Eye guidance in natural vision: Reinterpreting salience

Benjamin W. Tatler
School of Psychology, University of Dundee,
Dundee, UK

Mary M. Hayhoe
Center for Perceptual Systems, University of Texas
at Austin, Austin, TX, USA

Michael F. Land
School of Life Sciences, University of Sussex,
Sussex, UK

Dana H. Ballard
Computer Science Department, University of Texas
at Austin, Austin, TX, USA

Models of gaze allocation in complex scenes are derived mainly from studies of static picture viewing. The dominant framework to emerge has been image salience, where properties of the stimulus play a crucial role in guiding the eyes. However, salience-based schemes are poor at accounting for many aspects of picture viewing and can fail dramatically in the context of natural task performance. These failures have led to the development of new models of gaze allocation in scene viewing that address a number of these issues. However, models based on the picture-viewing paradigm are unlikely to generalize to a broader range of experimental contexts, because the stimulus context is limited, and the dynamic, task-driven nature of vision is not represented. We argue that there is a need to move away from this class of model and find the principles that govern gaze allocation in a broader range of settings. We outline the major limitations of salience-based selection schemes and highlight what we have learned from studies of gaze allocation in natural vision. Clear principles of selection are found across many instances of natural vision and these are not the principles that might be expected from picture-viewing studies. We discuss the emerging theoretical framework for gaze allocation on the basis of reward maximization and uncertainty reduction.

Keywords: salience, natural tasks, eye movements, reward, learning, prediction

Citation: 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, http://www.journalofvision.org/content/11/5/5, doi:10.1167/11.5.5.

Introduction

Visually guided behaviors require the appropriate allocation of gaze in both space and time. High acuity foveal vision must be directed to locations that provide information for completing behavioral goals. Behaviorally informative locations change with progress through a task, so this allocation of gaze must not only be to the right places but must also be at the right times to serve behavior. Understanding the principles that underlie the deployment of gaze in space and time is, therefore, important for understanding any visually guided behavior.

In this article, we review the current state of models of eye guidance for complex scene viewing and whether they can generalize to natural behavior. In particular, we review the dominant class of models that has emerged to explain gaze allocation in picture viewing: those that are based on low-level image properties, often operationalized as image salience. While this approach has provided insights into oculomotor selection and has given rise to a considerable volume of research, we argue that most current models offer only a limited description of human gaze behavior. Moreover, we argue that the dominant paradigm—that of picture viewing—is an inappropriate domain of explanation if we wish to understand eye movement behavior more generally. While most models have been built around a core of low-level feature conspicuity, some emerging models attempt to base selection on higher level aspects of scenes. We consider the direction that these models are taking and whether this will allow insights into vision in natural settings. We approach this by considering what a model of natural eye guidance should be able to explain. That is, we highlight the principles of fixation selection in natural tasks that can be found to generalize across many real-world situations; these are the components of eye movement behavior that need to be explained by any theoretical model. The common underlying principles for eye guidance suggest that behavioral relevance and learning are central to how we allocate gaze. These principles necessarily change the emphasis of what should be modeled and we suggest that a framework incorporating behavioral rewards will

Received December 21, 2010; published May 27, 2011

ISSN 1534-7362 © ARVO

provide a useful approach for understanding the manner in which models of eye guidance may be implemented in the future.

Image salience and eye movement behavior

The extensive psychophysical literature on visual search has demonstrated that basic visual features can capture and guide attention (see Wolfe, 1998). If a target differs from a set of distractors in just a single feature dimension, such as color or orientation, it can be detected very rapidly, and detection time remains fast irrespective of the number of distractors present (Treisman & Gelade, 1980). This "pre-attentive" capture ("popout") suggests that features can drive the allocation of attention. Similarly, more complex search, where targets are defined by the unique conjunction of two features, can successfully be explained using serial selection driven by image features. Models such as Treisman's feature integration theory (Treisman & Gelade, 1980) or Wolfe's (2007) guided search model produce human-like search behavior using only low-level featural information. A natural extension of this work was to ask whether this principle could be applied to understanding how attention is allocated in more complex scenes. These models of visual search underlie the most influential class of models of gaze patterns in picture viewing based on low-level image features. One computational implementation of this class of model is the notion of the "salience map," a spatial ranking of conspicuous visual features that could be candidates for covert or overt attention (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985). The salience map concept has had a profound influence on the research field and has become an integral component of many subsequent models of gaze allocation. In the original implementation of the salience model, when presented with a scene, low-level features are extracted in parallel across the extent of the viewed scene (Figure 1). Local competition across image space and feature scales results in feature maps for luminance contrast, color contrast, and orientation contrast. These individual feature maps are combined by weighted sum to create an overall distribution of local feature contrast, known as the "salience map." Attention is then allocated to locations in the scene according to the salience in the computed map using a winner-takes-all principle. To avoid attention becoming "stuck" at the most salient location, a local, transient inhibition is applied to each attended location. Each iteration of the model—a winner-takes-all selection of the most salient location followed by inhibition at the attended location—effectively represents a relocation of attention.

Explanatory power of the salience map

Visual conspicuity models such as Itti and Koch's salience map can explain aspects of human attention allocation. The salience model described in Figure 1 can localize popout targets in a single iteration of the model. However, conjunction targets can take several iterations of the model before they are selected, and the number of iterations depends upon the number of distractors present (Itti & Koch, 2000). This serial search behavior with search times dependent upon the distractor set size mirrors human

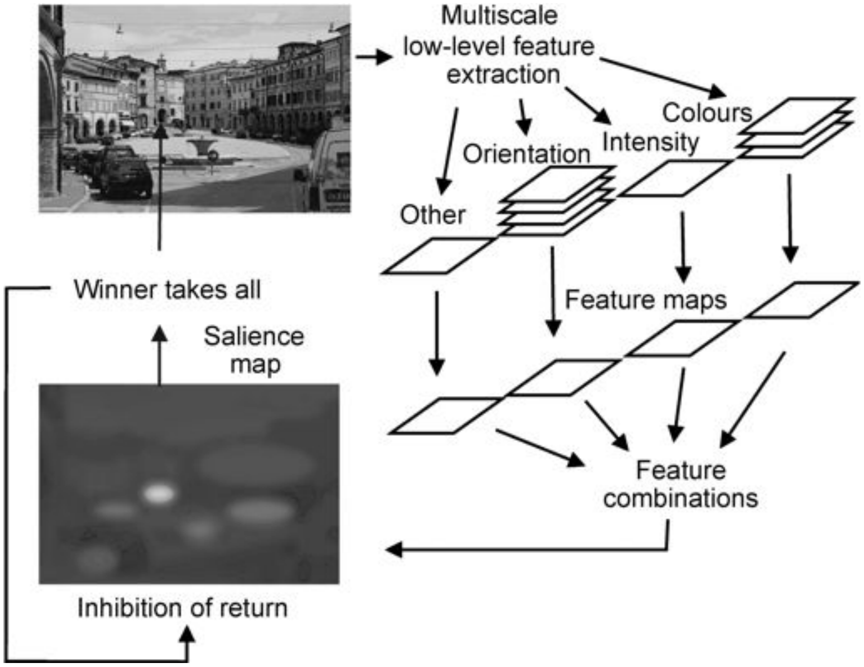

Figure 1. Schematic of Itti's salience model. Figure from Land and Tatler (2009); redrawn from Itti and Koch (2000).

search behavior. When presented with a complex photographic scene, the model predicts serial search behavior in which visually conspicuous locations are selected for "fixation," in a manner that appears superficially similar to human eye movement patterns.

Empirical evaluation of salience-based selection in complex scenes

It is clear that under certain circumstances image salience (or other feature-based models) can provide a good explanation of how humans orient their attention. However, the evidence so far discussed is derived almost exclusively from situations in which the stimulus is a simple visual search array or in which the target is uniquely defined by simple visual features. These studies provide a proof of principle for the notion that the visual system can select fixation targets on the basis of conspicuity. Indeed, the original goal of such models was to explain attentional capture rather than to provide a model of eye movements in complex scenes. It seems reasonable to ask whether the principles derived from simple paradigms might generalize to viewing more complex scenes. However, a real-world scene provides a much greater range of information types than these simple search displays. It is, therefore, important to empirically evaluate whether visual conspicuity contributes significantly to fixation selection when a greater range of information is available. Empirical evaluations of the salience model using complex, natural scenes show that more fixations lie within regions predicted by the salience model than would be expected by chance (e.g., Foulsham & Underwood, 2008) and salience at fixated locations is significantly higher than at control locations (e.g., Parkhurst, Law, & Niebur, 2002). Findings such as these are widespread, suggesting a correlation between low-level features in scenes and fixation selection by humans. However, as argued by previous researchers, these correlations alone should not be taken to imply any causal link between features and fixation placement (Henderson, 2003; Henderson, Brockmole, Castelhano, & Mack, 2007; Tatler, 2007).

Despite the widespread interest in this model and the considerable successes that it has had in predicting fixation selection at above-chance levels, it is important to consider just how much fixation behavior can be explained by a feature-based model of selection. Empirical evaluations of the extent of the difference in salience at fixated and control locations are informative in this respect. Using signal detection approaches, it is possible to consider not only whether statistically significant differences in the salience at fixated and control locations can be found but also the magnitude of these differences (see Tatler, Baddeley, & Gilchrist, 2005). Essentially, the technique can be used to determine the extent to which fixated and control locations can be discriminated on the basis of low-level feature information. The magnitude of the difference describes how well fixation selection can be described by low-level features. Such evaluations have found areas under the receiver–operator curve in the region of 0.55 to 0.65 (where 0.5 is chance), which suggests that the proportion of fixation behavior that can be accounted for by image salience is modest (e.g., Einhäuser, Spain, & Perona, 2008; Nyström & Holmqvist, 2008; Tatler et al., 2005). When the viewer's task is manipulated, this modest predictive power can disappear (e.g., Foulsham & Underwood, 2008; Henderson et al., 2007). The weak statistical support for low-level factors in fixation selection can be contrasted to the support offered for other factors in fixation selection. Using the same logic of attempting to discriminate between fixated and control locations on the basis of a particular source of information, areas under the receiver–operator curve for other factors can be compared to those found for image salience. Einhäuser, Spain et al. (2008) found that fixated and control locations can be better distinguished by object-level information than by image salience. Tatler and Vincent (2009) found that fixated and control locations can be better distinguished by biases in how we move our eyes than by image salience.

Despite these empirical shortcomings of the original implementation of the salience model (and of similar models), conspicuity-based accounts continue to feature prominently in much of the recent work on eye guidance (e.g., Xu, Yang, & Tsien, 2010; Yanulevskaya, Marsman, Cornelissen, & Geusebroek, 2010; Zehetleitner, Hegenloh, & Mueller, 2011; Zhao & Koch, 2011, and many others). Recent special issues of *Cognitive Computation* (Taylor & Cutsuridis, 2011) and of *Visual Cognition* (Tatler, 2009) reflect the continuing prominence of image salience and similar conspicuity-based factors in current research. Indeed, even recent emerging models often continue to retain a key role for visual conspicuity (e.g., Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Kanan, Tong, Zhang, & Cottrell, 2009), a point we will return to later in this article. However, this conspicuity-based class of computational model of eye guidance requires a set of assumptions that are conceptually and empirically problematic. In the section that follows, we highlight these assumptions and evaluate empirical evidence about their validity. Following this, we will consider the emerging models that overcome some of these limitations, including models that place less emphasis on visual conspicuity. We then consider situations in which conspicuity models may provide useful descriptions of human behavior.

