

LATEST: A Model of Saccadic Decisions in Space and Time

Benjamin W. Tatler

School of Psychology

University of Aberdeen, UK

James R. Brockmole

Department of Psychology

University of Notre Dame, USA

R. H. S. Carpenter

Department of Physiology, Development and Neuroscience

University of Cambridge, UK

Author Note

No funding bodies directly supported this work.

All authors contributed to the development of ideas, theoretical underpinnings and interpretation of the findings, and to the preparation of the manuscript.

The ideas behind this work and some preliminary findings were presented by BWT as a talk at the 15th European Conference on Eye Movements in 2013, with the abstract of this talk published in the Journal of Eye Movement Research.

There are no conflicts of interest of any form for any of the authors.

The authors are particularly grateful to Reinhold Kliegl for statistical advice.

Author for correspondence: Benjamin W. Tatler, e-mail: [b.w.tatler@abdn.ac.uk](mailto:b.w.tatler@abdn.ac.uk), tel: +44 (0)1224 273930.

## Abstract

Many of our actions require visual information and for this it is important to direct the eyes to the right place at the right time. Two or three times every second, we must decide both when and where to direct our gaze. Understanding these decisions can reveal the moment-to-moment information priorities of the visual system and the strategies for information sampling employed by the brain to serve ongoing behaviour. Most theoretical frameworks and models of gaze control assume that the spatial and temporal aspects of fixation point selection depend on different mechanisms. We present a single model that can simultaneously account for both when and where we look. Underpinning this model is the theoretical assertion that each decision to move the eyes is an evaluation of the relative benefit expected from moving the eyes to a new location compared to that expected by continuing to fixate the current target. The eyes move when the evidence that favours moving to a new location outweighs that favouring staying at the present location. Our model provides not only an account of when the eyes move, but also what will be fixated. That is, an analysis of saccade timing alone enables us to predict where people look in a scene. Indeed our model accounts for fixation selection as well as (and often better than) current computational models of fixation selection in scene viewing.

KEYWORDS fixation duration, eye movements, decision making, gaze control, fixation selection, scene viewing

## LATEST: A Model of Saccadic Decisions in Space and Time

Much that we do requires visual information. However, the manner in which our eyes sample the environment greatly limits the information that is available to us: the small window of clear vision at the centre of gaze can only be directed to three locations or so in the environment each second. The valuable resource of high quality vision must therefore be allocated with care in order to provide the right information at the right time. Thus, understanding the mechanisms that underlie fixation allocation in space and time can tell us about the moment-to-moment information priorities of the visual system and the strategies for information sampling employed by the brain to serve ongoing behaviour. A complete understanding of eye movement behaviour must therefore encompass what determines both *when* and *where* we look. To date, attempts to decode the mechanisms underlying the temporal and spatial aspects of gaze control have remained separate in the literature. In this paper, we will argue this independence has led to a false dichotomy and instead offer the novel theoretical argument that a single decision process determines *when* our eyes move as well as *where* they move.

The logic behind our proposal is straightforward yet entirely novel. The purpose of saccades is to acquire information about the outside world, and the outside world is complex, containing many potential sources of information. Questions regarding when and where to move the eyes can therefore be thought of in terms of competition between different potential sources of information in the environment. We will argue that this competition is resolved via a Bayesian-like process, in which decision signals that represent the relative expected benefit associated with different actions race against each other until one of the decision signals reaches a threshold criterion for making a choice. Viewed in this way, the decision signal that reaches threshold first will determine not only the choice of action but also the time of its occurrence.

A central, and novel, feature of our model is that choices regarding when and where to move the eyes are made by evaluating the expected benefit of prolonging the current

fixation (i.e., making a choice to *Stay*) in relation to the expected benefit of saccading to another location in the periphery (i.e., making a choice to *Go*). We will argue that, in actuality, at any given moment, and in parallel, the visual system makes a great many Stay-or-Go evaluations - one for each possible saccade target. Which of these Stay-or-Go evaluations is quickest to reach a point where the evidence in favour of a decision to go sufficiently outweighs that in favour of a decision to stay will determine when and where the eyes move next. Under this proposal, then, it should be possible to use the factors that describe variations in fixation duration to also predict where the eyes are most likely to be directed. This, therefore, is our goal.

To meet our goal, we divide our discussion into several parts, each focusing on a critical phase of our argument. In Part I, we formally develop the proposal that gaze control can be viewed as a series of Stay-or-Go decisions and that choices regarding where we look and when we look there can be construed, both conceptually and mathematically, as stemming from a single decision process. We will then capture this process in a formal model of gaze control which, for reasons that will become obvious later, we call LATEST (Linear Approach to Threshold Explaining Space and Time). In Part II, we focus on the temporal dynamics of fixations by identifying various factors that modulate fixation duration and tuning LATEST to model the manner in which they do so. Then, in Part III we turn our attention to the determination of fixation points in space. Specifically, we will use our purely duration-tuned model to additionally predict saccade endpoints. Of course, the success of our theoretical argument depends upon the success of this endeavour, so to evaluate the efficacy of LATEST, we will compare it to several established models of spatial selection in scenes that have previously been constructed without reference to the temporal aspects of gaze. In Part IV we will focus on particular “special cases” of gaze control behaviours observed during scene viewing that we will introduce later. Previewing our final conclusion, in all situations we will show that in terms of predicting actual behaviour, no existing model of fixation selection performs as well as LATEST, providing

strong evidence that we should no longer adhere to the current zeitgeist in which temporal and spatial aspects of gaze control are treated independently.

