Eye Movements from Laboratory to Life

Benjamin W. Tatler

The manner in which we sample visual information from the world is constrained by the spatial and temporal sampling limits of the human eye. High acuity vision is restricted to the small central foveal region of the retina, which is limited to just a few degrees of visual angle in extent. Moreover, visual sampling is effectively limited to when the retinal image is relatively stabilised for periods of fixation (Erdmann and Dodge 1898), which last on average around 200–400 ms when viewing text, scenes or real environments (Land and Tatler 2009; Rayner 1998). It is clear from these severe spatiotemporal constraints on visual sampling that high acuity vision is a scarce resource and, like any scarce resource, it must be distributed carefully and appropriately for the current situation.

The selection priorities that underlie decisions about where to direct the eyes have interested researchers since eye movement research was in its infancy. While stimulus properties were shown to influence fixation behaviour (McAllister 1905), it was soon recognised that the relationship between the form of the patterns viewed and the eye movements of the observer was not as close as early researchers had expected (Stratton 1906). Moreover, the great variation in fixation patterns between individuals (McAllister 1905) made it clear that factors other than stimulus properties were likely to be involved in allocating foveal vision.

In light of evidence gathered from observers viewing the Müller-Lyer illusion (Judd 1905), Poggendorff illusion (Cameron and Steele 1905) and Zöllner illusion (Judd and Courten 1905), Judd came to the conclusion that "the actual movements executed are in no small sense responses to the verbal stimuli which the subject receives in the form of general directions. The subject reacts to the demands imposed upon him by the general situation… The whole motive for movement is therefore not to be sought in the figures themselves" (Judd, 1905, p. 216–217).

The relative importance of external factors relating to the stimulus properties and internal factors relating to goals of the observer became a prominent theme in eye movement research and continues to underlie many aspects of contemporary eye movement research. While early research in this domain used simple patterns and

B. W. Tatler (✉)
University of Dundee, Dundee, UK
e-mail: b.w.tatler@activevisionlab.org

M. Horsley et al. (eds.), *Current Trends in Eye Tracking Research,*
DOI 10.1007/978-3-319-02868-2_2, © Springer International Publishing Switzerland 2014

line illusions (due to technological limitations in display and recording devices), more recent research has considered how we view complex scenes in an attempt to produce an ecologically valid account of eye guidance.

1 Eye Guidance in Scene Viewing

When viewing complex scenes, fixations are allocated preferentially to certain locations, while other locations receive little or no scrutiny by foveal vision (Buswell 1935). Moreover, the regions selected for fixations are similar between individuals: different people select similar locations in scenes to allocate foveal vision to (Buswell 1935; Yarbus 1967). Such similarity in fixation behaviour implies common underlying selection priorities across observers. Buswell (1935) recognised that these common selection priorities are likely to reflect a combination of common guidance by low-level information in scenes and by high-level strategic factors. However, what external factors are involved in prioritising locations for fixation and the manner in which low- and high-level sources of information combine to produce fixation behaviour were not clear. Since Buswell's seminal work, a considerable body of evidence has been accumulated regarding these issues and there now exist computational models of scene viewing that propose particular low-level features as prominent in fixation allocation, and specific ways in which high-level sources of information may be combined with low-level image properties in order to decide where to fixate.

1.1 Low-Level Factors in Eye Guidance

From the extensive literature on how humans search arrays of targets, it is clear that basic visual features can guide attention (Wolfe 1998) and models based solely on low-level features can offer effective accounts of search behaviour (Treisman and Gelade 1980; Wolfe 2007). Koch and Ullman (1985) proposed an extension of these feature-based accounts of visual search to more complex scenes, and this was later implemented as a computational model (Itti and Koch 2000; Itti et al. 1998). In this model, low-level features are extracted in parallel across the viewed scene using a set of biologically plausible filters. Individual feature maps are combined across features and spatial scales via local competition in order to produce a single overall visual conspicuity map referred to as a salience map (see Fig. 1). In this account, attention is allocated to the location in the scenes that corresponds to the most salient location in the salience map. Once attended, the corresponding location in the salience map receives transient local inhibition, and attention is relocated to the next most salient location. Thus, attention is allocated serially to locations in the scene in order of most to least conspicuous in the salience map.

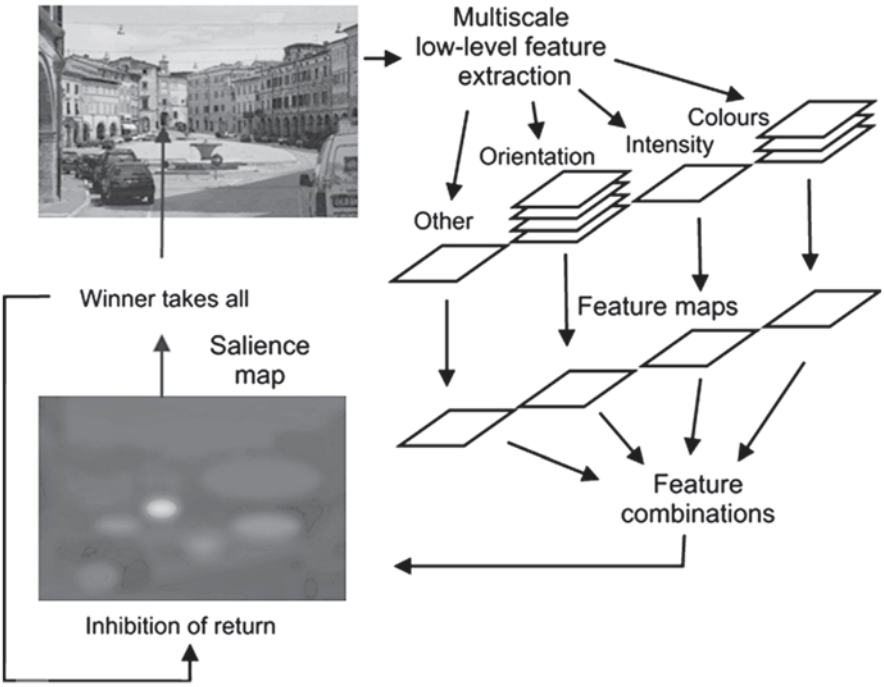

Fig. 1 Schematic of Itti and Koch's (2000) salience model, redrawn for Land and Tatler (2009)