Assumptions in models of scene viewing

Assumption 1: Pre-attentive features drive fixation selection

One of the essential assumptions behind salience models is that simple features are extracted pre-attentively at early levels of visual processing and that the spatial deviations of features from the local surround can, therefore, provide a basis for directing attention to regions of potential interest. While there exist statistically robust differences in the low-level content of fixated locations, compared with control locations (e.g., Mannan, Ruddock, & Wooding, 1997; Parkhurst et al., 2002; Reinagel & Zador, 1999), the magnitude of these differences tends to be small (see above), suggesting that the correlation between features and fixation is relatively weak. Furthermore, correlations are only found for small amplitude saccades (Tatler, Baddeley, & Vincent, 2006) and, crucially, disappear once the cognitive task of the viewer is manipulated (e.g., Foulsham & Underwood, 2008; Henderson et al., 2007). This does not mean that stimulus properties are unimportant. A high signal-to-noise ratio will make a variety of visual tasks such as search faster and more reliable. The question is whether simple stimulus features are analyzed pre-attentively and can, thus, form the basis for a bottom-up mechanism that can direct attention to particular locations. When walking around a real or virtual environment, feature-based salience offers little or no explanatory power over where humans fixate (Jovancevic, Sullivan, & Hayhoe, 2006; Jovancevic-Misic & Hayhoe, 2009; Sprague, Ballard, & Robinson, 2007; Turano, Geruschat, & Baker, 2003). In a virtual walking environment in which participants had to avoid some obstacles while colliding with others, image salience was not only unable to explain human fixation distributions but predicted that participants should be looking at very different scene elements (Rothkopf, Ballard, & Hayhoe, 2007). Humans looked at mainly the objects with only 15% of fixations directed to the background. In contrast, the salience model predicted that more than 70% of fixations should have been directed to the background. Thus, statistical evaluations of image salience in the context of active tasks confirm their lack of explanatory power. Hence, the correlations found in certain situations when viewing static scenes do not generalize to natural behavior. In ball sports, the shortcomings of feature-based schemes become even more obvious. Saccades are launched to regions where the ball will arrive in the near future (Ballard & Hayhoe, 2009; Land & McLeod, 2000). Crucially, at the time that the target location is fixated, there is nothing that visually distinguishes this location from the surrounding background of the scene. Even without quantitative evaluation, it is clear that no image-based model could predict this behavior. Similar targeting of currently empty locations is seen in everyday tasks such as tea making (Land, Mennie, & Rusted, 1999) and sandwich making (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003). When placing an object on the counter, people will look to the empty space where the object will be placed. As has been pointed out before, it is important to avoid causal inferences from correlations between features and fixations (Einhäuser & König, 2003; Henderson et al., 2007; Tatler, 2007), and indeed, higher level correlated structures such as objects offer better predictive power for human fixations (Einhäuser, Spain et al., 2008).

Assumption 2: There is a default bottom-up mode of looking

An implicit assumption in salience-based models is that there is a "default" task-free, stimulus-driven, mode of viewing and that vision for tasks is special in some way. The possibility of such a default viewing mode that can be overridden by other factors is discussed by several recent authors (e.g., Einhäuser, Rutishauser, & Koch, 2008; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006). Higher level factors are conceptualized as modulators of this basic mode of looking (see below). This assumption can be found at the heart of a wide range of studies and has motivated the use of "free-viewing" as a condition in studies of picture viewing, in an attempt to isolate task-free visual processing (e.g., Parkhurst et al., 2002). Here, the viewer is given no specific instructions during the experiment other than to look at the images. The assumption that "free-viewing" is a task-free condition for the viewer is questionable. It seems more likely that free-viewing tasks simply give the subject free license to select his or her own internal agendas (Tatler et al., 2005). A reasonable assumption about what people may be doing when asked to simply look at images is to recognize and remember the contents, but we cannot be sure of their internal priorities. Consequently, we are not studying viewing behavior while free of task, but rather we are studying viewing behavior when we have no real knowledge of what the viewer has chosen as the purpose of looking. Of course, the fixation behavior we engage in when "freely viewing" an image will be very different from that when engaged in a specific task such as search, but this does not imply that the former reflects any "default" task-free mode of looking. Not only is free-viewing a conceptually problematic task, but even when participants are freely viewing images, correlations between features and fixations are weak (Einhäuser, Spain et al., 2008; Nyström & Holmqvist, 2008; Tatler, 2007).

Assumption 3: Target selection from the map

Within the salience map framework, the decision about where to fixate arises from the computation of salience across the entire visual field, followed by a winner-takes-all process to select the most salient location. In order for this to allow more than one saccade, there is transient inhibition at each attended location. While this scheme seems like a reasonable computational solution to the problem of creating an iterative model of target selection, there exist at least two problems with this aspect of models.

Retinal sampling and eccentricity

In most accounts of salience-based schemes, the retinal position of image information is not accounted for; thus, decreasing retinal acuity in the periphery is overlooked (see Wischnewski, Belardinelli, & Schneider, 2010, for further information about the failure to consider peripheral sampling limits in most recent accounts of fixation selection). Some recent models do account for retinal sampling and we will consider these later. However, we first consider the problems associated with failing to account for this aspect of visual sampling. Vincent, Troscianko, and Gilchrist (2007) showed that feature coding becomes unreliable in the periphery once the variable resolution retina is taken into account. The feature maps and resultant salience maps generated when accounting for the variable spatial resolution outside the human fovea are very unlike those generated if uniform resolution sampling is assumed. This means that salience maps computed without taking into account the resolution of peripheral vision are biologically implausible. Moreover, salience computations that do account for spatial sampling heterogeneity fail to discriminate natural object targets in photographic scenes (Vincent et al., 2007). Retinal anisotropies in sampling result in tendencies to move the eyes in particular ways (Najemnik & Geisler, 2008). Humans tend to select nearby locations more frequently than distant locations as targets for their saccades (e.g., Bahill, Adler, & Stark, 1975; Gajewski, Pearson, Mack, Bartlett, & Henderson, 2005; Pelz & Canosa, 2001; Tatler et al., 2006). Similarly, when viewing pictures, horizontal saccades dominate (e.g., Bair & O'Keefe, 1998; Lappe, Pekel, & Hoffmann, 1998; Lee, Badler, & Badler, 2002; Moeller, Kayser, Knecht, & König, 2004). Incorporating these tendencies into models of fixation selection dramatically improves the predictive power of the model (Tatler & Vincent, 2009); indeed, these motor biases alone predicted fixation selection better than a model based on homogenous salience computation or homogenous edge feature extraction. We must, therefore, account for where information is in the retinal image rather than simply where peaks in any arbitrary whole-scene feature map might occur. Failing to account properly for where the winner is in the salience map results in distributions of saccade amplitudes that do not match human eye behavior (Figure 2).

It is interesting to compare the logic behind a winner-takes-all selection process and how we typically view the need to move the eyes. The general conception of the need to move the eyes is to bring the fovea to bear on information that is not fully available in the limited acuity peripheral vision. Thus, eye movements serve to provide new information about the surroundings, maximizing information gathering or reducing uncertainty about the visual stimulus (e.g., Najemnik & Geisler, 2005; Renninger, Verghese, & Coughlan, 2007). This contrasts with the winner-takes-all approach of selecting the region with the biggest signal as the next saccade target.

Inhibition of return

To allow attention to move on from the most salient peak in the salience map, transient inhibition of each attended location is included in the model. The inclusion of transient inhibition at attended locations is based on psychophysical experiments suggesting that there is an increase in latency when returning to recently attended locations (Klein, 1980, 2000; Klein & MacInnes, 1999; Posner & Cohen, 1984). However, empirical evidence suggests that there is no reduction in tendency to return to recently fixated locations when viewing photographic images (Smith & Henderson, 2009; Tatler & Vincent, 2008). Hooge, Over, van Wezel, and Frens (2005) found that while saccades back to the previously fixated location were preceded by longer fixation times (showing temporal IOR), there was no evidence of any decrease in the frequency of saccades back to previously fixated locations. Whether we observe something resembling inhibition of return or not depends upon the statistics of the dynamic environment being observed (Farrell, Ludwig, Ellis, & Gilchrist, 2010) and tasks that require refixations between objects show no evidence of IOR (Carpenter & Just, 1978). When specifically engaged in foraging behavior, refixations are rare (Gilchrist, North, & Hood, 2001), but it is not clear whether this is due to a low-level inhibitory mechanism, particular oculomotor strategies specific to foraging, or simply memory for previously visited locations (Gilchrist & Harvey, 2006). Indeed, Droll, Gigone, and Hayhoe (2007) demonstrated that locations are fixated more frequently if they are more likely to have the target.

The implementation of IOR in computational models of salience presents an obvious problem when attempting to simulate extended viewing. If the inhibition is long lasting, then refixations are impossible; if the inhibition is transient, then the model predicts cyclic eye movement behavior. Neither of these is compatible with human behavior. When viewing a picture of a face, participants will cycle around the triangle of central facial features (Yarbus, 1967). However, this cyclic behavior is not commonly found in more complex scenes and is certainly not an unavoidable consequence of looking at the same scene for more than a few seconds. Figure 3 compares fixation patterns for a human observer viewing a scene for an extended period to Itti and Koch's (2000) salience model inspecting the scene for the same number of fixations. Thus, it seems likely that a different mechanism is required to explain the transition from one fixation to

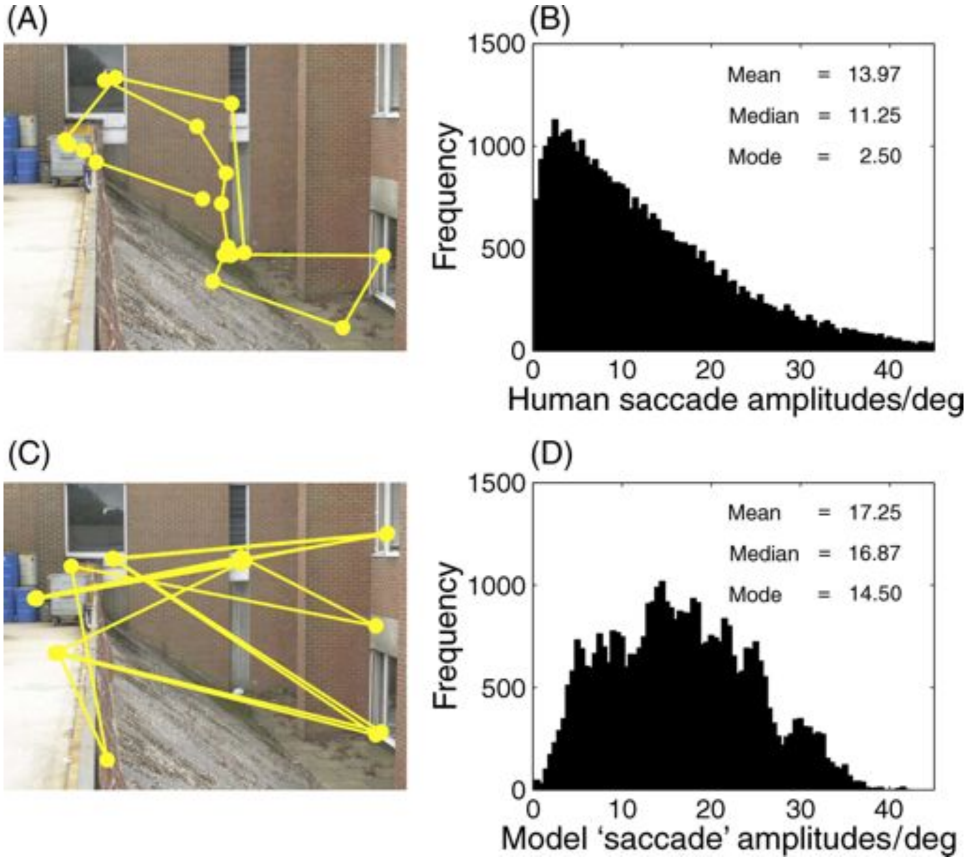

Figure 2. Saccade amplitudes from humans and the salience model. (A) Sample scan path from one participant looking at a photographic scene for 5 s. (B) Overall distribution of saccade amplitudes from humans looking at photographic scenes (*N* = 39,638 saccades). Data are taken from 22 participants, viewing 120 images for 5 s each. These data are drawn from the participants in Tatler and Vincent (2009) and full participant information can be found in this published paper. (C) Sample scan path from Itti's salience model. Simulation data are generated using the latest version of the saliency tool box downloaded from http://www.saliencytoolbox.net using the default parameters. Full details of the model can be found in Walther and Koch (2006). The simulation shown here was for the same number of "saccades" as recorded for the human data shown in (A). (D) Overall distribution of simulated saccade amplitudes from the salience model (*N* = 39,638 simulated saccades). Separate simulations were run for 22 virtual observers "viewing" the same 120 images as the human observers used in (B). For each image, the virtual observer made the same number of simulated saccades as the human observer had on that scene. The salience model produces larger amplitude saccades than human observers and does not show the characteristic positively skewed distribution of amplitudes.

the next. This is likely to be a more active mechanism, driven by a particular goal such as search or information acquisition.

