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretationof some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87
13. Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am.* 20, 1434–1448
14. Wacongne, C. *et al.* (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20754–20759
15. Kowler, E. *et al.* (2014) Davida Teller Award Lecture 2013: the importance of prediction and anticipation in the control of smooth pursuit eye movements. *J. Vis.* 14, 10
16. Altmann, G. and Kamide, Y. (1998) Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264
17. Lowder, M.W. and Ferreira, F. (2016) Prediction in the processing of repair disfluencies: evidence from the visual-world paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 1400–1416
18. Griffin, Z.M. and Bock, K. (1999) What the eyes say about speaking. *Psychol. Sci.* 11, 274–279
19. Gleitman, L.R. *et al.* (2007) On the give and take between event apprehension and utterance formulation. *J. Mem. Lang.* 57, 544–569
20. Henderson, J.M. (2007) Regarding scenes. *Curr. Dir. Psychol. Sci.* 16, 219–222
21. Itti, L. and Koch, C. (2001) Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 194–203
22. Borji, A. *et al.* (2014) Complementary effects of gaze direction and early saliency in guiding fixations during free viewing. *J. Vis.* 14, 3
23. Võ, M.L.H. and Henderson, J.M. (2009) Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *J. Vis.* 9, 24.1–24.15
24. Henderson, J.M. *et al.* (1999) The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 210–228
25. Hollingworth, A. and Henderson, J.M. (1998) Does consistent scene context facilitate object perception? *J. Exp. Psychol. Gen.* 127, 398–415
26. Hollingworth, A. and Henderson, J.M. (2000) Semantic informativeness mediates the detection of changes in natural scenes. *Vis. Cogn.* 7, 213–235
27. Chen, X. and Zelinsky, G.J. (2006) Real-world visual search is dominated by top-down guidance. *Vision Res.* 46, 4118–4133
28. Eckstein, M.P. *et al.* (2006) Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychol. Sci.* 17, 973–980
29. Malcolm, G.L. and Henderson, J.M. (2009) The effects of target template specificity on visual search in real-world scenes: evidence from eye movements. *J. Vis.* 9, 8.1–8.13
30. Henderson, J.M. *et al.* (2009) Searching in the dark: cognitive relevance drives attention in real-world scenes. *Psychon. Bull. Rev.* 16, 850–856
31. Malcolm, G.L. and Henderson, J.M. (2010) Combining top-down processes to guide eye movements during real-world scene search. *J. Vis.* 10, 4.1–4.11
32. Võ, M.L.H. and Wolfe, J.M. (2013) The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition* 126, 198–212
33. Castelhano, M.S. and Henderson, J.M. (2003) Flashing scenes and moving windows: an effect of initial scene gist on eye movements. *J. Vis.* 3, 67
34. Castelhano, M.S. and Henderson, J.M. (2007) Initial scene representations facilitate eye movement guidance in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 753–763
35. Võ, M.L.H. and Henderson, J.M. (2010) The time course of initial scene processing for eye movement guidance in natural scene search. *J. Vis.* 10, 14.1–14.13
36. Brockmole, J.R. *et al.* (2008) The role of meaning in contextual cueing: evidence from chess expertise. *Q. J. Exp. Psychol.* 61, 1886–1896
37. Brooks, D.I. *et al.* (2010) The nesting of search contexts within natural scenes: evidence from contextual cuing. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1406–1418
38. Brockmole, J.R. *et al.* (2006) Contextual cueing in naturalistic scenes: global and local contexts. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 699–706
39. Brockmole, J.R. and Henderson, J.M. (2006) Recognition and attention guidance during contextual cueing in real-world scenes: evidence from eye movements. *Q. J. Exp. Psychol.* 59, 1177–1187
40. Brockmole, J.R. and Henderson, J.M. (2006) Using real-world scenes as contextual cues for search. *Vis. Cogn.* 13, 99–108
41. Land, M.F. and McLeod, P. (2000) From eye movements to actions: how batsmen hit the ball. *Nat. Neurosci.* 3, 1340–1345
42. Mann, D.L. *et al.* (2013) The head tracks and gaze predicts: how the world's best batters hit a ball. *PLoS ONE* 8, e58289
43. Diaz, G. *et al.* (2013) Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *J. Vis.* 13, 20
44. Diaz, G. *et al.* (2013) Memory and prediction in natural gaze control. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20130064
45. Mital, P.K. *et al.* (2011) Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognit. Comput.* 3, 5–24
46. Land, M.F. *et al.* (1999) The roles of vision and eye movements in the control of activities of daily living. *Perception* 28, 1311–1328