The salience model replicates human search behaviour well when searching for feature singletons or conjunctions of two features (Itti and Koch 2000), and the extent to which it can explain attention allocation in more complex scenes has been the topic of a large volume of research. Most evaluations of the explanatory power of the salience model (and other similar models based on low-level feature-based attention allocation) use one of two approaches: measuring local image statistics at fixated locations (e.g. Reinagel and Zador 1999) or using the model to predict locations that should be fixated and seeing what proportion of human fixations fall within these predicted locations (e.g. Torralba et al. 2006). Both approaches seem to support a role for low-level information in fixation selection. Fixated locations have higher salience than control locations (e.g. Parkhurst et al. 2002), and more fixations are made within locations predicted by salience models than would be expected by chance (e.g. Foulsham and Underwood 2008). However, despite these apparently supportive results, the explanatory power of purely low-level models is limited: The magnitude of featural differences between fixated and control locations or how likely fixations are to fall within regions predicted by the models is typically small (Einhauser et al. 2008; Nyström and Holmqvist 2008; Tatler et al. 2005), suggesting that these models can only count for a limited fraction of fixation behaviour. Moreover, these basic results that appear to support low-level models must be interpreted

with caution. Correlations between low-level features and fixation selection may arise because of correlations between low-level features in scenes and higher-level scene content rather than because of a causal link between low-level properties and eye guidance (Henderson 2003; Henderson et al. 2007; Tatler 2007).

1.2 Higher-Level Factors in Eye Guidance

Low-level conspicuity tends to correlate with higher-level scene structure: Salient locations typically fall within objects in scenes (Elazary and Itti 2008). Moreover, the distribution of objects in a scene is a better account of fixation selection than salience. The locations that people select for fixation in photographic scenes are better described by the locations of objects in the scenes than by the peaks in a low-level salience map (Einhauser et al. 2008). Indeed, object-based descriptions may be a more appropriate level of scene description for understanding fixation selection than low-level feature descriptions (Nuthmann and Henderson 2010). It is possible that low-level visual conspicuity might offer a convenient heuristic for the brain to select locations that are likely to contain objects (Elazary and Itti 2008). However, semantically interesting locations are preferentially selected even when their low-level information is degraded: A blurred face will still attract fixations even though it has little signature in a salience map (Nyström and Holmqvist 2008). This result implies that even though low-level conspicuity tends to correlate with objects, it is not sufficient to explain why people select objects when viewing a scene.

In light of the shortcomings of purely low-level models of fixation selection, a number of models have been proposed that incorporate high-level factors. Navalpakkam and Itti (2005) suggested that higher-level knowledge might result in selective tuning of the various feature maps that make up the overall salience map. If the features of a target object are known, the corresponding channels in the salience map can be selectively weighted, and this should enhance the representation of the target object in the salience map. Other sources of knowledge about objects present potential candidates that may guide our search for them. Most objects are more likely to occur in some places than others—for example, clocks are more likely to be found on walls than on floors or ceilings. Torralba et al. (2006) suggested that these typical spatial associations between objects and scenes can be used to produce a contextual prior describing the likely location of an object in a scene. This contextual prior can then be used to modulate a low-level conspicuity map of the scene, producing a context-modulated salience map. Therefore, the suggestion is that, in general, gaze will be directed to locations of high salience that occur within the scene regions in which the target is expected to be found. Previous experience of objects can be used not only to form contextual priors describing where objects are likely to be found but also to produce "appearance priors" describing the likely appearance of a class of objects (Kanan et al. 2009). Again, if searching for a clock, we can use prior knowledge about the likely appearance of clocks to narrow down the search to clock-like objects in the scene irrespective of where they occur. Kanan

et al. (2009) proposed a model in which the appearance prior is used to modulate a low-level salience map in much the same way as Torralba et al. (2006) proposed for their context modulation. As such, in Kanan et al.'s (2009) model, gaze selects locations of high salience that coincide with scene regions that share properties characteristic of the target object's class. Modulating salience maps using context priors or appearance priors improves the performance of the model (Kanan et al. 2009; Torralba et al. 2006), suggesting that decisions about where to look when viewing scenes are likely to involve these types of information. Indeed, if both context and appearance priors are used to modulate a salience map, the resultant model is able to predict the likely locations that humans will fixate with remarkably high accuracy (Ehinger et al. 2009).

Many current models incorporate higher-level factors as modifiers of a basic low-level salience map. However, others suggest alternative cores to their models. In Zelinsky's (2008) target acquisition model, visual information is not represented as simple feature maps but as higher-order derivatives that incorporate object knowledge. Similarly, in Wischnewski et al.'s (2010) model, selection involves static and dynamic proto-objects rather than first-order visual features. Nuthmann and Henderson (2010) propose an object-level description as the core component of deciding where to look. These models each offer good explanatory power for scene viewing and demonstrate that basic visual features need not be the language of priority maps for fixation selection.

1.3 Behavioural Goals in Eye Guidance

Since they first proposed the salience model, Itti and Koch (2000) recognised that it would always be limited by its inability to account for the influence of behavioural goals on fixation selection. The importance of behavioural goals and the profound effect they have upon where people look have been recognised since the earliest work on illusions and scene viewing. As we have seen, Judd (1905) came to the conclusion that the instructions given to participants had more of an effect on where people fixated than did the stimuli when they were viewing simple line illusions. Buswell (1935) extended this idea to complex scene viewing. He showed that fixation behaviour when viewing a photograph of the Tribune Tower in Chicago with no instructions was very different from fixation behaviour by the same individual when asked to look for a face at one of the windows in the tower (Fig. 2). Yarbus (1967) later provided what has now become a classic demonstration of the profound effect task instructions have on viewing behaviour. A single individual viewed Repin's *They did not expect him* seven times, each time with a different instruction prior to viewing. Fixation behaviour was markedly different each time, and the locations fixated corresponded to those that might be expected to provide information relevant to the task suggested by the instructions (Fig. 3). These demonstrations provide a profound and important challenge for any model of fixation behaviour. Empirical evaluations of the explanatory power of low-level feature