Assumption 4: Time and target selection

A reasonable starting point when developing a model of eye movement behavior is to make the simplifying assumption that a first goal should be to explain spatial rather than temporal aspects of viewing behavior. It is becoming increasingly clear, however, that important information about the underlying mechanisms for saccade target selection also lies in the temporal domain. Fixation durations vary from a few tens of milliseconds to several hundred milliseconds and, in certain situations in real-world behaviors, can last for several seconds (Hayhoe et al., 2003; Land et al., 1999). Work on the importance of fixation duration in picture viewing is beginning to emerge (Henderson & Pierce, 2008; Henderson & Smith, 2009; Nuthmann, Smith, Engbert, & Henderson, 2010). Evidence from natural tasks emphasizes the need to consider fixation durations: fixation durations depend critically on the on the time required to acquire the necessary information for the current act (Droll, Hayhoe, Triesch, & Sullivan, 2005; Hayhoe, Bensinger, & Ballard, 1998; Hayhoe et al., 2003; Land et al., 1999). If fixation durations vary according to the information extraction requirements, then ignoring this source of information when evaluating and constructing models of eye guidance

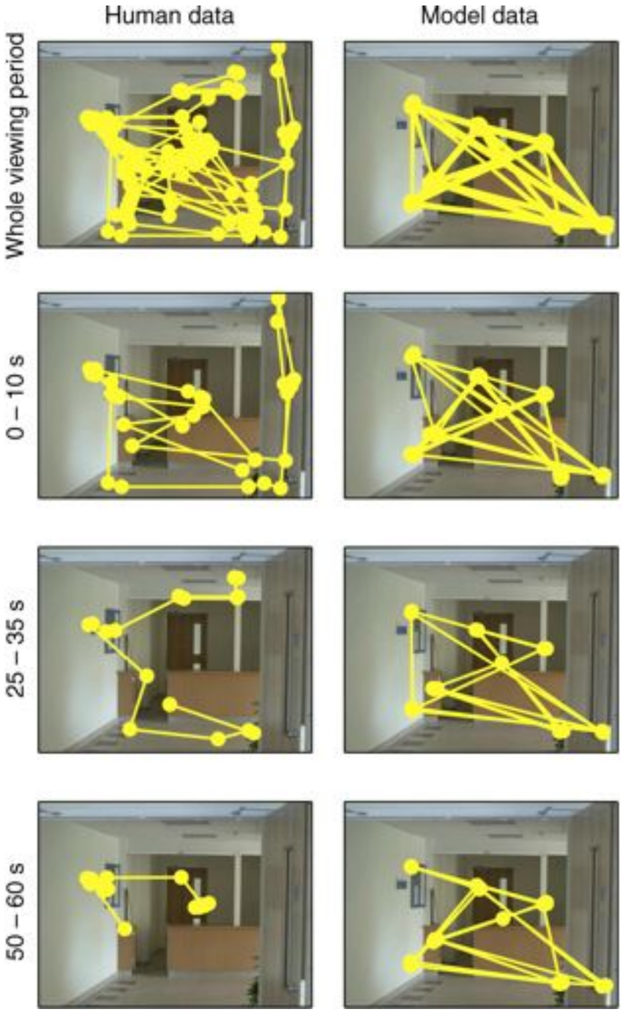

Figure 3. Behaviour during extended viewing for (left) a human observer and (right) the salience model. The human observer viewed the scene for 60 s with no instructions. The model simulated the same number of fixations as made by the observer during viewing (*N* = 129 fixations). Data are shown for the entire viewing period (top row). Note that the salience model simply cycles around a small number of locations, whereas the observer does not. The lower three rows show data divided into the first 10 s, the middle 10 s, and the final 10 s of viewing, with matched portions from the simulated sequence of fixations generated by the salience model. Simulation data were generated using the latest version of the saliency tool box downloaded from http://www.saliencytoolbox.net using the default parameters. Full details of the model can be found in Walther and Koch (2006).

misses a fundamental aspect of the control of attentional allocation.

Within the context of both simple laboratory paradigms and complex natural behavior, there is evidence that sequences of fixations may be planned in parallel (Zingale & Kowler, 1987). Unusually short fixations are often interpreted as implying that they are part of a pre-programmed sequence and the evidence for this in simple tasks is considerable. Express saccades are found predominantly when they are part of an ordered sequence of fixations in the same direction as each other (Carpenter, 2001). In the antisaccade task, erroneous prosaccades are frequently followed by short duration fixations before a corrective saccade to the opposite hemifield, a result that is interpreted as reflecting parallel programming of both the erroneous and corrective saccade before the error is initiated (Massen, 2004). The prevalence of very short fixation durations in some natural tasks (e.g., Hayhoe et al., 2003) could be interpreted in the same manner as short duration fixations have in simple viewing paradigms: as part of a pre-programmed sequence of eye movements.

From picture-viewing experiments, we know that the consistency between observers changes over time, such that different people will pick more similar locations for their first few fixations than for later fixations (Buswell, 1935). One possible explanation for this has been that the first few fixations in a viewing period are primarily driven by image properties (e.g., Parkhurst et al., 2002). However, subsequent studies have not supported this notion, suggesting that the correlations between features and fixations do not change over time (e.g., Nyström & Holmqvist, 2008; Tatler et al., 2005). Consequently, the changes in viewing behavior that are found across viewing time must come from higher level factors. Thus, time within a viewing epoch may prove to be an informative component for modeling the underlying target selection processes.

Assumption 5: Saccades precisely target locations for processing

Eye guidance in picture viewing is often assessed by comparing image statistics at fixated and control locations, extracting image properties over a small window (1–2 degrees of visual angle) centered at fixation (e.g., Parkhurst et al., 2002; Tatler et al., 2006; and many others). This approach assumes that the information at the center of gaze contains the intended target of each saccade. This seems plausible from the perspective of eye movement behavior in simple laboratory-based viewing paradigms. When required to fixate a small peripheral target, saccades that land short will almost always be corrected so that the fovea is brought to bear precisely upon the target (e.g., Becker, 1972, 1991; Carpenter, 1988; Deubel, Wolf, & Hauske, 1984; Kapoula & Robinson, 1986; Prablanc & Jeannerod, 1975). However, it is unclear whether such precision, evidenced by the presence of small corrective saccades, is a feature of natural image viewing (Tatler & Vincent, 2008).

In the context of more natural tasks, such precision may be unnecessary. When moving an object past an obstacle, getting the center of vision within about 3 degrees was sufficient: saccades that brought the foveae within 3 degrees of the obstacle were not corrected (Johansson, Westling, Backstrom, & Flanagan, 2001). Similarly, in tea making, saccades of amplitudes less than 2.5 degrees are very rare (Land et al., 1999). These findings suggest that getting the eye close to but not necessarily precisely on to a target is sufficient to serve many aspects of natural behavior, particularly when the objects being dealt with are large in the field of view. When making large relocations from one side of the room to another, gaze will sometimes be shifted in one large combined movement of eyes, head, and body. However, on other occasions, the relocation may involve one or more short duration fixations en route to the intended target (Land et al., 1999). In this case, the fixations made en route do not appear to land on any particular locations in the scene. It seems unlikely that these were intentionally targeted fixations; rather, they represent incidental stops during a planned relocation to the final, intended object. As such, the contents of these *en passant* fixations are unlikely to have played a key role in saccade targeting and modeling their visual characteristics of these fixations is likely to be misleading.

One question that arises when considering eye movements during natural behavior is whether all of the fixations we make are strictly necessary for serving the current behavioral goal or whether there is a certain amount of redundancy. Figure 4 shows an example of eye movements made while waiting for the kettle to boil. Many of these seem unlikely to be strictly necessary for the primary task and may reflect a variety of other purposes. It is entirely possible that these non-essential fixations are not targeted with the same precision or using the same selection criteria as other fixations. In general, the tight linking of fixations to the primary task will vary, depending on such factors as time pressure or behavioral cost. For example, fixations during driving may be more critical than when walking, where time is less critical. It is probably a mistake to think that every fixation must have an identifiable purpose and should be targeted with the same precision or selection criteria. It may be under conditions of reduced cognitive load that conspicuity-based fixations are most likely to be manifest.

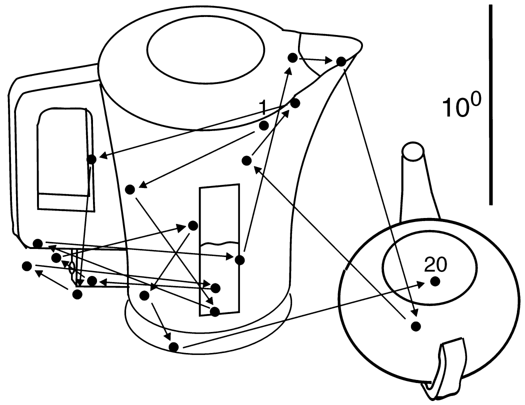

Figure 4. Profligacy in eye movement behavior. From Land and Tatler (2009).

Emerging alternative accounts

Not all of the issues identified above are fatal for existing approaches to the computational modeling of fixation selection. For example, incorporating peripheral acuity limits (Assumption 3) into models is tractable and several authors have incorporated aspects of this in computational models (e.g., Peters, Iyer, Itti, & Koch, 2005). Recent models emphasize the importance of inhomogeneous retinal sampling (e.g., Wischnewski et al., 2010; Zelinsky, 2008). Similarly models can incorporate information about when in a viewing epoch a fixation occurs or the duration of the fixation (Assumption 4). Models of fixation durations in scene viewing are beginning to emerge (Nuthmann et al., 2010).

Several recent models that attempt to incorporate higher level factors into accounts of fixation selection have been developed, a limitation of the original salience model that was recognized from the outset (Itti & Koch, 2000). One possibility is to suggest that top-down control is used to selectively weight the feature channels in the salience model to emphasize features that define the target of a search (Navalpakkam & Itti, 2005). A successful approach has been to incorporate prior knowledge of where particular objects are likely to be found in a scene in order to guide eye movements (Torralba, Oliva, Castelhano, & Henderson, 2006). In this model, a salience map of low-level conspicuity is modified by a contextual map of where particular targets are likely to occur. Contextual guidance and low-level features combine to provide good predictive power for human fixation distributions (Ehinger et al., 2009). In addition to using spatial expectations to refine the search space in a scene, prior knowledge of the appearance of objects of a particular class can be used (Kanan et al., 2009). Using the combination of a probabilistic appearance map, spatial contextual guidance and low-level feature salience can again be used to predict a sizeable fraction of human fixations (Kanan et al., 2009).