## PART I: THEORETICAL FOUNDATIONS OF LATEST

### Saccades as decisions

Before thinking about saccades specifically, it is useful to consider a broader relationship between decision making and reaction time (after all, saccades are, at their core, reactions that can be timed). In general, there are two features of reaction times that are often puzzling and that need mechanistic explanation. One is that they are so long in relation to what would be expected of the underlying neural processes such as synaptic delay, nerve conduction time, and so on. The other is that in a series of identical trials using identical stimuli, a latency whose average is say 200 ms can vary randomly on different occasions over a range of some 50 - 100 ms. This variability is obvious when response time histograms are plotted (Appendix Figure A1a). Additionally, a positive skewness is evident in plots of any reaction time distribution, whether saccadic or manual, and whether evoked by visual or other kinds of stimuli. However, if we consider not the reaction time itself but its reciprocal then the skewness disappears and we find that the resultant distribution is not only symmetrical but Gaussian or normal (Appendix Figure A1b). By considering the reciprocal of reaction time we effectively consider the *rate* at which responses are made and, hence, the implied rate of the underlying decision process which appears to be Gaussian. This can best be demonstrated by plotting cumulative histograms rather than conventional frequency histograms using a probit scale for probability. The result (called a reciprobit plot) will then be a straight line if the decision rate is Gaussian (Appendix Figure A1c-d).

Because reciprocal reaction time (or latency) reflects the rate of the underlying decision process, a first step in modelling decision processes is to consider whether we can model this distribution. The simplest model that will explain this normal distribution of

reciprocal latency is LATER: Linear Approach to Threshold with Ergodic Rate (Carpenter, 1981; Noorani & Carpenter, 2016). The basic principles of this approach are illustrated in Figure 1. A decision signal  $S$  rises linearly at a rate  $r$  from its initial value of  $S_0$  until it reaches a threshold value  $S_T$ , at which point the response is initiated. The variability in reaction times that is frequently observed comes about because the rate,  $r$ , by which information is accumulated for evaluating the decision varies from trial to trial with a normal distribution. This normal distribution of rates of accumulation can be described as having a mean rate of accumulation of  $\mu$  and variance of  $\sigma^2$ .

As described, this is of course merely an *ad hoc* empirical description; what strengthens the LATER model is that - apart from the built-in randomness - it also happens to represent the simplest possible implementation of an elementary quasi-Bayesian decision mechanism. More formally and generally, we can express its operation in terms of evidence  $E$ , supplied by sensory information, being accumulated to support competing hypotheses.

To illustrate, let's consider a situation in which an observer must determine if one or the other of a specific object (i.e. target) is present in a visual display. When looking at a kitchen, for example, an observer may need to determine if the object on a counter is a mug of tea or a bowl of soup. Hypothesis  $H_1$  would then be that a particular target (e.g., a mug of tea) is present and hypothesis  $H_2$  that another target (e.g., a bowl of soup) is present. More specifically, the likelihood ratio  $L$  for the two hypotheses, given by  $p(E | H_1)/p(E | H_2)$ , is used to update the prior odds,  $Q(H_1, H_2) = p(H_1)/p(H_2)$  to give the posterior odds  $Q'$ :

$$Q' = Q.L \tag{1}$$

To simplify various mathematical aspects of calculating likelihood ratios, Equation 1 can also be expressed in logarithmic form:



$$\log Q' = \log Q + \log L \quad (2)$$

where  $\log L$  is known as the *support* given to  $H_1$  (that a mug of tea is present) relative to  $H_2$  (that a bowl of soup is present) provided by E (Edwards, 1992). This process continues iteratively over time with the prior odds at time  $T$  being derived from the posterior odds at time  $T - 1$ . If the rate at which information is gathered remains constant throughout the decision process, the logarithm of the posterior likelihood will rise linearly. A decision to accept  $H_1$  can be made when the rising log odds reaches a criterion level - at this point the evidence in support of  $H_1$  sufficiently outweighs the evidence in support of  $H_2$ , and the observer determines that a mug of tea is on the counter. In other words, if we are prepared to identify LATER's decision signal  $S$  with log odds, we can see that the model provides a simple explanation for reaction times that is at the same time a kind of elementary decision mechanism (Carpenter, 2012; Carpenter & Williams, 1995; Noorani & Carpenter, 2011; Noorani & Carpenter, 2016).

It is the linear rise of the decision signal that distinguishes LATER from other models of accumulating evidence in neural decision processes (Ratcliff, 2001). This is because it is proposed that noise from integrating sensory information is unlikely to contribute significantly to decision times except when that information is very hard to gather, for example for visual information at very low contrast (Carpenter, Reddi, & Anderson, 2009). Conversely, under high-contrast conditions, when targets are easily detected, the random variability of reaction time to suddenly-presented targets is due to a process occurring at the decision stage rather than as a by-product of integration of sensory noise (Schall & Thompson, 1999).

### Applying LATER to saccadic decisions

While providing a framework for decision processes in general, LATER has been applied successfully to understand saccadic decision processes in particular, at least under

conditions where observers (sometimes monkeys) generate a single saccade within discrete trials to a pre-defined target within a simple visual display. Predictions based on the assumptions in LATER are borne out experimentally: Diminished supply of visual information (Reddi, Asrress, & Carpenter, 2003), reduced expectations (Carpenter & Williams, 1995), and lessened pressure to respond quickly (Reddi & Carpenter, 2000) all increase saccade latencies because they, respectively, slow the rate of rise of the decision signal ( $r$ ), diminish starting activation ( $S_0$ ), and increase decision thresholds ( $S_T$ ) within the model. Reward, or more generally the expected benefit of making a saccade to a particular location, for example in terms of how much information its fixation is likely to yield, would also be expected to contribute to the LATER decision process, and a number of experiments have demonstrated that this is in fact observed (Schütz, Trommershäuser, & Gegenfurtner, 2012; Takikawa, Kawagoe, Itoh, Nakahara, & Hikosaka, 2002; Watanabe, Lauwereyns, & Hikosaka, 2003).