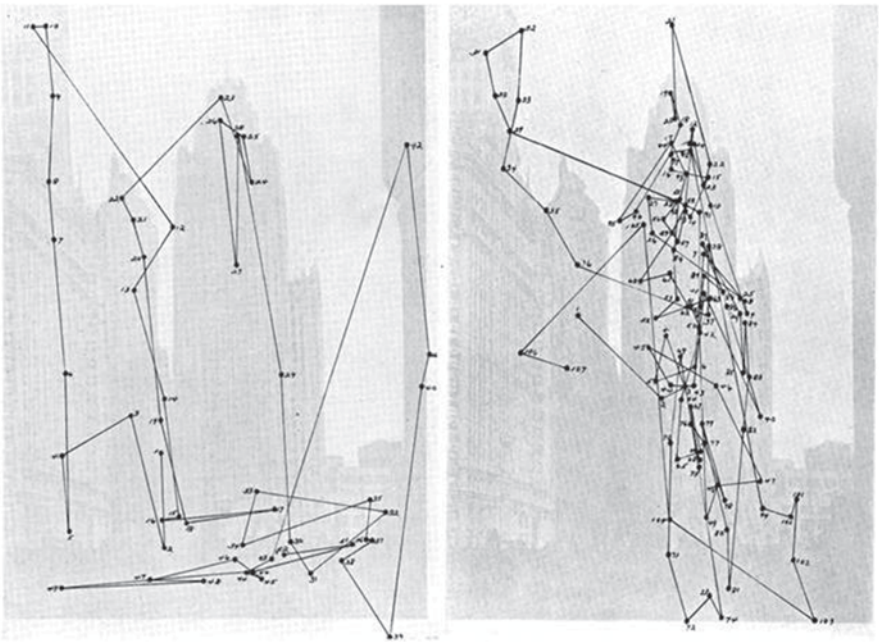

Fig. 2 *Left,* eye movements of an individual viewing the Chicago Tribune Tower with no specific instructions. *Right,* eye movements of the same individual when instructed to look for a face at a window in the tower. (Adapted from Buswell 1935)

salience during goal-directed looking tasks have shown that correlations between salience and selection are very low or absent when the observer is engaged in an explicit task such as search (Einhauser et al. 2008; Henderson et al. 2007; Underwood et al. 2006) or scene memorisation (Tatler et al. 2005). Where greater explanatory power has been found has been in cases where the task is not defined—the so-called free-viewing paradigm. In this task, participants are given no instructions other than to look at the images that they will be presented with. One motivation for employing this free-viewing paradigm is that it may be a way of isolating task-free visual processing, minimising the intrusion of higher-level task goals on fixation selection (Parkhurst et al. 2002). However, this paradigm is unlikely to produce task-free viewing in the manner hoped and is more likely to provide a situation where viewers select their own priorities for inspection (Tatler et al. 2005, 2011). It is also worth noting that even in such free-viewing situations, correlations between features and fixations are weak (Einhauser et al. 2008; Nyström and Holmqvist 2008; Tatler and Kuhn 2007).

1.4 Limits of the Screen

State-of-the-art models of scene viewing are able to make predictions that account for an impressive fraction of the locations fixated by human observers (Ehinger

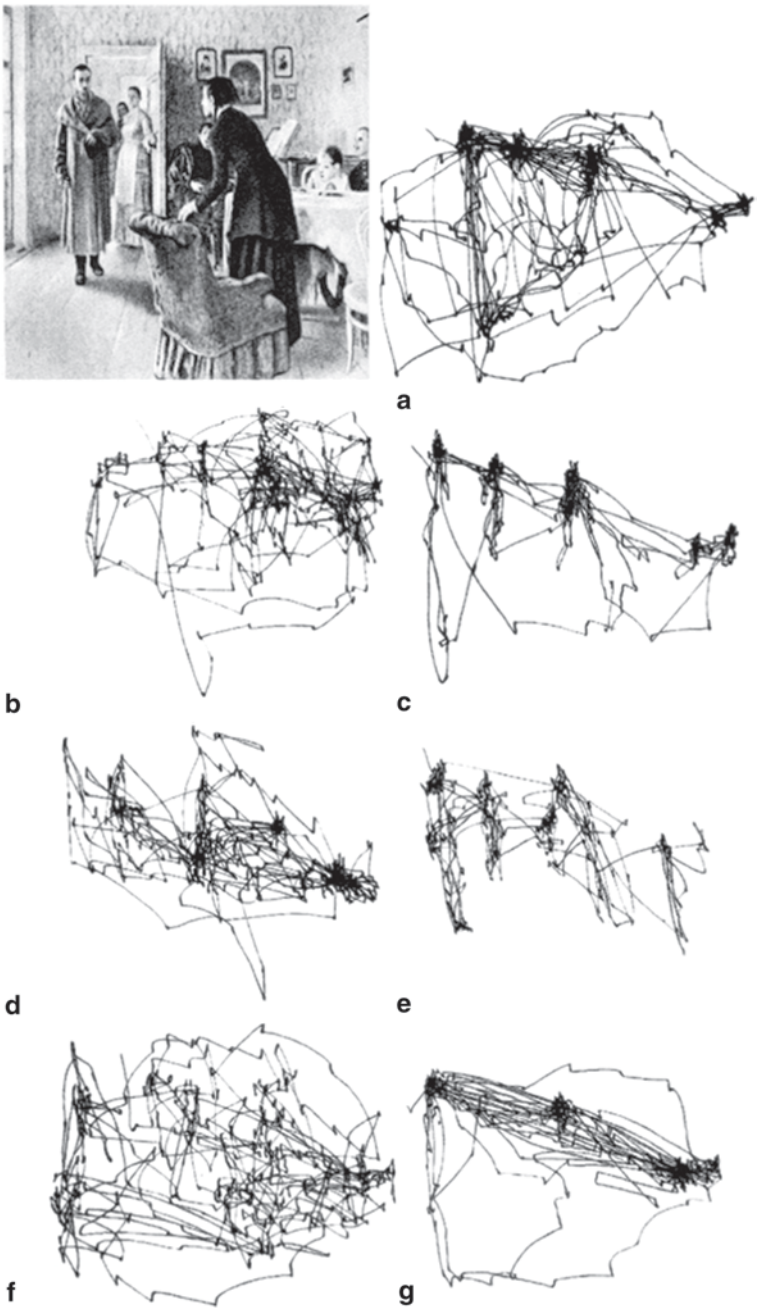

Fig. 3 Recordings of one participant viewing The Unexpected Visitor seven times, each with different instructions prior to viewing. Each record shows eye movements collected during a 3-minute recording session. The instructions given were (a) Free examination. (b) Estimate the material circumstances of the family in the picture. (c) Give the ages of the people. (d) Surmise what the family had been doing before the arrival of the unexpected visitor. (e) Remember the clothes worn by the people. (f) Remember the position of the people and objects in the room. (g) Estimate how long the unexpected visitor had been away from the family. (Illustration adapted from Yarbus, 1967, Figure 109, for Land and Tatler, 2009)