While the majority of recent computational models have retained a central place for low-level visual conspicuity, some models depart from this and build around alternative cores. The two most developed of these alternatives come from Wischnewski, Steil, Kehrer, and Schneider (2009; Wischnewski et al., 2010) and Zelinsky (2008). In Zelinsky's Target Acquisition Model, retinal inhomogeneity of sampling for the visual image is computationally implemented. Visual information is represented not as simple feature maps but as higher order derivatives, and knowledge of the target is incorporated. This model is successful at replicating human-like search of photographic scenes and the direction of the first saccade in a viewing epoch. The model can generalize to simpler stimuli and reproduce phenomena such as the center of gravity effect, where saccades land between potential targets (Zelinsky, Rao, Hayhoe, & Ballard, 1997).

Wischnewski et al.'s (2010) model builds upon Bundesen's (1990) Theory of Visual Attention. Wischnewski et al. attempt not only to move away from simple static visual features but also to overcome some of the problematic assumptions described above. In this model, retinal inhomogeneity of feature processing is included and the model centers around the integration of static features, dynamic features, proto-objects, and task. The emphasis in this model is not how static features are modified by other factors, but rather that the representation underlying saccade targeting is an integration across these levels of processing. These authors suggest that the different levels of information are integrated into an overall, retinotopic attention priority map. The notion of an attention priority map that integrates low-level and high-level cues has been suggested several times in the neurophysiological literature (Fecteau & Munoz, 2006). The neural implementations of such a priority map include the superior colliculus (McPeek & Keller, 2002), pulvinar (Robinson & Petersen, 1992), V1 (Li, 2002), V4 (Mazer & Gallant, 2003), LIP (Gottlieb, Kusunoki, & Goldberg, 1998), and the frontal eye field (Thompson & Bichot, 2005). Indeed, the emergence of a priority map to reflect the choice of either a target or an action in the posterior parietal cortex and subsequent areas is well supported and is clearly necessary to mediate targeted movements. It is also commonly accepted that both bottom-up and top-down signals contribute to such priority maps (Bichot & Schall, 1999; Gottlieb et al., 1998). The way that such activity emerges from the combination of stimulus and task context is unresolved, however, and beyond the scope of this review.

Wischnewski et al.'s notion of proto-objects as a key component in fixation selection is similar to recent suggestions by Henderson, Malcolm, and Schandl (2009). These authors suggested that selection proceeds from a representation of proto-objects ranked by cognitive relevance.

One other notable feature of Wischnewski et al.'s model is the incorporation of dynamic features. The need to account for dynamic stimuli and the inclusion of motion as a feature in models have been recognized for some time (see Dorr, Martinetz, Gegenfurtner, & Barth, 2010), and several versions of conspicuity-based models have incorporated dynamic features (e.g., Itti, 2005). However, it remains the case that the vast majority of studies of eye movements when viewing complex scenes use photographic images of real scenes, which necessarily fail to capture both the dynamics of real scenes and the complex, time-dependent, nature of task influences. Given this paradigmatic dominance of picture viewing, we will first consider whether this paradigm is a suitable domain in which to study eye guidance, before considering what can be learned from studying eye movements in dynamic and immersive contexts.

The picture-viewing paradigm

Can we learn about how we allocate gaze in natural environments and during natural behavior from how people look at pictures? While it is clear that models of picture viewing have utility for understanding tasks that involve looking at images on a computer monitor, it is important to consider whether we can use them to infer principles for fixation selection when behaving in natural environments. We are not the first to ask questions about the suitability of pictures as surrogates for real environments. Henderson (2003, 2006, 2007) has discussed this issue on several occasions. We wish to draw attention to two particular issues: biases introduced by the framing of the scene and effects of sudden scene onset.

The physical difference between photographs and real environments is obvious: the dynamic range of a photograph is much less than a real scene; many depth cues (stereo and motion parallax) are absent in static images; motion cues (both egomotion and external motion) are absent when viewing photographs; the observer's viewpoint in a still image is fixed and defined by the viewpoint of the photographer, which typically reflects compositional biases (Tatler et al., 2005). Not only is the field of view limited to the angle subtended by the display monitor, but also the scale of the image is typically undefined and depends on an inference by the observer. For example, a plate in a real setting might subtend 10 degrees, depending on the location of the observer, but in a picture of a scene it may subtend only a degree or two, and the subject must infer the viewpoint. This seems like a fairly sophisticated computation and is at odds with the essential idea of salience that low-level pre-attentive image features control gaze, with only limited perceptual analysis. Not only are the contents of photographs far removed from real images, but also placing the images within the bounds of the computer monitor's frame introduces strong biases in how the scenes are viewed. There is a strong tendency to fixate the center of images on a monitor irrespective of the scene's content (Tatler, 2007; Vincent, Baddeley, Correani, Troscianko, & Leonards, 2009). If, as Vincent et al. suggest, up to 34–56% of eye movements are best accounted for by a bias to fixate the screen center, then modeling the visual contents of these fixations will be very misleading.

Picture-viewing paradigms typically take the form of a series of trials characterized by the sudden onset of an image, followed by a few seconds of viewing, followed by the sudden offset of the image. Sudden onset may, in itself, influence inspection behavior. As discussed earlier, viewing patterns appear to change over time: there is inter-observer consistency in the locations fixated early in viewing, but this decreases with increasing viewing time (e.g., Parkhurst et al., 2002; Tatler et al., 2005). While this observation has given rise to a continuing debate about whether this arises because of an early dominance of salience followed by a switch toward more top-down control later in viewing, a more fundamental issue is what the implications of early differences are for the generalizability of findings. If viewing is different for the first few seconds after sudden scene onset (an observation that authors are in agreement about), then the selection criteria for these first few fixations are different for those later in viewing. The problem arises because there is no real-world analogue of the sudden onset of an entire scene, and it is known that the activity of neurons involved in target selection is very different for sudden onsets (Gottlieb et al., 1998). Even opening a door to a room is not like a sudden onset: here, the scene still emerges as the door opens. If we accept that sudden whole-scene onsets are peculiar to static scene paradigms, then the targeting decisions that underlie saccades made early in viewing periods may be specific to the sudden onset paradigms. Because such experiments typically only show scenes for a few seconds (in the region of 1–10 s in most studies), this could influence a sizeable fraction of the eye movements that are modeled.

It could be argued that the “purpose” of vision is very different when looking at a static scene to when engaging in real-world behavior. In natural tasks, a key goal of vision can be seen as extracting the visual information and coordinating the motor actions required to complete the task. However, when viewing photographic scenes, there is rarely a task that involves the active manipulation of objects in the environment. Rather, in static scene viewing, the task may be to search for a target, to remember the scene, or to make some judgement about the content of the scene. These classes of task are only a subset of the repertoire of behaviors we execute in the real world. Thus, the principles governing saccade targeting decisions in the tasks used in picture-viewing paradigms are most likely different from those used when engaged in active, real-world tasks.

Videos as surrogates for real-world settings

The shortcomings of static pictures as surrogates for real environments has been recognized by numerous investigators (e.g., Henderson, 2007; Shinoda, Hayhoe, & Shrivastava, 2001). As a result of this recognition, a growing number of studies are starting to use videos because these stimuli include dynamic information (e.g., Carmi & Itti, 2006; Itti, 2005; ’t Hart et al., 2009). Dynamic features can be strong predictors of eye movement behavior (Itti, 2005). However, this may not generalize to natural behavior because the frequent editorial cuts that are found in many movie sequences present an unusual and artificial situation for the visual system. Editorial cuts result in memorial and oculomotor disruptions to normal scene perception (Hirose, Kennedy, & Tatler, 2010). Moreover, such cuts result in behavior that is unlike how we view continuous movies with no cuts (Dorr et al., 2010; ’t Hart et al., 2009). When viewing continuous movies of a dynamic real-world environment, the predictive power of both static and dynamic feature cues was vanishingly small (Cristino & Baddeley, 2009). Thus, movie-style edited video clips may be problematic stimuli. It is also possible that the framing effects of the monitor continue to induce central biases to scene viewing that are ecologically invalid: while the central bias is weaker for continuous movies, it still remains and explains a considerable fraction of eye movement behavior (Cristino & Baddeley, 2009; Dorr et al., 2010; ’t Hart et al., 2009).

A role for visual conspicuity?

It should be reiterated at this point that the original goal of conspicuity models was not really to explain eye movements but rather to explain attentional capture, evaluating this by using eye movements. In this respect, such models were not really designed to explain eye movements in general and should not be expected to generalize to natural behavior. There is a large literature on attentional and oculomotor capture that we will not review here. In general, the findings of this literature are mixed. There is good evidence that specific stimuli such as sudden onsets, new objects, or motion transients have substantial power to attract attention (Franconeri & Simons, 2003; Gibson, Folk, Teeuwes, & Kingstone, 2008; Irwin, Colcombe, Kramer, & Hahn, 2000; Lin, Franconeri, & Enns, 2008; Theeuwes & Godijn, 2001). It is less clear whether certain classes of stimuli attract attention in an obligatory fashion, independently of the subject’s task set or ongoing cognitive goals (Jovancevic et al., 2006; Yantis, 1998). While much of natural behavior might be under task-driven control, there is clearly a need for a mechanism to capture attention and change the ongoing cognitive agenda. Many aspects of natural environments are unpredictable and there must be some mechanism to alert the observer to unexpected hazards. Our subjective impression that attention and gaze are reliably drawn to unusual stimuli or events in the environment argues for some mechanism like salience. It is a valid question whether salience models can work in these cases. The essential difficulty is that free viewing of static images is probably not a good paradigm either for attentional capture or for natural vision, as we have discussed. The problem in natural vision is that a stimulus that is salient in one context, such as peripheral motion with a stationary observer, may not be salient in another context, such as when the observer is moving and generating complex motion on the retina. To address this problem, Itti and Baldi (2006) suggested that "salient" events or locations are those that are unexpected or surprising, where surprise is defined as a statistical deviation from the recent history of visual stimuli. Surprising stimuli, therefore, correspond to statistical outliers in time, whereas salient stimuli are statistical outliers in space. A recent paper by Bruce and Tsotsos (2009) reflects this idea in the space domain by defining salience as a "surprisal" value or the extent to which a region differs from its neighborhood. Some kind of surprise mechanism is essential for attracting attention to stimuli that are important but not encompassed by the current task set. There is only a little work on the statistical basis for the formation of a surprise signal. Itti and Baldi conjecture that the visual system learns the statistics of images by estimating the distribution of parameters of probability distributions that can explain recent image feature data. In the context of video sequences, as subsequent image frames are processed, Bayesian inference updates the priors with the posterior from the previous frame. They measure surprise as the shift between the posterior and prior probabilities of model parameters. Itti and Baldi's model is a complex multi-parameter simulation of early visual processing and works on very short time scales (100s of ms). Thus, it is unlikely to reflect the long-term memory factors involved in natural behavior. Most scenes are highly familiar and observers have the opportunity to build extensive long-term memory representations built up over thousands of fixations. Brockmole and Henderson (2005, 2008) and Matsukura, Brockmole, and Henderson (2009) showed that subjects are more likely to fixate changes in scenes when they have previously viewed the scene for a few seconds. Uke-Karacan and Hayhoe (2008) showed that several minutes experience in a virtual environment led to increased fixations on changed objects in the scene. Thus, stimuli that are surprising with respect to a prior expectation might constitute a robust means of attracting attention.