LATER appears to correspond to aspects of what is known about the neural architecture of saccade generation. Recordings from the parietal cortex of monkeys carrying out saccadic decisions have confirmed that the neurons do indeed code for log odds (Yang & Shadlen, 2007) as would be necessary for a decision process of the form proposed in LATER. In monkey frontal eye fields, activity in movement neurons rises steadily prior to saccade initiation with a rate that varies randomly for each saccade, with neuronal activation for each saccade reaching a relatively consistent final level (Hanes & Schall, 1996). Comparison of the time-course of activity in stimulus-related neurons in the same area shows that this random variation is gratuitously ‘injected’ somewhere between the stimulus- and movement-related neurons, rather than being a consequence of sensory noise (Thompson, Bichot, & Schall, 1997, 1996). Similar linearly rising signals can be seen at several other locations during the latent period for the response, notably in the superior colliculus, where the starting level appears to reflect prior expectation (Basso & Wurtz, 1998; Krauzlis, Dill, & Kornyló, 2002, 2004), as LATER would predict, and in parietal

cortex (Huk & Shadlen, 2005; Lynch, Mountcastle, Talbot, & Yin, 1977; Shadlen & Newsome, 2001); these rising signals are almost certainly the cause of the well-known *Bereitschaftspotentialen* or readiness potentials, that can be recorded in humans by averaging EEG in synchronization with a subsequent voluntary response (Becker, Hoehne, Iwase, & Kornhuber, 1972; Kornhuber & Deecke, 1965). For more discussion of the neural correlates of LATER see Noorani and Carpenter (2016).

Despite the evidence reviewed above, there is one feature of saccadic responses seen in some situations - and, importantly for the development of our new model, in scene viewing - that cannot be explained by a single accumulation process in LATER (Figure 1). With large data sets, particularly under conditions of high expectation or urgency, one often sees a small sub-population of very short-latency responses that cannot be explained by the simple LATER model. These responses are infrequent - often less than 10% of the total responses - but more frequent than would be expected for a single underlying decision process. These fast responses form a distinct LATERian population of their own: when plotted on a reciprobital scale, these early responses lie on a different straight line that is much shallower than the main distribution and usually extrapolates back to a zero intercept at infinite time, implying a large value of  $\sigma$  and a  $\mu$  of zero (Figure 2). They can be explained by a relatively simple model in which two LATER units operate in parallel, the first of them to reach threshold determining the time of the response. One unit is the sensible decision unit  $(\mu, \sigma)$  that generates the main part of the distribution; the other is a kind of maverick unit, which because of the large  $\sigma$  and despite the zero  $\mu$ , just occasionally wins, generating an early response. The early saccades they generate are a prominent feature of the latency distributions observed during reading, nystagmus and scene viewing (Carpenter, 1994; Carpenter & McDonald, 2007; McDonald, Carpenter, & Shillcock, 2005; Roos, Calandrini, & Carpenter, 2008). Therefore, any model of saccade timing in scene viewing needs to account for the over-representation of very short duration fixations, and in our decision-based model we achieve this by proposing a basic model architecture

comprising two competing decision signals - the sensible (discussed in Parts II and III) and maverick (discussed in Part IV) decision processes outlined here and illustrated in Figure 2.

### **From LATER to LATEST: modelling spontaneous saccades in scene viewing**

The experiments that led to successful modelling of the variability of saccadic response times, described above, were all based on evoked responses in discrete trials using simple visual displays. Because of the constrained circumstances under which LATER was developed, existing LATER accounts of saccadic decisions are simply inadequate when one wants to consider real-world scene viewing. In the real world we are faced with a very different and less artificial situation than that which LATER was designed to deal with: the potential targets are there all the time, and there are a lot of them. Furthermore, LATER describes information processing and accumulation at peripheral retinal locations, with no involvement of information processing in central vision. However, information in central vision is known to modulate fixation duration in reading (Rayner, 1998) and scene viewing (Henderson, Weeks Jr, & Hollingworth, 1999; Underwood, Humphreys, & Cross, 2007). Hence, any framework for understanding saccade timing during scene viewing should therefore incorporate not only information processing in peripheral vision but also that within central vision. This requirement underpins a major conceptual deviation of our proposed LATEST model from previous LATER models of saccadic decisions.

We propose a novel form of the decision process that incorporates peripheral and central visual information processing by considering the decision as an evaluation of the relative merits of saccading to a candidate target location in the periphery compared with maintaining fixation at the target currently being fixated. We can then re-frame the comparison in Equation 2, which pitches competing hypotheses about the information gathered at a peripheral location against each other and has underpinned all previous LATER models of saccadic decisions, in a new way: as an evaluation of the hypothesis that behavioural goals will best be served by moving the eyes to the candidate location (GO to

the new location) relative to the hypothesis that behavioural goals will best be served by maintaining fixation (STAY at the current location):

$$\log Q_T(\mathbf{H}_{\text{GO}}, \mathbf{H}_{\text{STAY}}) = \log Q_{T-1}(\mathbf{H}_{\text{GO}}, \mathbf{H}_{\text{STAY}}) + \text{support}(\mathbf{H}_{\text{GO}}, \mathbf{H}_{\text{STAY}}) \quad (3)$$

This Stay-or-Go form of the underlying decision process offers a key conceptual departure from previous decision models, and allows the decision process to be conceptualised in a way that is more appropriate for complex scenes with many potential targets for each saccade and the evaluation of evidence at current and potential fixation locations.