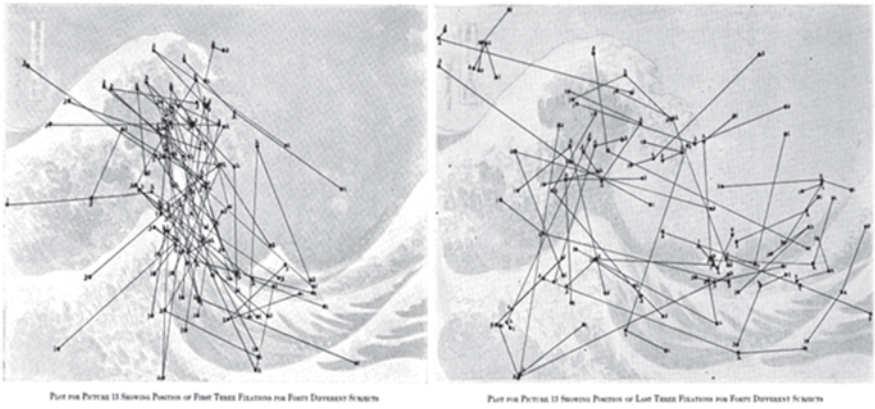

Fig. 4 Left, eye movements of 40 subjects during the first second of viewing *The Wave*. Right, eye movements of 40 subjects during the final second of viewing *The Wave*. From Buswell (1935).

et al. 2009). However, it is important to remember that the majority of evidence regarding the control of fixation selection in scene viewing comes from studies in which participants view static photographic (or photorealistic) images displayed on computer monitors. Static scenes are, of course, very different from real environments in many ways and it is important to ask the extent to which the principles of fixation selection identified in such studies generalise beyond the limits of the computer screen. There are at least four key aspects of static scene-viewing paradigms that must be considered. First, scenes typically appear with a sudden onset, are viewed for a few seconds and then disappear again. Second, the viewed scene is wholly contained within the frame of the monitor. Third, static scenes necessarily lack the dynamics of real environments. Fourth, the tasks that we engage in when viewing images on screens are rather unlike those that we engage in in more natural contexts.

Viewing behaviour is very different in the first second or two following scene onset than it is later on in the viewing period (Buswell 1935; Fig. 4). Locations selected for fixation are more similar across observers soon after scene onset than they are after several seconds of viewing (Buswell 1935; Tatler et al. 2005). Early consistency across participants followed by later divergence in fixation selection could imply that early fixations are more strictly under the control of low-level salience (Carmi and Itti 2006; Parkhurst et al. 2002) or alternatively that higher-level strategies for viewing are common soon after scene onset but later diverge (Tatler et al. 2005). Whatever the underlying reasons for these changes in viewing behaviour over time, the mere fact that viewing behaviour is very different soon after scene onset than it is later on raises concerns about the generalisability of findings from scene-viewing paradigms. It seems likely that the priorities for selection are rather different in the first second or two of viewing than they are for subsequent fixations. Given that sudden whole-scene onsets are not a feature of real-world environments, it may be that the factors that underlie saccade-targeting

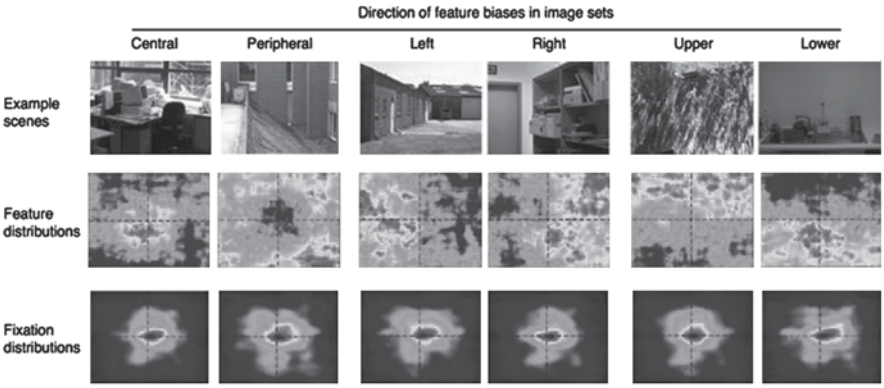

Fig. 5 The central fixation bias in fixation behaviour when viewing images on a computer monitor. Fixation distributions (*bottom row*) show a strong central tendency irrespective of the distribution of features in the images (*middle row*). (Redrawn from Tatler 2007)

decisions soon after scene onset do not reflect those that underlie natural saccade target selection. As such, this potentially limits the utility of models developed using these data.

When viewing scenes on a monitor, observers show a marked tendency to fixate the centre of the scene more frequently than the periphery (e.g. Parkhurst et al. 2002). Compositional biases arising from photographers' tendencies to put objects of interest near the centre of the viewfinder mean that images typically used in static scene-viewing paradigms often have centrally weighted low-level feature distributions. However, the distribution of low-level features in scenes is not sufficient to explain this tendency to preferentially fixate the centre of scenes (Tatler 2007). When viewing scenes with feature distributions that are not centrally biased, the tendency to fixate the centre of the scene persists, and indeed the overall distribution of fixation locations is not shifted by the distribution of features across the scene (Fig. 5). Not only is this result challenging for low-level salience models but also it raises a more serious concern for screen-based experiments: that these central fixation tendencies exist irrespective of the content of the scenes shown to the observers. There are a number of reasons that this tendency to look at the screen centre may be adaptive—it provides an optimal view of the whole scene, a good starting point for scene exploration and a location where objects of interest are expected given previous experience of photographs—but the factors that underlie these decisions to look at the screen centre are not strictly visual. As such, attempting to model these selections on the basis of the targeted visual information may be rather misleading.

Of course, static scenes necessarily lack the dynamics of real environments, but one potential solution here is to use dynamic moving images to overcome this shortcoming. By passively recording a movie of a scene from a single static viewpoint (Dorr et al. 2010) or recording a head-centred view of an environment (Cristino and Baddeley 2009), it is possible to produce dynamic scenes that have less pronounced

compositional biases than static scenes and no sudden whole-scene onsets beyond that at the start of the movie. However, even for head-centred movies, Cristino and Baddeley (2009) found that viewing behaviour was dominated by scene structure, with fixations showing a spatial bias related to the perceived horizon in the scene.