It is, therefore, clear that there are circumstances in which conspicuity-based models of eye guidance and attention can provide explanations of human behavior. When the visual signal in the environment is large (as is the case in simple feature-based search arrays and sudden onset paradigms or when an unexpected event occurs), then this signal will drive eye movement behavior. It is an empirical question whether attentional capture by large signals, that is, the mechanisms of surprise, constitutes a significant portion of ordinary oculomotor behavior. Learned strategies such as searching for Stop signs at intersections can certainly deal with many of the vicissitudes of the natural world, but clearly some attention-getting mechanism is essential. Understanding how the visual world is coded in memory to form the basis of prior expectations and allow reliable detection of surprising stimuli is an important question that needs to be resolved. A related question is the extent to which mechanisms of surprise might be modulated by behavioral goals. For example, one can imagine that the visual system might have the task of looking for surprising stimuli as a priority in many circumstances, or alternatively, vision might only prioritize surprising stimuli when there is no other pressing demand. The answer to these questions would help determine the extent to which results from picture viewing might generalize to natural behavior.

Eye guidance in natural behavior

We have argued that the conspicuity-based theoretical models are unable to explain many aspects of human fixation behavior and that picture viewing (and perhaps movie viewing) is a problematic paradigm for understanding eye movement behavior. Given that a fundamental function of vision is to provide information necessary for survival, if we are to understand the principles that underlie fixation selection, we must consider eye movements in the context of behavioral goals, where the requirement is to seek out relevant information at the time when it is needed. Most contemporary models of fixation selection acknowledge the importance of accounting for cognitive control of eye movements. However, few engage with the need to consider visual selection as being fundamentally and intricately linked to action. One exception to this is Schneider's (1995) Visual Attention Model, which distinguishes "what" and "where" components of target selection, with the latter considering selection for action. Despite its conceptual and empirical strengths (Deubel & Schneider, 1996), the importance of selection for action in models of eye guidance has not featured prominently in more recent models.

Empirical evaluations show that conspicuity-based theoretical models lack explanatory power in the context of natural behavior (e.g., Rothkopf et al., 2007). Thus, we argue that conspicuity-based approaches are not a suitable theoretical framework for understanding eye movements in the context of natural behavior. The challenge, therefore, for this field is to develop a suitable theoretical alternative. Moreover, models that make empirically testable predictions of fixation selection are required. In the sections that follow, we first consider the key findings from studies of natural tasks that are common across multiple instances of behavior. Our aim in this section of the article is to bring together common findings from a range of different natural task settings in order to identify common principles for fixation selection rather than to provide extensive details on any one natural task. Understanding the common observations allows us to identify general principles that underlie eye movements in natural tasks. From these principles, it is clear that the issues that must be explained by any theory of natural eye guidance are rather different from those typically considered in current models. The principles identified here offer the essential elements from which theoretical models might be built.

Spatial coupling between gaze and behavioral goal

Cognitive control of eye movements was well established before the development of salience models (e.g., Buswell, 1935; Kowler, 1990; Yarbus, 1967). This case has been strengthened by more recent work in natural tasks. All studies of eye movements during natural behavior show that there is an intimate link between where we look and the information needed for the immediate task goals (Epelboim et al., 1995; Hayhoe et al., 2003; Land & Furneaux, 1997; Land et al., 1999; Patla & Vickers, 1997; Pelz & Canosa, 2001). The link between our behavioral goals and the allocation of overt visual attention is highlighted by the fact that when engaged in a natural task essentially all the fixations fall on task-relevant objects (Hayhoe et al., 2003; Land et al., 1999), whereas before beginning a task (such as sandwich making) the distribution of fixations between task-relevant and -irrelevant objects is about equal (Hayhoe et al., 2003; Rothkopf et al., 2007). The extent to which fixation placement is driven by the information-gathering requirements for an interaction with an object was demonstrated by Rothkopf et al. (2007) in an immersive virtual reality environment. Here, fixations on identical objects varied considerably depending upon whether the participant was attempting to approach or avoid the object (Figure 5). This result highlights the importance of understanding the function of each fixation for understanding fixation placement.

Similarity between different individuals

The intimate link between vision and action is reflected in the consistency that is observed between individuals who complete natural tasks. Different individuals show a high degree of consistency in where and when they look at informative locations while engaged in natural behaviors. Drivers look consistently at or close to the tangent point of a bend or the lane ahead, with around 50% of fixations made by three drivers falling within an area subtending only about 3 degrees in diameter (Land & Lee, 1994). Fixations on other pedestrians when walking are very consistent across individuals: despite a lack of any explicit instructions, there was a high degree of consistency in when and for how long oncoming pedestrians were fixated (Jovancevic-Misic & Hayhoe, 2009). When cutting a sandwich, subjects always fixate the initial point of contact with the knife and move their gaze along the locus of the cut, just ahead of the knife (Hayhoe et al., 2003). The similarity in fixation sequences of different individuals when taking the kettle to the sink to fill it is illustrated in Figure 6 (Land et al., 1999).

A similarly impressive degree of inter-observer consistency can be found when recording gaze behavior of

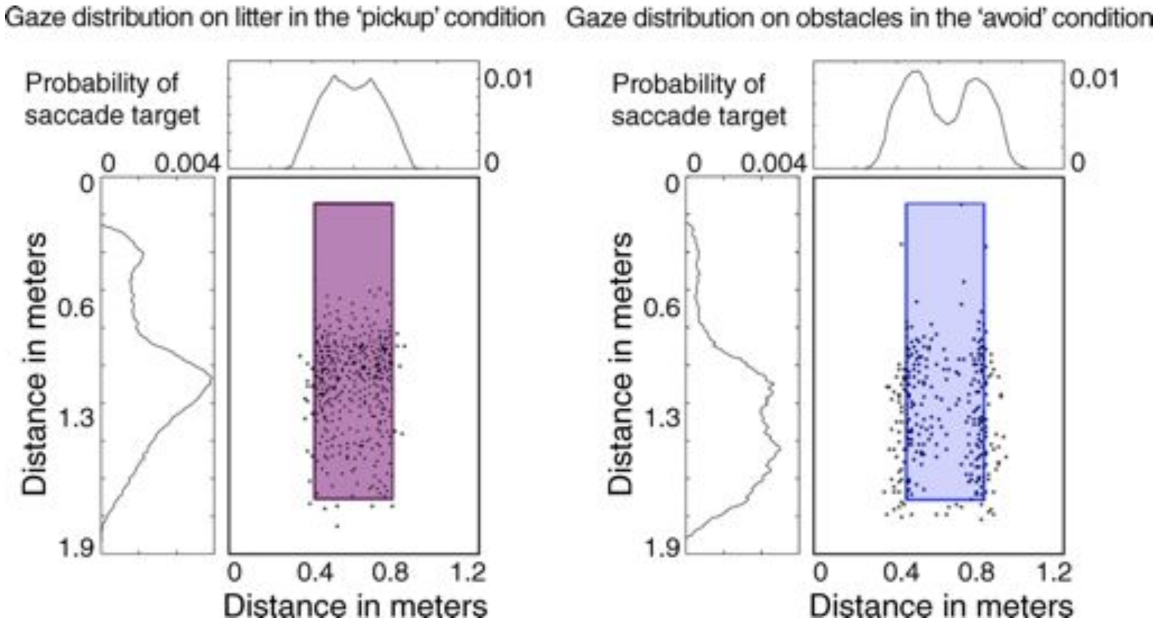

Figure 5. When subjects navigating a virtual environment are told to approach and pick up an object, their fixations tend to be centered on the object, but when the subjects are told to avoid the object, their fixations hug the edge of the object. The salience of the object is identical, but its associated uses have changed, dramatically changing the fixation distribution characteristics. From Rothkopf et al. (2007).

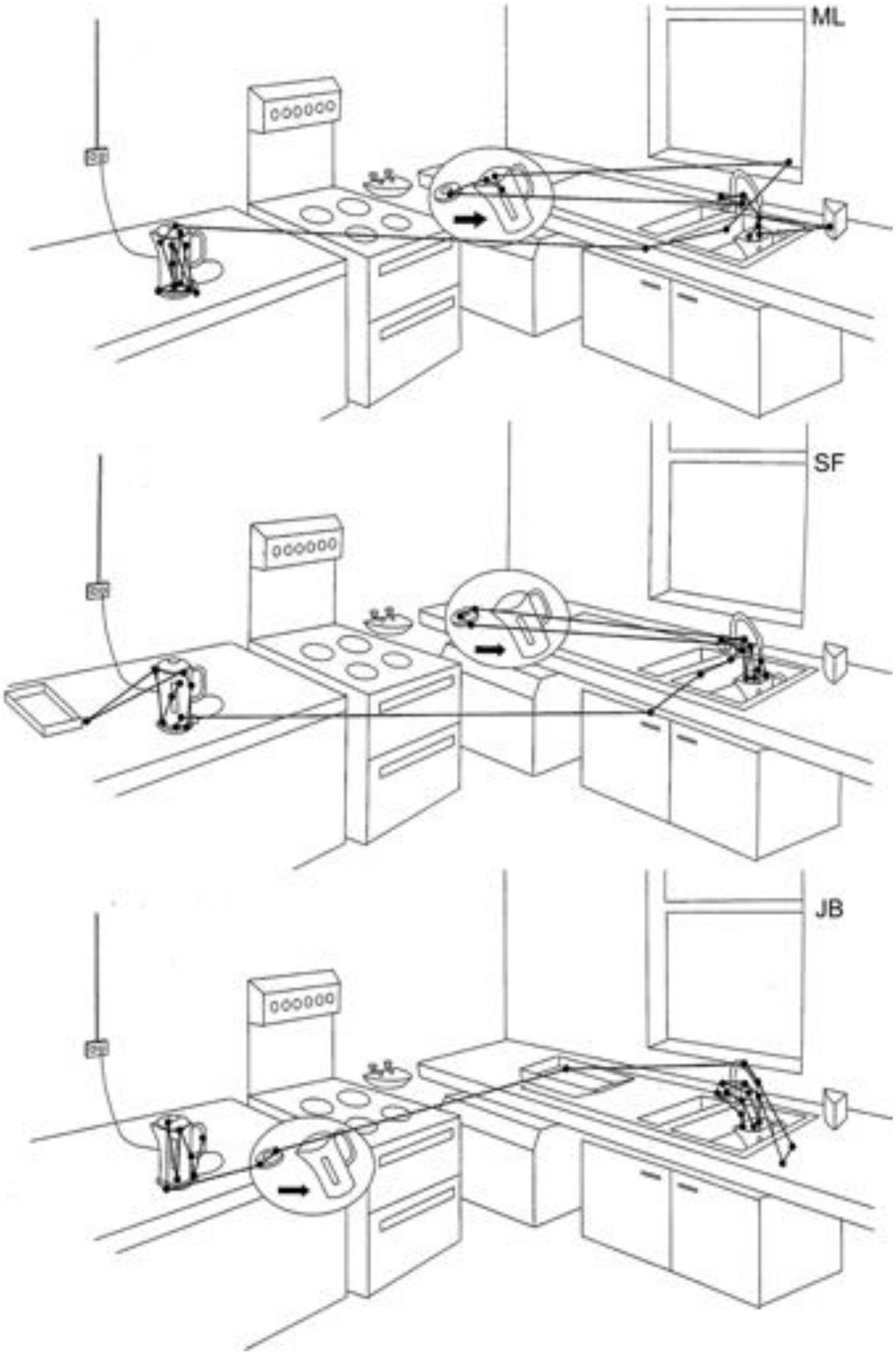

Figure 6. Scan patterns of three people taking the kettle to the sink in order to fill it prior to making tea (from Land et al., 1999).

observers watching a magician perform a trick. When making a cigarette and lighter "disappear," observers consistently fixate certain locations at the crucial moments in the performance (Kuhn & Tatler, 2005). This trick is based on the simple principle of distracting the observer while first the lighter and then the cigarette is dropped. At these crucial points, the observers consistently look to the opposite hand to that being used to drop the object (Kuhn & Tatler, 2005; Tatler & Kuhn, 2007). This misdirection to the inappropriate hand is tightly controlled in space and time, so that about 300 ms before the object is dropped, most participants will be looking at the same (inappropriate) location (Tatler & Kuhn, 2007). Of course, the question must be asked as to how the magician ensures the consistent misdirection of the audience at these crucial moments. These authors have shown that it is the magician's own direction of gaze that is the key to successfully misdirecting the observer (Kuhn, Tatler, & Cole, 2009).