Within this novel Stay-or-Go theoretical framework, when evaluating a decision to move the eyes from the current fixation point (1) to a possible future location (2), the mean rate of rise of the decision signal,  $\mu$ , can be decomposed into independent factors, representing the tendency to stay at the current location  $\mu_{\text{STAY}}$  and the tendency to go to the new target  $\mu_{\text{GO}}$  (Figure 3).  $\mu_{\text{GO}}$  can in turn be decomposed into a component which is a function of the target location, in particular the visual information ( $\mu_2$ ) at the location, and another determined by the retinocentric position of the target, i.e. relative to the current fixation ( $\mu_{12}$ ).  $\mu_{\text{STAY}}$  on the other hand can be expected to be a function only of the content of the current fixation ( $\mu_1$ ). In Part II of the present paper we characterise the factors that contribute to  $\mu_{\text{STAY}}$  and  $\mu_{\text{GO}}$

Of course, from any fixation during scene viewing there are many potential locations for the next saccade to target. If we assume that the visual system is able to gather information in parallel across the extent of the visual field, then this phrasing of the decision has the advantage of allowing multiple Stay-or-Go decision signals to rise toward threshold simultaneously. Thus, for each potential peripheral target location a separate Stay-or-Go decision can be evaluated, with the location at which support for Go reaches threshold first being the one to which the saccade is directed (Figure 3). Not only does this conceptualisation of spontaneous saccade generation in scene viewing allow both

information at fixation and in the periphery to contribute to the decision processes, but it also makes a strong prediction that decisions about where and when to move the eyes are intimately linked.

A race between multiple peripheral Stay-or-Go decisions essentially describes a form of spatial decision map underlying saccade generation in scene viewing. However, unlike previous such propositions (e.g., Borji & Itti, 2013; Itti, Koch, & Niebur, 1998), in our proposal the underlying spatial decision map arises from the temporal dynamics of the decision processes for when to move the eyes, essentially comprising the rising log odds of the Stay-or-Go evaluation at each location. If we can model saccade timings in scene viewing we should therefore be able to use this model to predict spatial selection as well, thus providing a single underlying mechanism for deciding when and where to look based on LATER-like ballistic accumulation of evidence (a proposition we will test empirically in Part III). LATER or LATER-like models have never previously been applied to predict the endpoint of saccades; the model presented in this paper is the first to do so. We call this model LATEST (Linear Approach to Threshold Explaining Space and Time).

In LATEST, the inclusion of information in central vision contributing to saccadic decision times necessarily means that the resulting decision map is fundamentally retinotopic in nature. Thus, each fixation will be underpinned by a different decision map, which evolves over time from the simultaneous computation of Stay-or-Go evaluations throughout peripheral vision. Because deciding whether to saccade to each peripheral location involves the relative evaluation of evidence from both the current and candidate location, the manner in which the log odds rise toward threshold at any particular peripheral location may vary considerably from fixation to fixation. For example, the decision signal associated with potentially saccading to Location 2 in Figure 3 will rise differently when the participant is fixating Location 1 than it would if the participant was fixating other locations (say Location 3). In this way it is not solely the information at Location 2 that determines whether and how quickly it is selected, but the interplay

between the information at that location and in central vision. This proposed antagonism between fovea and periphery is a prominent feature of many of the underlying neural structures concerned with the initiation of saccades. In the colliculus, there are mechanisms of mutual inhibition between neurons in the region corresponding to the fovea, and thus the current point of fixation, and those coding for peripheral targets (Munoz & Istvan, 1998; Munoz & Wurtz, 1995), and a similar antagonistic arrangement has been reported in monkey frontal eye fields (Hanes & Wurtz, 2001), though there may not be the sharp distinction between a ‘fixation zone’ and the rest of the colliculus that was originally believed (Goffart, Hafed, & Krauzlis, 2012; Nummela & Krauzlis, 2011).

In the present paper we evaluate both the appropriateness of our proposed Stay-or-Go theoretical foundation of LATEST and the ability of a single model to explain both when and where observers look. We do this by first producing a descriptive model of the factors that contribute to modulation in saccade timing in scene viewing (Part II) and then using the outcome of this modelling of when observers move the eyes to predict where people look (Part III).

## **PART II: MODELLING *WHEN* WE MOVE OUR EYES**

Despite considerable evidence regarding factors that appear to modulate fixation duration, our theoretical understanding of the control of fixation duration in complex scene viewing is relatively underdeveloped, with only a small number of models having been proposed.

### **Contrasts Between LATEST and previous models of fixation duration**

At least four key factors define LATEST and differentiate it from existing models of fixation duration. First, LATEST proposes that the time taken to move the eyes reflects the time taken to gather sufficient evidence to support a decision to move the fovea to a peripheral location. In this way, the decisions about where and when to move the eyes are intricately and inevitably linked and arise from the same underlying processes. No other

model of fixation duration makes this claim. Many models of fixation placement remain silent on the determinants of fixation duration and vice versa (e.g., Borji & Itti, 2013; Nuthmann, Smith, Engbert, & Henderson, 2010). Those few models that do address both the spatial and temporal aspects of gaze control propose a mechanistic division between the two. For example, Findlay and Walker (1999) proposed that when and where the eyes move arises from competition between a ‘fixate centre’ mechanism that governs how long a fixation should be maintained and a ‘move centre’ mechanism that identifies where the eyes should move to next. Even models that propose that temporal and spatial factors interact to influence target selection rely on two separate processes for time and space that compete with each other rather than a single underlying process that encompasses both when and where we look. For example, to explain correlations between fixation duration and saccade amplitude in scene viewing, Unema, Pannasch, Joos, and Velichkovsky (2005) proposed that selection arises from the combination of a race to threshold to trigger a saccade to a salient object and an inhibitory process to decrease local activation in the salience map shortly after the start of the fixation.

Second, in LATEST, decisions about when and where to move the eyes arise from Stay-or-Go evaluations of the relative benefit offered by moving the eyes to a new location or by staying at the current foveal target. This allows both central and peripheral information to contribute to the decision and thus to influence decisions in space and time. Most models of fixation duration are not so inclusive. For example, Mackay, Cerf, and Koch (2012), who in other respects have provided an account of fixation duration that is closest to the one we propose in this article, only considered visual information at the saccade target location with no contribution of central information to saccade timing. (As we will see, LATEST also includes a broader spectrum of visual features whereas Mackay et al., only considered salience as described by Walther & Koch, 2006, and faces.)