Screen-based viewing paradigms—using either static or dynamic scenes—are also limited in the types of tasks that observers can engage in. In such situations, task manipulations typically involve responding to different instructions, such as to freely view, search or memorise scenes. However, these tasks lack a fundamental component of natural behaviour: interaction with the environment. In natural tasks, we typically employ gaze in a manner that is intricately linked to our motor actions (see Land and Tatler 2009). The lack of motor interaction with the scene in picture-viewing paradigms may well have fundamental effects upon how gaze is deployed (Steinman 2003). Epelboim et al. (1995, 1997) showed that many aspects of gaze coordination change in the presence of action, including the extent to which gaze shifts involve head as well as eye movements, the extent to which the eyes converge on the plane of action and the relationship between saccade amplitude and peak velocity. The limitation of using screen-based paradigms to study real-world behaviours was highlighted by Dicks et al. (2010) in a task that required goalkeepers to respond to either a real person running to kick a football or a life-sized video of the same action. Furthermore, the nature of the response was varied such that the goalkeepers responded verbally, moved a lever or moved their body to indicate how they would intercept the ball's flight. The locations fixated by the goalkeeper differed between real and video presentations and also with the type of response required. Importantly, viewing behaviour was different when observing a real person and responding with a whole body movement than in any other condition. This highlights the importance of studying visual selection in a natural task setting and suggests that any removal of naturalism can result in fixation behaviour that is unlike that produced in real behaviour.

2 Eye Guidance in Natural Tasks

From its evolutionary origins, a fundamental function of vision has been to provide information that allows the organism to effectively and appropriately carry out actions necessary for survival. Decisions about when and where to move the eyes in real-world situations are therefore likely to be intimately linked to the information demands of the current actions. Thus, it is appropriate to consider gaze not as an isolated system but as part of a broader network of vision, action and planning as we interact with the environment (Fig. 6). Thus, if we are to produce an ecologically valid account of the factors underlying fixation selection, we must consider whether models developed using laboratory-based paradigms can be extended to more natural settings.

To date, the computational models developed for scene-viewing paradigms have rarely been tested in the context of natural behaviour. One exception to this comes

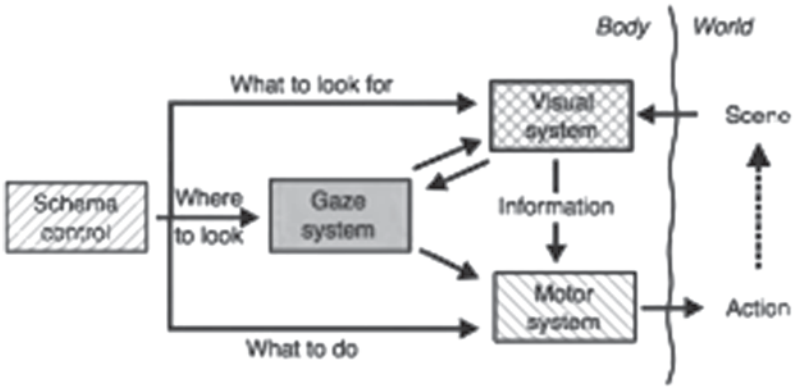

Fig. 6 Schematic illustration of interplay between gaze control, visual processing, motor action and schema planning in natural behaviour

from Rothkopf et al. (2007) who showed that in a virtual reality walking task, low-level salience was unable to account for fixation selection. Instead, fixations were made to task-relevant objects and locations in the environment irrespective of their low-level visual salience. While more state-of-the-art models incorporating higher level factors (Ehinger et al. 2009; Kanan et al. 2009; Torralba et al. 2006) have yet to be tested in natural settings, the fundamental failure of the pure salience model in a naturalistic setting raises concerns about the utility of these types of model, which retain visual conspicuity as their core. An alternative, and necessary, approach is to consider what principles for fixation selection can be identified from studies of eye movements during natural tasks and use these to specify the aspects of behaviour that any model of fixation selection in natural tasks must be able to account for.

Eye movements have been studied in a wide variety of real-world activities from everyday domestic tasks to driving, to ball sports (see Land and Tatler 2009). Across all of these tasks, it is clear that where we look is intimately linked to our actions. This simple and universal finding itself clearly demonstrates the fundamental influence that the active task requirements place on guiding eye movement behaviour. The intricate link between our behavioural goals and the allocation of overt visual attention is highlighted by the fact that when engaged in a natural task, we rarely fixate objects that are not relevant to our overall behavioural goals (Hayhoe et al. 2003; Land et al. 1999). In comparison, before beginning the task we are equally likely to fixate objects that will later be task relevant or irrelevant (Hayhoe et al. 2003). But the influence of natural behaviour on viewing is not simply to impose a preference to look at objects relevant to the overall goals of the behaviour. What is clear is that the eyes are directed to the locations that are relevant to the task on a moment-to-moment basis. That is, at each moment in time we look at the locations that convey information that allows us to act upon the environment in order to complete our current motor acts (Ballard et al. 1992; Hayhoe et al. 2003; Land et al. 1999; Land and Furneaux 1997; Patla and Vickers 1997; Pelz and Canosa 2001).

For example, when approaching a bend in the road, drivers fixate the tangent point of the bend, and this location provides key information required to compute the angle that the steering wheel should be turned (Land and Lee 1994). In table tennis (Land and Furneaux 1997) and cricket, we look at the point where the ball will bounce (Land and McLeod 2000), and this point offers crucial information about the likely subsequent trajectory that the ball will follow. These findings illustrate that spatial selection is intimately linked to the current target of manipulation. Thus, in order to understand where people look, we must first understand the nature of the behaviour they are engaged in and the structure of the task. Of course, this means that spatial selection will be somewhat parochial to the particular task that a person is engaged in. The type of information that is required to keep a car on the road is likely to be very different from that required to make a cup of tea. As such, the type of information that governs spatial selection by the eye is likely to be very different in different tasks.