A clear implication of the spatial and temporal consistency that is found between participants in natural tasks is that the decisions about where and when to allocate gaze must be governed by the same underlying principles in different individuals. Given the role of eye movements in gathering information to accomplish tasks, it makes sense that fixation patterns between individuals should be similar, as they will reflect the physical and dynamic properties of the environment as well as common behavioral goals. This stability in fixation patterns makes the investigation of natural behavior unexpectedly accessible to experimental investigation. The high degree of consistency found in gaze allocation in natural settings is in contrast to the often quite low consistencies found between individuals when viewing static scenes. Especially after a few seconds from the onset of a static scene, there is often very little consistency in where different observers fixate (Tatler et al., 2005). Consequently, one could argue that the principles underlying fixation selection during natural tasks may be more robust than those that researchers have tried to capture in models of eye guidance when viewing static scenes.

Timing of gaze shifts

A striking feature of natural behavior is that there is not only a tight spatial coupling between the eye and the target of the current motor act but also there is a tight temporal coupling between vision and action. This was elegantly demonstrated by Johansson et al. (2001), who measured the departure time of the eye relative to the hand as the subject maneuvered an object past an obstacle. Gaze moved onto the next target just at the point that the object cleared the obstacle. Similar time locking of the fixations and actions has been observed in driving (Land & Lee, 1994; Land & Tatler, 2001), making tea or sandwiches (Hayhoe et al., 2003; Land et al., 1999), music sight reading (Furneaux & Land, 1999), walking (Patla & Vickers, 2003), and reading aloud (Buswell, 1920). The ubiquity of this eye–action temporal coupling underlines the necessity to consider placement of the eyes in time as well as in space. Moreover, it may well be that the correct temporal placement of the eyes is more crucial to successful completion of behaviors than precise spatial placement and that skilled performance is as dependent upon the correct allocation of gaze in time as in space (Land & McLeod, 2000).

The roles of learning

Implicit in much of the research on natural tasks is the finding that people must learn what to look at and when (Chapman & Underwood, 1998; Land, 2004; Land & Furneaux, 1997; Land & Tatler, 2009). For example, in a virtual driving environment, Shinoda et al. (2001) asked participants to look for Stop signs while driving an urban route. Approximately 45% of fixations fell in the neighborhood of intersections during this task, and as might be expected from this, participants were more likely to detect Stop signs placed near intersections than those placed in the middle of a block. This result suggests that drivers have learned that traffic signs are more likely around intersections and so to preferentially allocate their gaze to these regions. At a more detailed level, people must learn the optimal location for the specific information they need. For example, where on the kettle a subject will look depends on what they need to do with that kettle. When waiting for it to boil, they will look mainly at the fill level indicator and switch (Figure 4). When placing it on its base, fixations will alternate between the bottom of the kettle and the fixings protruding from the base (on the work surface). When pouring water from the kettle, fixations will be made to the water stream in the receiving vessel. People must learn not only the locations at which relevant information is to be found but also the order in which the fixations must be made in order to accomplish the task. Thus, when making a sandwich an individual must locate the peanut butter and the bread before picking them up, pick up the knife before spreading, and so on. This means that a complete understanding of eye movements in natural behavior will require an understanding of the way that tasks are learned and represented in the brain, much of which presumably occurs over long time periods during development. In adult life, skills can be learned more rapidly, because they build on related skills already acquired.

In a study that explored the development of eye–hand coordination in a novel task, Sailer, Flanagan, and Johansson (2005) used a mouse-like control task to show that initially the eyes lagged behind action, apparently providing feedback information about the success of the last maneuver. However, once skilled at this task (after about 20 min), the eyes led the movement of the mouse cursor systematically by about 0.4 s, anticipating the next goal of the cursor on the screen. Similarly, learner drivers fixate just ahead of the car when cornering, whereas more experienced drivers look into the bend and fixate points on the road that will be reached as much as 3 s later, thus anticipating any need for future action (Land, 2006; Land & Tatler, 2009).

In stable environments, the observer needs only to update the locations of items that are moved or monitor items that are changing state. In dynamic environments, such as driving, walking, or in sports, more complex properties must be learned. In walking, humans need to know how pedestrians typically behave and how often to look at them. The fact that humans do indeed learn such statistics was demonstrated by Jovancevic-Misic and Hayhoe (2009). In a real walking setting, they were able to actively manipulate gaze allocation by varying the probability of potential collisions. Manipulation of the probability of a potential collision by a risky pedestrian (i.e., one with a past record of attempting collisions) was accompanied by a rapid change in gaze allocation. Subjects learned new priorities for gaze allocation within a few encounters and looked both sooner and longer at potentially dangerous pedestrians. This finding generalizes earlier work, for example, by He and Kowler (1989), showing the sensitivity of saccades to stimulus probability.

Further evidence for learning the dynamic properties of the environment comes from the fact that saccades are often proactive, that is, they are made to a location in a scene in advance of an expected event. In walking, subjects looked at risky pedestrians before they veered onto a collision course. In cricket, squash, and catching balls, players anticipate the bounce point of the ball by 100 ms or more (Land & McLeod, 2000). This ability to predict where the ball will bounce depends on previous experience of the ball's trajectory in combination with current sensory data. This suggests that observers have learned models of the dynamic properties of the world that can be used to position gaze in anticipation of a predicted event. Indeed, given neural delays between the eye and cortex, in time-critical behaviors such as driving and ball sports, action control must proceed on the basis of predictions rather than perceptions.

It is clear from these examples that the types and time scales of learning in the above examples vary considerably. Thus, any theoretical model must be able to explain learning across this broad range.

Reward-based models of gaze allocation

If we are to place learning at the center of theoretical accounts of eye guidance, it is important to consider how it might be implemented in the brain. The reward system, which has been implicated in a variety of aspects of learning, offers a suitable system for implementing the learning that is required for deploying gaze in natural behavior.

Neural substrates for learning gaze allocation in task execution

It has become increasingly clear that the brain's internal reward mechanisms are intimately linked to the neural machinery controlling eye movements. Schultz et al. have shown that dopaminergic neurons in the basal ganglia signal the reward expected from an action. The role of dopamine in expected reward is signaled as it is handed out in anticipation of the result of a behavior (e.g., Schultz, Tremblay, & Hollerman, 2000). Sensitivity to reward is manifest throughout the saccadic eye movement circuitry. Caudate cell responses reflect both the target of an upcoming saccade and the reward expected after making the movement (Hikosaka, Nakamura, & Nakahara, 2006). Saccade-related areas in the cortex (LIP, FEF, SEF, and DLPF) all exhibit sensitivity to reward (Dorris & Glimcher, 2004; Glimcher, 2003; Glimcher, Camerer, Fehr, & Poldrack, 2009; Platt & Glimcher, 1999; Stuphorn & Schall, 2006; Stuphorn, Taylor, & Schall, 2000; Sugrue, Corrado, & Newsome, 2004). The neurons involved in saccadic targeting respond in a graded manner to both the amount of expected reward and the probability of a reward in the period prior to execution of the response. Sensitivity to both these variables is critical for learning and, consequently, for linking fixation patterns to task demands. The cortical saccade-related areas converge on the caudate nucleus in the basal ganglia, and the cortical–basal ganglia–superior colliculus circuit appears to regulate the control of fixation and the timing of planned movements. Such regulation is a critical requirement for task control of fixations.

The relevance of the neurophysiological work on reward may not be immediately obvious for ordinary human behavior. In neurophysiological paradigms, usually a primary reward such as juice or a raisin is delivered after the animal performs an action. This, of course, does not happen in real life when one makes an eye movement. However, eye movements are for the purpose of obtaining information, and this information is used to achieve behavioral goals, such as making a sandwich, that are ultimately important for survival. Thus, visual information acquired during a fixation can be thought of as a secondary reward and can mediate learning of gaze patterns by virtue of its ultimate significance for adaptation and survival. Indeed, several researchers have quantified the intrinsic reward associated with looking at particular visual stimuli. Deaner, Khera, and Platt (2005) and Shepherd, Deaner, and Platt (2006) measured how much liquid reward monkeys were willing to give up in order to obtain visual information about members of their social group. In this case, liquid is the measurable, external equivalent of an internal reward resulting from gaze. Thus, the dopaminergic machinery appears to be intimately related to the sensitivity of eye movement target selection to behavioral outcomes.

Modeling eye movements using reward

The reward sensitivity of the eye movement circuitry provides the neural underpinnings for reinforcement learning models of behavior (Montague, Hyman, & Cohen, 2004; Schultz, 2000). The mathematics of reinforcement learning is potentially useful for understanding how complex gaze patterns might be generated (Sutton & Barto, 1998). Dopaminergic cells signal the reward expected from an action, and reinforcement learning models are pertinent because they allow an agent to learn what actions or action sequences will lead to reward in the future. Given a set of possible states, and actions that might be associated with those states, reinforcement learning algorithms allow an agent to learn a policy for selecting actions that will ultimately maximize reward.

There have been few attempts to model the eye movements observed in complex behavior. However, one