Third, fixation times in LATEST are assumed to reflect decision time alone with no additional processes (such as saccade programming) contributing to their duration. This



omission of saccadic programming time as a central factor in determining saccade timing in LATEST stands in contrast to other models of saccade duration. For example, in their CRISP (Controlled Random-walk with Inhibition for Saccadic Planning) model, Nuthmann et al. (2010) propose that saccades are generated by a random walk to threshold which is followed by the initiation of a saccade program. The saccade program comprises first a labile stage during which the saccade can be cancelled, and then a non-labile stage during which cancellation is not possible (see also Becker & Juergens, 1979). The random walk for the next saccade begins as soon as the labile stage of the saccade program is initiated. Visual and cognitive factors exert their influence on fixation duration via alterations in the random walk process as well as cancellations of saccade programs during their labile phase. The lack of saccadic programming time in LATEST stems from the fact that very few neurons appear to be involved between the neural trigger signal to move the eyes and the oculomotor muscles (Büttner, Hepp, & Henn, 1977), thus neural programming seems unlikely to contribute significantly to the time it takes to decide to move the eyes. Indeed, a wide variety of neurophysiological studies have demonstrated very short latencies between visual stimulation and the appearance of responses in various visual areas, and between electrical stimulation in the oculomotor system and saccadic responses (Robinson, 1972; Sparks, 1986; Sylvestre & Cullen, 1999).

Finally, in LATEST, the sub-population of very short duration fixations commonly observed in scene viewing and other studies of saccade latencies and fixation durations arises from competition between the main decision process and a maverick saccade generator. The former is under the influence of factors that govern information gathering and evaluation, whereas the latter varies greatly in rate of rise but does so around a mean rate of rise of zero, meaning that occasionally it will rise to threshold faster than the main decision unit, generating unexpectedly short fixation durations. This proposal again stands in contrast to existing models. For example, in CRISP (Nuthmann et al., 2010) short duration fixations occur because a saccade programme initiated during the previous

fixation has already reached its non-labile phase by the time the current fixation begins and thus cannot be cancelled, resulting in a saccade being executed soon after the start of the current fixation.

### **Identifying factors that modulate fixation duration in LATEST**

What information is used to evaluate the Stay-or-Go decision? Equation 3 suggests that decisions reflect the evidence in support of moving to a peripheral location ( $\mu_{GO}$ , Figure 3) relative to the evidence in support of maintaining fixation at the current foveal target ( $\mu_{FO}$ , Figure 3). Thus information processing at these two locations should contribute to decision times. In scenes, visual information can be described at a variety of levels, from basic low-level features to higher level semantic understanding. Prior research suggests that fixation durations may be modulated by information across this range of levels.

Substantial work has explored which specific visual features in an environment influence fixation duration. Focusing on scene viewing, degrading visual content by lowering luminance (Loftus, 1985) or eliminating high-spatial frequencies (Mannan, Ruddock, & Wooding, 1995) result in prolonged fixations. However, not all degradations of low-level information are associated with increased fixation duration. For example, local image contrast appears to be independent of fixation duration (Einhäuser & König, 2003; Guo, Mahmoodi, Robertson, & Young, 2006). Hence, the degree to which fixation durations are systematically linked to individual low-level image properties remains equivocal. Alternatively, rather than defining scene information according to variation in individual features, another approach is to describe the properties of scenes in terms of feature constellations. This approach is most commonly used to model fixation locations (Borji & Itti, 2013; Itti et al., 1998; Judd, Durand, & Torralba, 2012), but the general principles can be used to characterize fixation durations. For example, scenes with higher clutter (a metric based on the covariance of several features within a scene, see Rosenholtz,

1999, 2001) are associated with longer search times (Henderson, Chanceaux, & Smith, 2009; Rosenholtz, Li, & Nakano, 2007). Because such approaches combine features they can be thought of as intermediate levels of description.

While the relationship between basic visual features in scenes and fixation duration is uncertain and somewhat inconsistent, semantic factors have well-documented influences on fixation duration. Objects that are difficult to recognise (De Graef, Christiaens, and d'Ydewalle (1990) or that are semantically inconsistent with the scene (Becker, Pashler, & Lubin, 2007; Bonitz & Gordon, 2008; Castelhana & Heaven, 2011; De Graef et al., 1990; Henderson et al., 1999; Rayner, Castelhana, & Yang, 2009; Underwood, Templeman, Lamming, & Foulsham, 2008) all receive longer fixations. These results imply that the semantic informativeness of an object may underlie fixation duration. Behavioural goals also modulate fixation duration, with shorter fixation durations during search than during scene memorisation (Henderson et al., 1999; Nuthmann et al., 2010).

It is clear that visual information defined at basic, intermediate or semantic levels in scenes is likely to impact saccadic decision times and as such present important candidates for components of any model of saccade timing in scene viewing. Moreover, it is clear that these sources of information are not constrained to the current target of fixation, but distributed over the current and to-be-fixated location. For example, we know that basic features can be extracted extrafoveally (Treisman & Gelade, 1980) and such extrafoveal processing may be sufficient to allow object recognition (Li, VanRullen, Koch, & Perona, 2002; Spotorno, Malcolm, & Tatler, 2015; Underwood et al., 2008). This implies that extrafoveal targets can receive quite high levels of processing.