While spatial selection is, in some ways, parochial to the task, temporal allocation of gaze is strikingly similar across many real activities. For many activities, gaze tends to be directed to an informative location around 0.5–1 s before the corresponding action. In tea making, the eyes fixate an object on average 0.5–1 s before the hands make contact with the object. In music reading (Furneaux and Land 1999) and speaking aloud (Buswell 1920), the eyes are typically 0.5–1 s ahead of key presses and speech respectively. During locomotion, the eyes fixate locations about 0.5–1 s ahead of the individual, and this is found when walking (Patla and Vickers 2003), driving at normal speed (Land and Lee 1994) or driving at high speed (Land and Tatler 2001). The correspondence in eye-action latency across such different tasks suggests that this temporal allocation of gaze is not only under strict control but also under common control in many real-world activities. As such, any account of gaze allocation in natural tasks must be able to explain this temporal coupling between vision and action in which gaze is allocated in anticipation of the upcoming action.

Of course, there are exceptions to the typical 0.5–1 s eye-action latency found in many natural tasks. In particular, in ball sports like cricket, squash and table tennis, there simply is not enough time to keep the eyes this far ahead of action. In these situations, anticipatory allocation of gaze is still seen albeit over rather different timescales to other tasks. In cricket (Land and McLeod 2000) and table tennis (Land and Furneaux 1997), gaze is directed to the point in space where the ball will bounce about 100 ms before the ball arrives. Similarly, in squash the eyes arrive at the front wall about 100 ms ahead of the ball (Hayhoe et al. 2011). If the ball bounces off a wall, gaze is allocated to a location that the ball will pass through shortly after it bounces off the wall with an average of 186 ms before the ball passes through this space (Hayhoe et al. 2011).

The examples described above illustrate that gaze is used to acquire information required for ongoing action and is allocated ahead of action. Correct spatiotemporal allocation of gaze is central to successful task performance in many situations. For example, in cricket both a skilled and an unskilled batsman were found to look at the same locations (the release of the ball and the bounce point), but the skilled

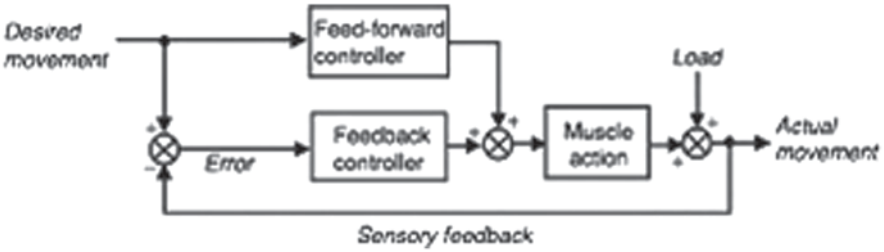

Fig. 7 Action control using feedforward and feedback loops. (From Land and Tatler 2009)

batsman looked at the bounce point about 100 ms before the ball arrived, whereas the unskilled batsman fixated this location at or slightly after the ball arrived at the bounce point (Land and McLeod 2000). Given the importance of appropriate spatiotemporal allocation of gaze in natural behaviours, what internal processes might underlie this visuomotor co-ordination in space and time? Anticipatory allocation of gaze ahead of ongoing action could be achieved if we allocate gaze on the basis of internal predictive models (Hayhoe et al. 2011; Land and Tatler 2009). The idea that the brain constructs internal predictive models of external events has been around for some time (e.g. Miall and Wolpert 1996; Wolpert et al. 1995; Zago et al. 2009). An elegant example of the importance of both feedback and prediction in visuomotor control was provided by Mehta and Schaal (2002). When balancing a 1-m pole on a table tennis bat, visual feedback alone was inadequate: If the tip of the pole was touched, disturbing the pole, the delay between visual sampling of this event and an appropriate motor response was slower (220 ms) than the maximum possible delay for normal balancing (160 ms). This suggests that to balance the pole effectively, visual feedback was too slow, and so task performance must be reliant on internal prediction. The use of forward models in this behaviour was underlined by the finding that participants were able to continue to balance the pole even when vision was removed for periods of up to 500–600 ms. Mehta and Schaal (2002) explained this behaviour as involving a Kalman filter where raw sensory feedback is compared to a copy of the motor command to the muscles in order to provide an optimised prediction of the consequences of action (Fig. 7). Such a scheme has the advantage of being able to use prediction alone in the absence of visual feedback and, thus, can tolerate brief interruptions to sensory feedback.

However, the scheme illustrated in Fig. 7 is unlikely to be sufficient for more complex tasks like the ball sports and everyday activities discussed earlier. In these situations, gaze acquires information about the future state of the world by looking at locations where action is about to occur: Objects are fixated 0.5–1 s before they are manipulated; the space where an object will be set down is fixated about half a second before the object is placed there; the spot where a ball will soon pass through is fixated 100–200 ms before the ball arrives. These anticipatory allocations of gaze certainly involve internal predictive models, but these models are not predictors in the sense described in Fig. 7. Rather, these models are mechanisms for providing

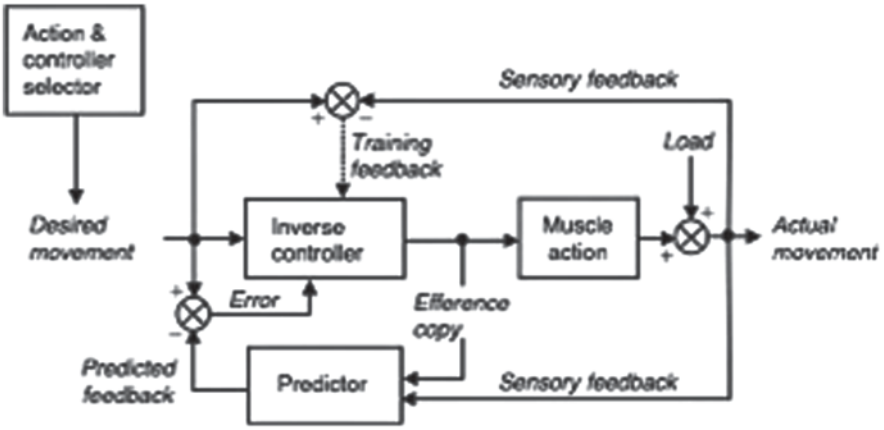

Fig. 8 Control of action using an inverse controller to refine task performance, together with a predictor in the feedback loop that provides delay-free feedback. (Modified from the "motor control system based on engineering principles" of Frith et al. 2000 for Land and Tatler 2009)