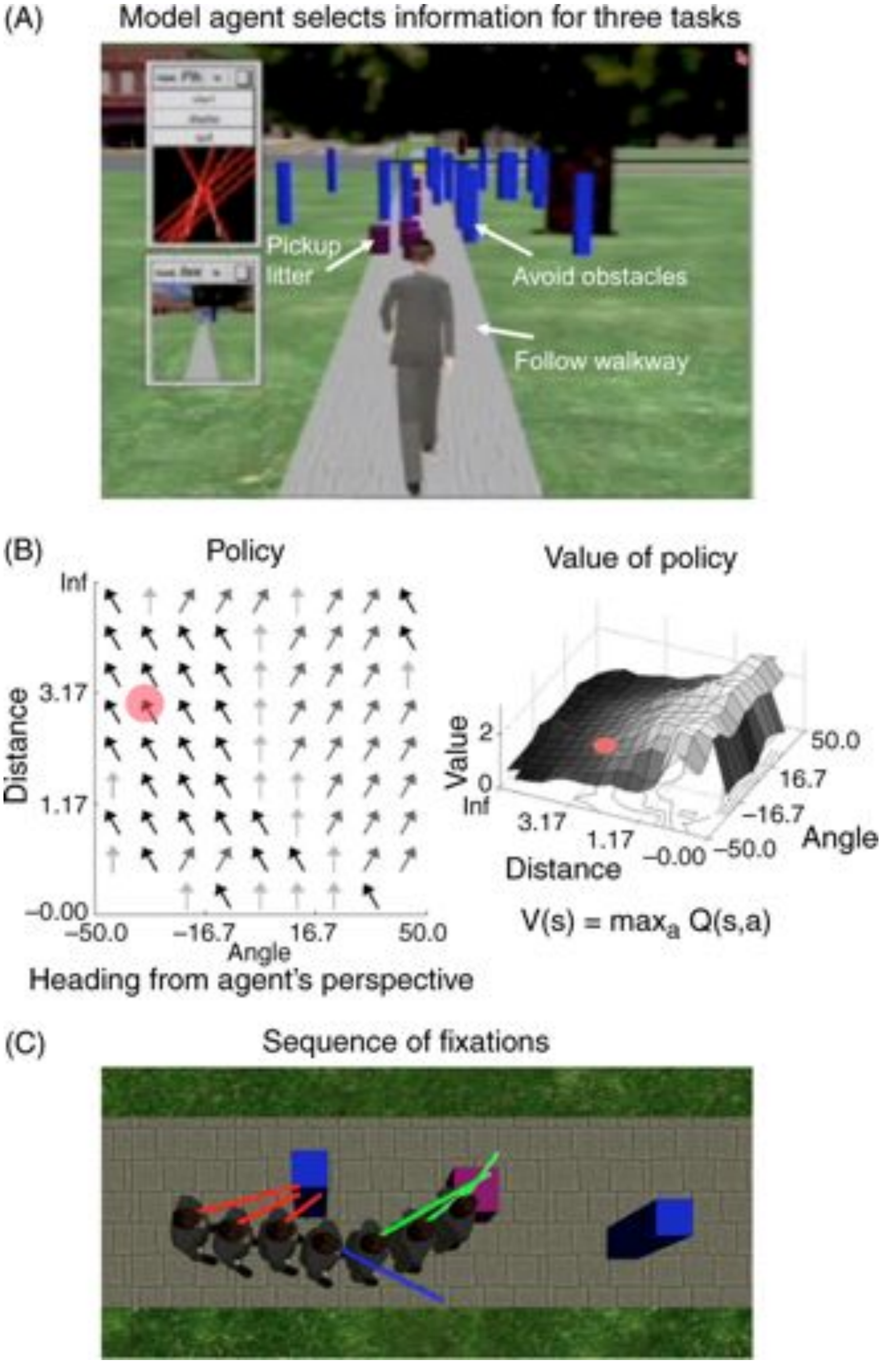

Figure 7. The model of Sprague et al. (2007). (A) A virtual agent in a simulated walking environment. The agent must extract visual information from the environment in order to do three subtasks: staying on the sidewalk, avoiding blue obstacles, and picking up purple litter objects (achieved by contacting them). The inset shows the computation for staying on the path. The model agent learns how to deploy attention/gaze at each time step. (B) The agent learns a policy for choosing an action, given the current state information from gaze for a given task. Each action has an associated value, and the agent chooses the option with the highest value. (C) Seven time steps after learning. The agent chooses the task that reduces uncertainty of reward the most. The red lines indicate that the agent is using visual information to avoid the obstacle. The blue line indicates that the agent is using information about position on the sidewalk, and the green lines show the agent using vision to intersect the purple litter object.

such model, by Sprague et al. (2007; Figure 7), shows how a simulated agent in a virtual environment can learn to allocate gaze to avoid obstacles and control direction in walking (see also Ballard & Hayhoe, 2009; Rothkopf & Ballard, 2009; Rothkopf et al., 2007). The model assumes that visual computations required in the real world can be broken down into a set of subtasks, or modules, such as controlling direction, avoiding obstacles, and so on. Each subtask is associated with some reward value. For example, obtaining visual information that allows avoidance of an obstacle presumably provides secondary reward. These authors have provided a computational account of how we can successfully distribute attention and gaze between these visual subtasks in a dynamic environment. Their chosen paradigm involves walking along a virtual path with three simultaneous tasks: stay on the path, avoid obstacles, and pick up “litter.” The proposed model assumes that we can only attend to one location at any moment in time and that our uncertainty about unattended tasks grows over time. The decision about which task to attend to is based on the expected reward of switching attention to another task, evaluated every 300 ms. To choose between ongoing competing tasks, in their model, uncertainty increases (together with an attendant cost) when gaze is withheld from an informative scene location. The model assumes that eye movements are selected to maximize reward by reducing uncertainty that could result in suboptimal actions. Framing the decision about where to look in terms of uncertainty reduction has been effective in explaining aspects of static scene viewing (Najemnik & Geisler, 2005, 2008; Renninger, Coughlan, & Vergheese, 2005) as well as dynamic scene viewing.

Reward is a central component of recent applications of statistical decision theory to understanding control of body movements. In this approach, the concepts of *reward* (costs and benefits of the outcome of the action), *uncertainty* (of both sensory state and outcome), and *prior knowledge* (probability distributions associated with world states) are central to understanding sensory-motor behavior (e.g., Tassinari, Hudson, & Landy, 2006; Trommershäuser, Maloney, & Landy, 2008). When reward is externally defined (e.g., by monetary reward), it has been shown that subjects making rapid hand movements learn a complicated spatially distributed target reward system and behave in a nearly optimal manner to maximize reward (e.g., Seydell, McCann, Trommershäuser, & Knill, 2008; Trommershäuser, Maloney, & Landy, 2003). Similar targeting experiments using saccadic eye movements with monetary rewards and losses showed that reward affected saccadic targeting, although stimulus strength also affected the movements particularly at short latency (Stritzke, Trommershäuser, & Gegenfurtner, 2009). Other evidence for the role of reward in saccade targeting has been demonstrated by Navalpakkam, Koch, Rangel, and Perona (2010) who showed that subject’s saccade behavior in a visual search is consistent with an ideal Bayesian observer, taking into account both rewards and stimulus detectability. Thus, it is plausible that the patterns of eye movements observed in the natural world takes into account both the reward structure of the environment and stimulus uncertainty (Trommershäuser, Glimcher, & Gegenfurtner, 2009).

Models that use reward and uncertainty as central components are in their relative infancy and are not yet at the stage of providing a computational model that explains eye movements across multiple instances of natural behavior. However, such models offer the potential to include ubiquitous aspects of fixation selection that cannot be explained within conspicuity-based models. For example, the common tendency to look into empty spaces in anticipation of an event is very problematic for conspicuity models but can be explained if gaze allocation is based on expected (secondary) reward. Developing eye guidance models based on reward is a difficult endeavor because it essentially requires a model of task execution. Not only this, but as we have seen the types and time scales of learning that we must be able to model vary considerably. At present, models based on reward focus on the immediate time scale of the current behavioral situation but reflect the outcome of longer time scales of learning. Reinforcement learning, for example, presumably functions on a developmental time scale, so adults’ gaze patterns would reflect the end product of such models. Many fundamental questions require empirical support. For example, is it appropriate to model behavior as a set of semi-independent subtasks? This assumption of behavioral modules is critical to make the problem computationally tractable (Rothkopf & Ballard, 2010), but it is not known whether it is a good model of sensory-motor behavior. However, it is clear that reward is intrinsic to many aspect of cortical function (Glimcher et al., 2009) so the reward-based approach seems likely to provide a key building block from which to develop future theories and models of gaze behavior.

Conclusions

Investigation of eye guidance in scenes has been driven largely by studies of static scene viewing. The latest models of this behavior can be thought of as modifications to the image salience framework, where a core bottom-up mode of looking is modified by various high-level constraints. We argue that the basic assumptions at the heart of such studies are problematic if we wish to try to generalize these models to how gaze is allocated in natural behavior. That is, models developed from static scene-viewing paradigms may be adequate models of how we look at pictures but are unlikely to generalize to gaze behavior in other situations. Developing computational models of gaze allocation that can generalize across many instances of natural behavior is a difficult goal. However, we see already from studies of gaze selection in natural behavior that there is a consistent set of principles underlying eye guidance involving behavioral relevance, or reward, uncertainty about the state of the environment, and learned models of the environment, or priors. These factors control the decision mechanisms that govern what we should attend to on the basis of where we will gain information for fulfilling the current behavioral goals.

Acknowledgments

The authors would like to thank Werner Schneider and an anonymous reviewer for their helpful comments and suggestions. We thank Brian Sullivan for his comments on an earlier draft. Work in this article was supported by two NIH Grants (NIH R01 EY05729 and NIH R01 EY019174) and two grants from the Leverhulme Trust (F/00 143/O and F/00 143/J). We thank Sharon Scrafton for supplying the tea making footage used in the icon for this article and Sarah Walls for contributing the eye movement data used in Figure 3.

Commercial relationships: none.

Corresponding author: Benjamin W. Tatler.
Email: b.w.tatler@dundee.ac.uk.
Address: School of Psychology, University of Dundee, Dundee DD1 4HN, UK.

References

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology, 14,* 468–469.

Bair, W., & O'Keefe, L. P. (1998). The influence of fixational eye movements on the response of neurons in area MT of the macaque. *Visual Neuroscience, 15,* 779–786.

Ballard, D. H., & Hayhoe, M. M. (2009). Modeling the role of task in the control of gaze. *Visual Cognition, 17,* 1185–1204.

Becker, W. (1972). The control of eye movements in the saccadic system. *Bibliotheca Ophthalmologica, 82,* 233–243.

Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Vision & visual dysfunction: Eye movements* (vol. 8, pp. 95–137). Basingstoke, UK: Macmillan.

Bichot, N. P., & Schall, J. D. (1999). Effects of similarity and history on neural mechanisms of visual selection. *Nature Neuroscience, 2,* 549–554.

Brockmole, J. R., & Henderson, J. M. (2005). Prioritization of new objects in real-world scenes: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 857–868.

Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Effects of object-scene consistency. *Visual Cognition, 16,* 375–390.

Bruce, N. D. B., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision, 9*(3):5, 1–24, http://www.journalofvision.org/content/9/3/5, doi:10.1167/9.3.5. [PubMed] [Article]

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547.

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.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46,* 4333–4345.

Carpenter, P. A., & Just, M. A. (1978). Eye fixations during mental rotation. In J. Senders, R. Monty, & D. Fisher (Eds.), *Eye movements and psychological functions II* (pp. 115–133). Hillsdale, NJ: Erlbaum.

Carpenter, R. H. S. (1988). *Movements of the eyes* (2nd ed.). London: Pion.

Carpenter, R. H. S. (2001). Express saccades: Is bimodality a result of the order of stimulus presentation? *Vision Research, 41,* 1145–1151.

Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception, 27,* 951–964.

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

Deaner, R. O., Khera, A. V., & Platt, M. L. (2005). Monkeys pay per view: Adaptive valuation of social images by rhesus macaques. *Current Biology, 15,* 543–548.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36,* 1812–1837.

Deubel, H., Wolf, W., & Hauske, G. (1984). The evaluation of the oculomotor error signal. In A. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 55–62). Amsterdam, The Netherlands: North Holland.

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, http://www.journalofvision.org/content/10/10/28, doi:10.1167/10.10.28. [PubMed] [Article]

Dorris, M.-C., & Glimcher, P.-W. (2004). Activity in posterior parietal cortex is correlated with the subjective desirability of an action. *Neuron, 44,* 365–378.

Droll, J. A., Gigone, K., & Hayhoe, M. M. (2007). Learning where to direct gaze during change detection. *Journal of Vision, 7*(14):6, 1–12, http://www.journalofvision.org/content/7/14/6, doi:10.1167/7.14.6. [PubMed] [Article]

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1416–1438.

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,* 945.

Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience, 17,* 1089–1097.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*(2):2, 1–19, http://www.journalofvision.org/content/8/2/2, doi:10.1167/8.2.2. [PubMed] [Article]

Einhäuser, 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. [PubMed] [Article]

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

Farrell, S., Ludwig, C. J. H., Ellis, L. A., & Gilchrist, I. D. (2010). The influence of environmental statistics on inhibition of saccadic return. *Proceedings of the National Academy of Sciences, 107,* 929–934.

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences, 10,* 382–390.

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, http://www.journalofvision.org/content/8/2/6, doi:10.1167/8.2.6. [PubMed] [Article]

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65,* 999–1010.

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 B: Biological Sciences, 266,* 2435–2440.

Gajewski, D. A., Pearson, A. M., Mack, M. L., Bartlett, F. N., & Henderson, J. M. (2005). Human gaze control in real world search. In L. Paletta, J. K. Tsotsos, E. Rome, & G. W. Humphreys (Eds.), *Attention and performance in computational vision* (pp. 83–99). Heidelberg, Germany: Springer-Verlag.