In addition to the low- intermediate- and higher-level information at both at fovea and in peripheral vision, other, less strictly visual factors are also known to influence saccade timings and should be included in our list of potential factors that might modulate decision times. In Figure 3, we proposed that factors related to the spatial relationship between the current fixation and the selected target location should also be considered as

possible contributors to saccadic decision times (expressed as  $\mu_{12}$ ). Based on previous evidence, we need to consider the eccentricity of the targeted information as greater eccentricity of targets will necessarily limit information supply and thus influence the accumulation of evidence at the peripheral location. Furthermore, the amplitudes of incoming and outgoing saccades can influence how long a fixation lasts, although the nature and direction of this relationship is not consistently observed (Cohen & Ross, 1977; Pelz & Canosa, 2001; Tatler & Vincent, 2008; Unema et al., 2005). Not only the amplitude of the outgoing saccade matter, but also its direction in relation to the incoming saccade. When two consecutive saccades are in the same direction, the duration of the fixation that separates them tends to be shorter than when the two saccades are not in the same direction (Anderson, Yadav, & Carpenter, 2008; Carpenter, 2001; Tatler & Vincent, 2008). One additional non-visual factor contributes to fixation durations is ordinal fixation number. In some of the earliest recordings of fixation behaviour when viewing complex scenes and patterns Buswell (1935) observed that fixation durations tend to increase over the course of viewing an image for several seconds, an effect that has been replicated frequently in subsequent years

## Experimental Method

**Participants.** Seventy volunteers (mean age 23.3, 20 male) from the University of Dundee participated in the experiment for course credit or monetary compensation. All had normal or corrected-to-normal vision and were naive to the purposes of the experiment. The study was approved by the University of Dundee Research Ethics Committee (SREC 10026, “Eye movements while viewing scenes”).

**Stimuli and Apparatus.** Sixty-four images of everyday indoor and outdoor scenes were used as stimuli (Appendix Figure A2), selected from image sets previously used in Tatler (2007) and Brockmole and Henderson (2006). Images were displayed in 800 x 600 pixel format on a ViewSonic G90f-4 19-inch CRT monitor running at a refresh rate of 100

Hz with the display resolution set to 1024 x 768 pixels. At a constrained viewing distance of 63.5 cm, the screen viewing area subtended approximately 31.8 x 23.8 degrees of visual angle. Images subtended approximately 24.8 x 18.6 degrees. Gaze position was recorded at a rate of 1000 Hz with an EyeLink 1000 (SR Research, Canada) eye tracker. Viewing was binocular but data were recorded only from the dominant eye. A chin and forehead rest stabilised the head.

**Procedure.** Participants were informed that they would see a series of photographs of everyday indoor and outdoor scenes, and that they should remember as much as they could about the scenes and their contents for a memory test that would follow the final image. The memory test was not administered but rather served to provide a common task across participants (see Tatler, Hayhoe, Land, & Ballard, 2011, for a discussion). The experimental trials were preceded by a 9-point calibration procedure that was used to ensure that gaze estimation was better than 0.5 degrees on average over the 9 calibration points and no worse than 1 degree on any of the individual calibration points. If necessary, equipment set-up was modified and calibration was repeated until these accuracy criteria were satisfied. Each trial then began with a central fixation point. By interrogating the spatial correspondence between this point and estimated gaze position, calibration could be monitored throughout the experiment and adjusted when necessary. The scene was then displayed for 10 seconds during which participants were free to move their eyes.

**Data Quality and Preparation.** To parse the gaze record into fixations and saccades, we used the SR Research algorithm with default sensitivity settings<sup>1</sup>. Across participants, this algorithm identified 127,197 fixations that began and ended while the scene was visible on the display. However, we excluded 448 fixations (.35%) that were located off the scene image and 10,007 fixations (8.2%) that were considered unreliable<sup>2</sup> from our analyses.

**General modelling approach.** Our modelling approach followed two distinct steps. The first step was to conduct an essential test of our basic model architecture by

verifying that the distribution of fixation durations for each individual was consistent with two LATER-like linear accumulation processes working in competition: one a sensible unit and the other a maverick unit (see Figure 2). However, distribution level modelling is not the goal of this work - rather we are interested in whether we can model the factors that modulate each observed fixation duration. Thus, the second step in our modelling approach was to model the variation in fixation duration on a fixation by fixation basis. We now explain each of these two modelling approaches in detail.

**Step 1: Distribution modelling procedure.** Data were modelled using the software application SPIC (Carpenter, 1994) to obtain estimates of three parameters describing the underlying distribution(s) of fixation durations: the mean and standard deviation of the rate of the main distribution of fixation durations and the standard deviation of the rate for any identified sub-population of unusually short fixation durations (these constitute the three parameters in LATER, described earlier). Durations were modelled in 10-ms bins and Kolmogorov-Smirnov (Kolmogorov, 1941) tests were performed to compare the observed distributions of fixation durations to those predicted by a LATER model generating saccades using the parameter estimates. Best-fit estimates of three distribution parameters were obtained by minimising the Kolmogorov-Smirnov statistic. Because fixation durations vary considerably between individuals (Andrews & Coppola, 1999) it is important to calculate separate fits for each participant.

**Step 2: Modelling variation in fixation duration.** The distribution modelling procedure allows us to identify any fixations likely to have been terminated by a saccade generated by the maverick saccade generator in our basic two-unit model of saccade generation (Figure 2). If we assume that the sub-population of very short duration fixations arises from a stochastic process, not influenced by visual or cognitive factors, it is reasonable to suggest that any attempts to model the factors that modulate fixation duration should not attempt to encompass these essentially randomly terminated fixations. Rather, modelling efforts should focus on understanding the factors that influence decision

times in the main population of saccades. We therefore constrained our modelling efforts to those fixation durations likely to have arisen from the main decision process and likely to have involved an accumulation of evidence for evaluating the decision to move the eyes or remain fixated at the current location (we will provide a complete analysis of the sub-population of short fixations in Part IV).

We used the three parameter estimates derived from the distribution modelling procedure to simulate saccadic decision times from two competing LATER decision units - the main and maverick units - and from this generate a simulated set of fixation durations. For any simulated fixation duration, we then calculated the probability that this duration would have been derived from the main distribution (see Appendix Figure A3); this was calculated for each participant due to the high variability in LATER parameter estimates (see Appendix Table A1). Our calculated likelihoods were used to assign a probability that each observed fixation duration in our dataset was generated by the main decision unit. For analysis, fixations were included if the estimated probability that they were drawn from the main distribution of durations was greater than 60%.

Our final dataset for modelling comprised fixations that were likely to have been terminated by a saccade generated by the main decision process, that passed our quality assurance criteria and that were followed by a fixation that also passes our quality assurance criteria. These strict requirements resulted in a final dataset of 55,341 fixations for modelling.