feedforward input to the motor controllers, which manage the relationship between the desired goal and the motor commands required to achieve that goal. The model illustrated in Fig. 8 depicts a situation that is suitable for understanding complex skilled behaviour. The inclusion of an inverse controller provides a mechanism for learning by transforming the desired sensory consequences of an action back into the motor commands that will produce those consequences. The mismatch between the desired and actual sensory consequences of the actions produced by the inverse controller provides the signal with which the controller can be improved. This model places learning at the heart of visuomotor co-ordination. Initially, for a novel visuomotor task, this system should operate essentially by trial and error, using feedback to improve performance. But after sufficient training, the controller can operate in an open-loop manner using the desired result as its input. Evidence in support of this scheme was provided by Sailer et al. (2005) who studied eye–hand co-ordination while learning a novel visuomotor task in which a manual control device was manipulated in order to move a cursor to targets on a computer monitor. Initially, the eyes lagged the movements of the cursor. In this phase, gaze was presumably deployed to provide feedback about the consequences of motor acts. However, after sufficient training, participants were able to perform the task well and gaze was deployed ahead of action, with the eyes leading the movements of the cursor by an average of about 0.4 s.

Not only can the scheme illustrated in Fig. 8 be used to explain visuomotor skill acquisition, but also it can provide a framework for online refinement of the internal models in the light of incoming sensory evidence. In cricket, a general model of how the ball will behave at the bounce point can be built up over years of experience, but the general model must be flexible enough to be adapted to the current pitch conditions for any given innings. The defensive play that batsmen typically

engage in at the start of their innings presumably reflects this refinement of the general model based on sensory input for the current conditions (Land and McLeod 2000). Similar online adaptations of internal models based on current experience have been found when unexpected changes are made during ongoing behaviour.

Hayhoe et al. (2005) provide a nice example of how we are able to adapt our internal forward models to an unexpected change in the environment. Three people stood in a triangular formation and threw a tennis ball to each other. Like cricket, when receiving a ball, participants first fixated the release point of the ball before making an anticipatory saccade to the predicted bounce point, and then tracked the ball after its bounce. However, after several throws, one of the participants surreptitiously switched the tennis ball for a bouncier ball. When this happened, the usual oculomotor tracking of the ball broke down on the first trial with the new ball; instead, participants reverted to making a series of saccades. However, the flexibility of the internal predictors was demonstrated first by the fact that participants still caught this unexpected ball, and second by the adaptation in behaviour that followed over the next few trials with the new ball. Over the next six trials, arrival time at the bounce point advanced such that by the sixth throw with the new ball the participant was arriving at the bounce point some 100 ms earlier than on the first trial. Furthermore, the pursuit behaviour was rapidly reinstated, with pursuit accuracy for the new, bouncier ball about as good as it had been for the tennis ball by the third throw of the new ball. Thus, not only do the results demonstrate a reliance on forward models for task performance and the allocation of gaze, but they also demonstrate that these models can rapidly adapt to change in the environment.

When observers walk toward other people who they have encountered previously and who may attempt to collide with them, Jovancevic-Misic and Hayhoe (2009) showed that observers can use prior experience of these individuals to allocate gaze on the basis of the predicted threat the individual poses. Those people who the observers predicted were likely to collide with them were be looked at for longer than those who observers predicted were unlikely to collide with them, based on previous encounters. Moreover, if after several encounters, the behaviour of the oncoming individuals changed such that those who were previously of low collision threat were now trying to collide with the observer and vice versa, gaze allocation rapidly adapted to these changed roles over the next couple of encounters.

The model of visuomotor co-ordination outlined in Fig. 8 provides a framework for understanding spatiotemporal allocation of gaze for the actions required to serve ongoing behavioural goals. This model can be used to explain how gaze is allocated ahead of action in skilled behaviour and places emphasis on the importance of learning and online refinement of internal models. Learning in the proposed inverse controller can be achieved via simple reinforcement. Reward mechanisms therefore may play a crucial role not only in the development of these internal models but also in the moment-to-moment allocation of gaze. In support of this possibility, the eye movement circuitry is sensitive to reward (Montague and Hyman 2004; Schultz 2000) and, therefore, reward-based learning of gaze allocation is neurally plausible. Sprague and colleagues (e.g. Sprague et al., 2007) have begun to develop reward-based models of gaze behaviour in complex tasks. In a walking task that involves

three concurrent sub-goals (avoid obstacles, collect "litter" and stay on the path), some reward value can be assigned to each sub-task. Gathering information for a sub-task is therefore rewarded. In this model, attention can only be allocated to one sub-task at a time, and uncertainty about non-attended sub-tasks increases over time. As uncertainty increases, so does the amount of information (i.e. the reduction in uncertainty) that will be gained by attending to that sub-task. The model allocates attention over time on the basis of the expected reward associated with attending to each sub-task and reducing uncertainty about that sub-task (Sprague et al. 2007). This model offers a proof of principle that gaze allocation in natural tasks can be explained using reward-based models.

Reward-based explanations of sensorimotor behaviour are emerging across a variety of experimental settings (e.g. Tassinari et al. 2006; Trommershäuser et al. 2008). Hand movements are optimised to maximise externally defined reward (e.g. Seydell et al. 2008; Trommershäuser et al. 2003). Saccadic eye movements show similar sensitivity to external monetary reward (Stritzke et al. 2009) and are consistent with an ideal Bayesian observer that incorporates stimulus detectability and reward (Navalpakkam and Itti 2010). It seems likely therefore that reward-based underpinnings to saccadic decisions may become increasingly important to our understanding of eye movements in laboratory and real environments. Moreover, reward-based models of fixation selection provide a promising new direction for research and language for describing the priority maps that are likely to underlie decisions about when and where to move the eyes.

References

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *337*(1281), 331–338.

Buswell, G. T. (1920). *An experimental study of the eye-voice span in reading*. Chicago: Chicago University Press.

Buswell, G. T. (1935). *How people look at pictures: A Study of the Psychology of Perception in Art*. Chicago: University of Chicago Press.

Cameron, E. H., & Steele, W. M. (1905). The poggendorff illusion. *Psychological Monographs*, *7*(1), 83–111.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, *46*(26), 4333–4345. doi:10.1016/j.visres.2006.08.019.

Cristino, F., & Baddeley, R. J. (2009). The nature of the visual representations involved in eye movements when walking down the street. *Visual Cognition*, *17*(6–7), 880–903.