Gibson, B., Folk, C., Teeuwes, J., & Kingstone, A. (2008). Introduction to special issue on attentional capture. *Visual Cognition, 16,* 145–154.

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition, 14,* 704–715.

Gilchrist, I. D., North, A., & Hood, B. (2001). Is visual search really like foraging? *Perception, 30,* 1459–1464.

Glimcher, P. (2003). The neurobiology of visual-saccadic decision making. *Annual Review of Neuroscience, 26,* 133–179.

Glimcher, P., Camerer, C., Fehr, E., & Poldrack, R. (2009). *Neuroeconomics: Decision making and the brain*. London: Academic Press.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature, 391,* 481–484.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research, 38,* 125–137.

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):6, 49–63, http://www.journalofvision.org/content/3/1/6, doi:10.1167/3.1.6. [PubMed] [Article]

He, P. Y., & Kowler, E. (1989). The role of location probability in the programming of saccades—Implications for center-of-gravity tendencies. *Vision Research, 29,* 1165–1181.

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

Henderson, J. M. (2006). Eye movements. In C. Senior, T. Russell, & M. Gazzaniga (Eds.), *Methods in mind* (pp. 171–191). Cambridge, MA: MIT Press.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16,* 219–222.

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.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin and Review, 16,* 850–856.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15,* 566–573.

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition, 17,* 1055–1082.

Hikosaka, O., Nakamura, K., & Nakahara, H. (2006). Basal ganglia orient eyes to reward. *Journal of Neurophysiology, 95,* 567–584.

Hirose, Y., Kennedy, A., & Tatler, B. W. (2010). Perception and memory across viewpoint changes in moving images. *Journal of Vision, 10*(4):2, 1–19, http://www.journalofvision.org/content/10/4/2, doi:10.1167/10.4.2. [PubMed] [Article]

Hooge, I. T. C., Over, E. A. B., van Wezel, R. J. A., & Frens, M. A. (2005). Inhibition of return is not a foraging facilitator in saccadic search and free viewing. *Vision Research, 45,* 1901–1908.

Irwin, D. E., Colcombe, A. M., Kramer, A. F., & Hahn, S. (2000). Attentional and oculomotor capture by onset, luminance and color singletons. *Vision Research, 40,* 1443–1458.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12,* 1093–1123.

Itti, L., & Baldi, P. (2006). Bayesian surprise attracts human attention. In Y. Weiss, B. Schölkopf, and J. Platt (Eds.), *Advances in Neural Information Processing Systems, (NIPS 2005)* (vol. 18, pp. 547–554). Cambridge, MA: MIT Press.

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

Johansson, R. S., Westling, G. R., Backstrom, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *Journal of Neuroscience, 21,* 6917–6932.

Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). Control of attention and gaze in complex environments. *Journal of Vision, 6*(12):9, 1431–1450, http://www.journalofvision.org/content/6/12/9, doi:10.1167/6.12.9. [PubMed] [Article]

Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *Journal of Neuroscience, 29,* 6234–6238.

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

Kapoula, Z., & Robinson, D. A. (1986). Saccadic undershoot is not inevitable: Saccades can be accurate. *Vision Research, 26,* 735–743.

Klein, R. M. (1980). Does oculomotor readiness mediate cognitive control of visual attention? In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 259–276). Hillsdale, NJ: Lawrence Erlbaum.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences, 4,* 138–147.

Klein, R. M., & MacInnes, J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science, 10,* 346–352.

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

Kowler, E. (1990). The role of visual and cognitive processes in the control of eye movement. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 1–70). Amsterdam, The Netherlands: Elsevier.

Kuhn, G., & Tatler, B. W. (2005). Magic and fixation: Now you don't see it, now you do. *Perception, 34,* 1155–1161.

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition, 17,* 925–944.

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research, 159,* 151–160.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research, 25,* 296–324.

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

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369,* 742–744.

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

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28,* 1311–1328.

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology, 11,* 1215–1220.

Land, M. F., & Tatler, B. W. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. Oxford, UK: Oxford University Press.

Lappe, M., Pekel, M., & Hoffmann, K. P. (1998). Optokinetic eye movements elicited by radial optic flow in the macaque monkey. *Journal of Neurophysiology, 79,* 1461–1480.

Lee, S. P., Badler, J. B., & Badler, N. I. (2002). Eyes alive. *ACM Transactions on Graphics, 21,* 637–644.

Li, Z. P. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences, 6,* 9–16.

Lin, J. Y., Franconeri, S., & Enns, J. T. (2008). Objects on a collision path with the observer demand attention. *Psychological Science, 19,* 686–692.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision, 11,* 157–178.

Massen, C. (2004). Parallel programming of exogenous and endogenous components in the antisaccade task. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57,* 475–498.

Matsukura, M., Brockmole, J. R., & Henderson, J. M. (2009). Overt attentional prioritization of new objects and feature changes during real-world scene viewing. *Visual Cognition, 17,* 835–855.

Mazer, J. A., & Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search: Evidence for a ventral stream visual salience map. *Neuron, 40,* 1241–1250.

McPeek, R. M., & Keller, E. L. (2002). Superior colliculus activity related to concurrent processing of saccade goals in a visual search task. *Journal of Neurophysiology, 87,* 1805–1815.

Moeller, G. U., Kayser, C., Knecht, F., & Konig, P. (2004). Interactions between eye movement systems in cats and humans. *Experimental Brain Research, 157,* 215–224.

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioral control. *Nature, 431,* 760–767.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434,* 387–391.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision, 8*(3):4, 1–14, http://www.journalofvision.org/content/8/3/4, doi:10.1167/8.3.4. [PubMed] [Article]

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45,* 205–231.

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 5232–5237.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117,* 382–405.

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. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 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,* 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, 48,* 133–138.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research, 41,* 3587–3596.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45,* 2397–2416.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 400,* 233–238.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 531–554). Hillsdale, NJ: Erlbaum.

Prablanc, C., & Jeannerod, M. (1975). Corrective saccades: Dependence of retinal reafferent signals. *Vision Research, 15,* 465–469.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network—Computation in Neural Systems, 10,* 341–350.

Renninger, L. W., Coughlan, J., & Vergheese, P. (2005). An information maximization model of eye movements. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (vol. 17, pp. 1121–1128). Cambridge, MA: MIT Press.

Renninger, L. W., Vergheese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision, 7*(3):6, 1–17, http://www.journalofvision.org/content/7/3/6, doi:10.1167/7.3.6. [PubMed] [Article]

Robinson, D. L., & Petersen, S. E. (1992). The pulvinar and visual salience. *Trends in Neurosciences, 15,* 127–132.

Rothkopf, C. A., & Ballard, D. H. (2009). Image statistics at the point of gaze during human navigation. *Visual Neuroscience, 26,* 81–92.

Rothkopf, C. A., & Ballard, D. H. (2010). Credit assignment in multiple goal embodied visuomotor behavior. *Frontiers in Psychology, 1,* 173.

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, http://www.journalofvision.org/content/7/14/16, doi:10.1167/7.14.16. [PubMed] [Article]

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye–hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience, 25,* 8833–8842.

Schneider, W. X. (1995). VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition, and space-based motor action. *Visual Cognition, 2,* 331–375.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature reviews: Neuroscience, 1,* 199–207.

Schultz, W., Tremblay, L., & Hollerman, J. R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex, 10,* 272–283.

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.

Shepherd, S. V., Deaner, R. O., & Platt, M. L. (2006). Social status gates social attention in monkeys. *Current Biology, 16,* 119–120.

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research, 41,* 3535–3545.

Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition, 17,* 1083–1108.

Sprague, N., Ballard, D. H., & Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Transactions on Applied Perception, 4,* 11.

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.

Stuphorn, V., & Schall, J. D. (2006). Executive control of countermanding saccades by the supplementary eye field. *Nature Neuroscience, 9,* 925–931.

Stuphorn, V., Taylor, T. L., & Schall, J. D. (2000). Performance monitoring by the supplementary eye field. *Nature, 408,* 857–860.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science, 304,* 1782–1787.

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

't Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., Konig, P., et al. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition, 17,* 1132–1158.

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. [PubMed] [Article]

Tatler, B. W. (Ed.) (2009). *Eye guidance and natural scenes*. Hove, UK: Psychology Press.

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

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research, 46,* 1857–1862.

Tatler, B. W., & Kuhn, G. (2007). Don't look now: The magic of misdirection. 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. 697–714). Oxford, UK: Elsevier.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2,* 1–18.

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition, 17,* 1029–1054.

Taylor, J. G., & Cutsuridis, V. (2011). Saliency, attention, active visual search, and picture scanning. *Cognitive Computation, 3,* 1–3.

Theeuwes, J., & Godijn, R. (2001). Attentional and oculomotor capture. In C. Folk & B. Gibson (Eds.), *Attraction, distraction, and action: Multiple perspectives on attentional capture* (pp. 121–150). Amsterdam, The Netherlands: Elsevier.

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Progress in Brian Research, 147,* 249–262.

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,* 766–786.

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

Trommershäuser, J., Glimcher, P. W., & Gegenfurtner, K. R. (2009). Visual processing, learning and feedback in the primate eye movement system. *Trends in Neurosciences, 32,* 583–590.

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.

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research, 43,* 333–346.

Uke-Karacan, H., & Hayhoe, M. (2008). Is attention drawn to changes in familiar scenes? *Visual Cognition, 16,* 346–374.

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,* 321–342.

Vincent, B. T., Baddeley, R. J., Correani, A., Troscianko, T., & Leonards, U. (2009). Do we look at lights? Using mixture modelling to distinguish between low- and high-level factors in natural image viewing. *Visual Cognition, 17,* 856–879.

Vincent, B. T., Troscianko, T., & Gilchrist, I. D. (2007). Investigating a space-variant weighted salience account of visual selection. *Vision Research, 47,* 1809–1820.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19,* 1395–1407.

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,* 326–343.

Wischnewski, M., Steil, J., Kehrer, L., & Schneider, W. (2009). Integrating inhomogeneous processing and proto-object formation in a computational model of visual attention. In H. Ritter, G. Sagerer, R. Dillmann, & M. Buss (Eds.), *Cognitive Systems Monographs* (vol. 6, pp. 93–102).

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

Wolfe, J. M. (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: Oxford.

Xu, J., Yang, Z., & Tsien, J. Z. (2010). Emergence of visual saliency from natural scenes via context-mediated probability distributions coding. *PLoS ONE, 5,* e15796.

Yantis, S. (1998). Control of visual attention. In H. Pashler (Ed.), *Attention* (pp. 233–256). Hove, UK: Psychology Press.

Yanulevskaya, V., Marsman, J. B., Cornelissen, F., & Geusebroek, J. (2010). An image statistics-based model for fixation prediction. *Cognitive Computation, 3,* 94–104.

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

Zehetleitner, M., Hegenloh, M., & Mueller, H. J. (2011). Visually guided pointing movements are driven by the salience map. *Journal of Vision, 11*(1):24, 1–18, http://www.journalofvision.org/content/11/1/24, doi:10.1167/11.1.24. [PubMed] [Article]

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115,* 787–835.

Zelinsky, G., Rao, R., Hayhoe, M., & Ballard, D. (1997). Eye movements reveal the spatio-temporal dynamics of visual search. *Psychological Science, 8,* 448–453.

Zhao, Q., & Koch, C. (2011). Learning a saliency map using fixated locations in natural scenes. *Journal of Vision, 11*(3):9, 1–15, http://www.journalofvision.org/content/11/3/9, doi:10.1167/11.3.9. [PubMed] [Article]

Zingale, C. M., & Kowler, E. (1987). Planning sequences of saccades. *Vision Research, 27,* 1327–1341.