Because fixation durations are not normally distributed, we modelled their reciprocal. As described earlier, the reciprocal of fixation duration conveniently describes the rate of rise of the decision signal,  $\mu$ , and thus can be used to consider the influence of the fixed effects in our model upon the rate of rise of the decision process. We refer to this as the decision rate or  $\mu$  in the results section that follows;  $\mu$  can be assumed to represent the linear sum of all the component factors that contribute to it.

We used Linear Mixed Models to predict fixation durations, using R Studio

(RStudio Team, 2015) to run the lme4 package<sup>3</sup> (Bates, Maechler, Bolker, & Walker, 2014) in the R statistical analysis environment (R Core Team, 2014). This modelling approach allows between-subject and between-item variance to be estimated simultaneously, is able to handle mixtures of discrete and continuous predictors, and is suitable for handling large numbers of covariates. These advantages are desirable for the present work because previous literature identifies a number of different factors that appear to be associated with modulations in fixation duration and as such it is important to be able to evaluate a range of possible factors in the model, while simultaneously accounting for correlations between these factors. Based on previous studies (reviewed above) we included measures of low-, intermediate- and high-level visual information in scenes as fixed effects.

For low-level information we computed orientation information in the images which has been shown to account for spatial fixation selection better than other sources of low-level information in scenes (Baddeley & Tatler, 2006). Orientation information was computed using four oriented odd-phase Gabor patches (oriented at 0 deg, 45 deg, 90 deg, and 135 deg), taking the absolute values in order to capture unsigned difference from the mean. These four maps were combined and normalised (see Baddeley & Tatler, 2006; Tatler, Baddeley, & Gilchrist, 2005; Tatler & Vincent, 2009, for a discussion). Orientation maps were constructed for high-frequency information using Gabors with envelope standard deviation of 20 cycles per degree.

For intermediate-level information in scenes we used a model of image salience. Many such models exist (Borji & Itti, 2013; Judd et al., 2012). We described image salience using RARE2012<sup>4</sup> (Riche et al., 2013). This salience model describes feature rarity at multiple scales within colour and orientation feature channels. By computing rarity at multiple scales, the model highlights locations that differ from local context but are also globally rare in the image. The overall salience map combines rarity maps across features and scales.

Describing higher-level information in scenes is not trivial. When attempting to model high-level, semantic effects in scene viewing, we can consider global changes



associated with different task settings (as in Nuthmann et al., 2010), or we can try to capture semantic effects within tasks. One could assume that describing where objects lie in scenes would be an appropriate description of semantic factors in images. Indeed object-level descriptions of scenes can offer good accounts of where people look in scenes (Einhäuser, Spain, & Perona, 2008; Nuthmann & Henderson, 2010). However, it is not always clear how objects should be defined, and object definitions are typically and necessarily subjective (Einhäuser et al., 2008; Nuthmann & Henderson, 2010). Furthermore, semantically informative locations in scenes may be objects or may be background locations (e.g., Charness, Reingold, Pomplun, & Stampe, 2001).

An alternative approach - and the one that we used here - is to collect subjective ratings from participants, asking them to indicate the most semantically informative locations in scenes. In a rating study, 27 participants from the University of Dundee and the University of Notre Dame (none of whom participated in our main eye movement study) were shown each of the 64 images used in the main study and asked to select the five most semantically interesting locations in the image. Participants were asked to select locations based on how meaningful they were for the scene and were explicitly instructed to ignore aspects such as the colours or brightnesses in the images. Images were displayed using custom-coded GUIs in Matlab and selections were made by clicking on five locations in each image. Participants were able to re-select any of their five selections before moving on to the next image, but were not permitted to revise selections once they had moved on to a new image. Images were displayed in random order. We used the selections to create “semantic interest” maps by centring Gaussians with full width at half maximum of two degrees around each selected location.

Figure 4 shows examples of orientation, salience and semantic rating maps for five sample images used in the present study. Information for subsequent modelling was extracted from these maps by computing the mean value in each map within a 1 degree x 1 degree patch centred around fixation. Because we were interested in the impact of

information at fixation and at potential peripheral target locations, we extracted information not only for the current fixation but also for the next fixation location (i.e. the target of the observed saccade).

Given previous evidence regarding factors that appear to influence fixation duration we included a further set of fixed effects in our model of fixation duration. These were the ordinal fixation number in each trial, the change in direction between the incoming and outgoing saccades, and the amplitudes of the incoming and outgoing saccades. Because retinal sampling limits will necessarily restrict information availability in peripheral vision,, we also included fixed effects for the interaction between the amplitude of the outgoing saccade (which describes the eccentricity of the target about to be fixated) and each of our three measures of visual information in scenes.

The effects of each fixed effect in the final LMM are presented graphically as partial effects using the `remef()` function provided by Hohenstein and Kliegl (see Hohenstein & Kliegl, 2014, for example use) for `lmer()` outputs. This approach allows us to plot the relationship between predictor and outcome once any effects of other factors in the model have been removed, providing a better reflection of the actual underlying relationship between each fixed effect and the outcome.

## Results

**Distribution modelling.** Figure 5 shows reciprobbit plots of distributions of fixation durations for six of our 70 participants (for the distributions of all 70 see Appendix Figure A4). The prominence of the population of early saccades is clear, suggesting that there are two underlying distributions of fixation durations: one comprising a relatively few fixations (typically well under 10% for each participant) of short duration and the other comprising longer ones and contributing the majority of observations for each participant. These findings support our proposed basic architecture for LATEST of two competing LATER-like decision units: the main unit generating sensible decisions and the maverick

unit that competes with it.