Dicks, M., Button, C., & Davids, K. (2010). Examination of gaze behaviors under in situ and video simulation task constraints reveals differences in information pickup for perception and action. *Attention, Perception, & Psychophysics*, *72*(3), 706–720. doi:10.3758/APP.72.3.706.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), *28*, 1–17. doi:10.1167/10.10.28.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling Search for People in 900 Scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(6–7), 945-978. doi:10.1080/13506280902834720.

Einhauser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14):18, 1–26, http://www.journalofvision.org/ content/8/14/18, doi: 10.1167/8.14.18.

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*(3), 3.1–15. doi:10.1167/8.3.3.

Epelboim, J. L., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*(23–24), 3401–3422.

Epelboim, J. L., Steinman, R. M., Kowler, E., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1997). Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Research*, *37*(18), 2597–2607.

Erdmann, B., & Dodge, R. (1898). *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage*. Halle: Niemeyer.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6.1–17. doi:10.1167/8.2.6.

Frith, C.D., Blakemore, S-J., & Wolpert, D.M. (2000). Abnormalities in the awareness of and control of action. *Philosophical Transactions of the Royal Society of London B* 355, 1771–1788.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *266*(1436), 2435–2440.

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2011). Predictive eye movements in natural vision. *Experimental Brain Research*, *217*(1), 125–136. doi:10.1007/s00221-011-2979-2.

Hayhoe, M., Mennie, N., Sullivan, B., & Gorgos, K. (2005) The role of internal models and prediction in catching balls. In: Proceedings of AAAI fall symposium series.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49–63. doi:10:1167/3.1.6.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. L. (2007). Visual saliency does not account for eye movements during search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford, UK: Elsevier.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254–1259.

Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive Gaze Control in Natural Environments. *Journal of Neuroscience*, *29*(19), 6234–6238. doi:10.1523/JNEUROSCI.5570-08.2009.

Judd, C. H. (1905). The Müller-Lyer illusion. *Psychological Monographs*, *7*(1), 55–81.

Judd, C. H., & Courten, H. C. (1905). The Zöllner illusion. *Psychological Monographs*, *7*(1), 112–139.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*(6–7), 979–1003.

Koch, C., & Ullman, S. (1985). Shifts in selective visual-attention—towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *352*(1358), 1231–1239.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483), 742–744. doi:10.1038/369742a0.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience*, *3*(12), 1340–1345. doi:10.1038/81887.

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology*, *11*(15), 1215–1220.
Land, M. F., & Tatler, B. W. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. Oxford: OUP.
Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.
McAllister, C. N. (1905). The fixation of points in the visual field. *Psychological Monographs*, *7*(1), 17–53.
Mehta, B. & Schaal, S. (2002). Forward models in visuomotor control. *Journal of Neurophysiology* 88, 942–953.
Miall, R.C. & Wolpert, D.M. (1996). Forward models for physiological motor control. *Neural Networks*, 9, 1265–1279.
Montague, P., & Hyman, S. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*, 760–767.
Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research,* 45, 205–231.
Navalpakkam, V., & Itti, L. (2010). A goal oriented attention guidance model. *Biologically Motivated Computer Vision*, 81–118.
Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, 10(8) 20, 1–19. doi:10.1167/10.8.20.
Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing-both initially and overall. *Journal of Eye Movement Research*, 2, 1–11.
Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123.
Patla, A. E., & Vickers, J. N. (1997). Where and when do we look as we approach and step over an obstacle in the travel path? *Neuroreport*, *8*(17), 3661–3665.
Patla, A. E., & Vickers, J. N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale*, *148*(1), 133–138. doi:10.1007/s00221-002-1246-y.
Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*(25–26), 3587–3596.
Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological bulletin*, *124*(3), 372–422.
Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network (Bristol, England)*, 10(4), 341–350.
Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14), 16.1–20. doi:10.1167/7.14.16.
Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. The Journal of neuroscience: The official journal of the Society for. *Neuroscience*, *25*(39), 8833–8842. doi:10.1523/JNEUROSCI.2658-05.2005.
Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience*, *1*(3), 199–207.
Seydell, A., McCann, B. C., Trommershäuser, J., & Knill, D. C. (2008). Learning stochastic reward distributions in a speeded pointing task. *Journal of Neuroscience*, *28*, 4356–4367.
Sprague, N., Ballard, D. H., & Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Transactions on Applied Perception*, 4, 11.
Steinman, R. M. (2003). Gaze control under natural conditions. In: Chalupa, L.M., & Werner, J.S. (Eds). *The Visual Neurosciences*. pp. 1339–1356. Cambridge: MIT Press.
Stritzke, M., Trommershäuser, J., & Gegenfurtner, K. R. (2009). Effects of salience and reward information during saccadic decisions under risk. *Journal of the Optical Society of America A*, *26*, B1–B13.
Stratton, G. M. (1906). Symmetry, linear illusions, and the movements of the eye. *Psychological Review*, *13*, 82–96.

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining priors and noisy visual cues in a rapid pointing task. *Journal of Neuroscience*, *26*, 10154–10163.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4, 1–17, http:// www.journalofvision.org/content/7/14/4, doi:10.1167/ 7.14.4.

Tatler, B. W., & Kuhn, G. (2007). Don't look now: the misdirection of magic. In: van Gompel, R., Fischer, M., Murray, W., & Hill, R. (Eds). *Eye Movement Research: Insights into Mind and Brain*. pp. 697–714. Amsterdam: Elsevier.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, *45*(5), 643–659.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11(5)*, *5* 1–23. doi:10.1167/11.5.5.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America A*, *20*, 1419–1433.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning, and statistical decision theory. *Trends in Cognitive Sciences*, *12*, 291–297.

Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. European Journal of. *Cognitive Psychology*, *18*(3), 321–342. doi:10.1080/09541440500236661.

Wischnewski, M., Belardinelli, A., & Schneider, W. (2010). Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. *Cognitive Computation*, *2*(4), 326–343.

Wolfe, J. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York: OUP.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*(1), 33–39.

Wolpert. D.M., Ghahramani, Z., & Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Zago M, McIntyre J, Patrice Senot P, & Lacquaniti F (2009) Visuo-motor coordination and internal models for object interception. *Experimental Brain Research* 192(4):571–604.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*(4), 787–835. doi:10.1037/a0013118.