47. Hayhoe, M.M. *et al.* (2003) Visual memory and motor planning in a natural task. *J. Vis.* 3, 49–63
48. Land, M.F. and Hayhoe, M.M. (2001) In what ways do eye movements contribute to everyday activities? *Vision Res.* 41, 3559–3565
49. Hayhoe, M.M. and Ballard, D. (2005) Eye movements in natural behavior. *Trends Cogn. Sci.* 9, 188–194
50. Hayhoe, M. and Ballard, D. (2014) Modeling task control of eye movements. *Curr. Biol.* 24, R622–R628
51. Friston, K. *et al.* (2012) Perceptions as hypotheses: saccades as experiments. *Front. Psychol.* 3, 1–20
52. Glimcher, P.W. (2011) Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc. Natl Acad. Sci. U.S.A.* 108 (Suppl.), 15647–15654
53. Ranganath, C. and Rainer, G. (2003) Cognitive neuroscience: neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* 4, 193–202
54. Kim, G. *et al.* (2014) Pruning of memories by context-based prediction error. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8997–9002
55. Henson, R.N. and Gagnepain, P. (2010) Predictive, interactive multiple memory systems. *Hippocampus* 20, 1315–1326
56. Trommershäuser, J. *et al.* (2008) Decision making, movement planning and statistical decision theory. *Trends Cogn. Sci.* 12, 291–297
57. Najemnik, J. and Geisler, W.S. (2005) Optimal eye movement strategies in visual search. *Nature* 434, 387–391
58. Chaumon, M. *et al.* (2014) Visual predictions in the orbitofrontal cortex rely on associative content. *Cereb. Cortex* 24, 2899–2907
59. Zhang, J. and Rowe, J.B. (2014) The neural signature of information regularity in temporally extended event sequences. *Neuroimage* 107, 266–276
60. Levy, R. (2008) Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177
61. Kuperberg, G.R. and Jaeger, T.F. (2015) What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 3798, 1–70
62. Hale, J. (2016) Information-theoretical complexity metrics. *Lang. Linguist. Compass* 10, 397–412
63. Traxler, M.J. (2014) Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. *Trends Cogn. Sci.* 18, 605–611

64. Pickering, M.J. and Garrod, S. (2007) Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* 11, 105–110
65. Itti, L. and Baldi, P. (2009) Bayesian surprise attracts human attention. *Vision Res.* 49, 1295–1306
66. Levy, R. *et al.* (2009) Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21086–21090
67. Henderson, J.M. *et al.* (2016) Language structure in the brain: a fixation-related fMRI study of syntactic surprisal in reading. *Neuroimage* 132, 293–300
68. Vossel, S. *et al.* (2014) Attention, predictions and expectations, and their violation: attentional control in the human brain. *Front. Hum. Neurosci.* 8, 3389
69. Summerfield, C. and de Lange, F.P. (2014) Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* 15, 745–756
70. Petersen, S.E. and Posner, M.I. (2012) The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* 35, 73–89
71. Thompson, K.G. *et al.* (2005) Neuronal basis of covert spatial attention in the frontal eye field. *J. Neurosci.* 25, 9479–9487
72. Casarotti, M.M. *et al.* (2012) Paying attention through eye movements: a computational investigation of the premotor theory of spatial attention. *J. Cogn. Neurosci.* 24, 1519–1531
73. Henderson, J.M. (2003) Human gaze control during real-world scene perception. *Trends Cogn. Sci.* 7, 498–504
74. Henderson, J.M. (2011) Eye movements and scene perception. In *The Oxford Handbook of Eye Movements* (Liversedge, S.P. *et al.*, eds), pp. 593–606, Oxford University Press
75. Nuthmann, A. *et al.* (2010) CRISP: a computational model of fixation durations in scene viewing. *Psychol. Rev.* 117, 382–405
76. Henderson, J.M. and Smith, T.J. (2009) How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Vis. Cogn.* 17, 1055–1082
77. Henderson, J.M. and Pierce, G.L. (2008) Eye movements during scene viewing: evidence for mixed control of fixation durations. *Psychon. Bull. Rev.* 15, 566–573
78. Henderson, J.M. *et al.* (2009) The influence of clutter on real-world scene search: evidence from search efficiency and eye movements. *J. Vis.* 9, 32.1–32.8
79. Võ, M.L.H. and Henderson, J.M. (2011) Object–scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Atten. Percept. Psychophys.* 73, 1742–1753
80. Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422
81. Clifton, C. *et al.* (2016) Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *J. Mem. Lang.* 86, 1–19
82. Reichle, E.D. *et al.* (1998) Toward a model of eye movement control in reading. *Psychol. Rev.* 105, 125–157
83. Reichle, E.D. *et al.* (2013) Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Dev. Rev.* 33, 110–149
84. Apel, J.K. *et al.* (2012) Targeting regressions: do readers pay attention to the left? *Psychon. Bull. Rev.* 19, 1108–1113