For all participants, fixation durations were well-fitted by our proposed LATEST basic model architecture of a main decision unit together with an early component with zero intercept (Kolmogorov-Smirnov,  $p > 0.1$ , one-sample test; see Figure 6 for two examples of simulated data based on parameter fits). Thus fits had three free parameters: the mean  $\mu$  and standard deviation  $\sigma$  of the rate of rise for the main population and the standard deviation  $\sigma_\epsilon$  for the early population (which had a mean rate of rise of 0). The parameter estimates (Appendix Table A1) and reciprobbit plots (Appendix Figure A4) reveal the large variation between participants.

**Modelling fixation duration in the main unit.** Table 1 shows the output for our LMM describing saccade timing in scene viewing for those fixations likely to have been terminated by a saccade generated by the main Stay-or-Go decision unit in our two-process competitive model of saccade generation.

Decision rate (and thus fixation duration) varied with ordinal fixation number, the angular change in direction between the incoming and outgoing saccades, and the amplitudes of the incoming and outgoing saccades (Figure 7). The effects of incoming and outgoing saccade amplitude were qualified by interactions with change in saccade direction. These findings are largely in line with previous studies that have considered the relationship between these variables and fixation duration, which have found that fixation durations increase with viewing time and thus ordinal fixation number (Buswell, 1935; Unema et al., 2005), increase with increasing angular change between surrounding saccades (Anderson et al., 2008; Carpenter, 2001; Smith & Henderson, 2009; Tatler & Vincent, 2008), and have also demonstrated a non-linear relationship between fixation duration and the amplitude of the outgoing saccade (Unema et al., 2005; Velichkovsky, Rothert, Kopf, Dornhöfer, & Joos, 2002). Unlike previous results, however, we also found a strong relationship between incoming saccade amplitude and decision rate, with larger saccades being followed by longer duration fixations (contrary to Tatler & Vincent, 2008, who found no relationship).

Decision rate varied significantly with edge content at current and target locations, with increasing edge content at fixation being associated with decreased decision rate, but increasing edge content at the target location being associated with increased decision rate (Figure 8). For visual salience, there was a significant relationship between salience at fixation and decision rate (Figure 8), with higher salience at fixation reducing decision rate. Decision rate was strongly predicted by semantic interest at the current and next fixation, with decision rate decreasing as semantic interest at fixation increased and decision rate increasing as semantic interest at the saccade target location increased (Figure 8); the effect of semantic interest at the target location was qualified by an interaction with outgoing saccade amplitude (thus target eccentricity) suggesting that increasing retinal eccentricity reduced the effect of semantic interest on decision rate.

## Discussion

The distribution modelling results suggest that the durations of fixations observed while participants viewed images of real-world scenes in our study can be well described as arising from a process of linear accumulation of evidence for a decision process, operating in competition with a maverick signal that gives rise to a small number of very fast responses when this maverick signal reaches threshold activation faster than the rising sensible decision signal (we will investigate the maverick saccades in detail in Part IV). Thus, our observed data confirm the proposed basic architecture of LATEST: Two competing LATER-like decision units, one that conducts a sensible evaluation of evidence, the other that produces maverick responses.

An important prediction of LATEST is that because of the inverse relation between Stay and Go, a factor that increases  $\mu$  at the fixation point will be expected to reduce it when present at the next location. As can be seen in Figure 8, in the case of both edge information and semantic interest, this prediction appears to be fulfilled: at fixation,  $\mu$  falls both with increasing edge information and semantics, whereas the opposite is true in

respect of the next location.

If we return to the schematic of the decision process presented in Figure 3, we can use the results of the LMM to suggest the likely contributors to the Stay and Go components of the evaluation (namely  $\mu_1$ ,  $\mu_2$  and  $\mu_{12}$ ). Importantly, the LMM approach allows variance to be attributed appropriately, ensuring that reported contributions of each fixed effect are those that can be attributed uniquely to that fixed effect after the contributions of all other fixed effects in the model have been accounted for.

The finding that ordinal fixation number influenced decision rate further implies that we need to include an additional component for factors unrelated to the current or candidate (next) location (Figure 9), which we will call  $\mu_0$ . We might also attribute the influence of the incoming saccade amplitude to this category of effects as it is not strictly related to information processing in central or peripheral vision or to planning the next saccade. Given our theoretical assertion that the decision is an evaluation of two competing hypotheses (in this case Stay versus Go), it is helpful to attribute any factors not directly attributable to the current or next location to one of the two hypotheses being evaluated. Because these factors are by definition not related to the content of either the current or the next location they could be assigned to either of our two hypotheses. However, we prefer to attribute them to the Stay component of the evaluation for two reasons. First, they are common across all decision processes that we propose are being evaluated simultaneously when selecting from multiple candidate locations for the next saccade and thus share this common contribution that factors at the current location do to each decision process (see Figure 3). Second, the results of our LMM show that as ordinal fixation number increases so decision rate decreases; thus this factor contributes by promoting elongated stays at the currently fixated location as viewing progresses.

**Factors influencing  $\mu_{\text{STAY}}$ .** If we start by decomposing  $\mu_{\text{STAY}}$  into factors related to visual processing at the current foveal location  $\mu_1$  and factors not strictly related to current visual processing  $\mu_0$  we can use the findings of our LMM to suggest the likely

contributors to these components:

$$\mu_0 = aN + bA + cA^2 \quad (4)$$

where  $N$  is the ordinal fixation number in viewing,  $A$  is the amplitude of the incoming saccade and  $A^2$  is the squared amplitude of the incoming saccade;  $a$ ,  $b$  and  $c$  represent the weights of these factors, derived from the model output shown in Table 1.

$$\mu_1 = dI_1 + eS_1 + fO_1 \quad (5)$$

where  $I$  is the semantic interest rating,  $S$  is the salience and  $O$  is the orientation or edge content at the current foveal target location (1); each weighted ( $d$ - $f$ ) according to the LMM output in Table 1.

Thus, our findings show that low-, intermediate- and higher-level information at fixation all contribute to decision time by promoting perseverance at the current fixation location, with decision rate decreasing when the current location was rich in edge information, high in visual salience or was rated as semantically meaningful. In prior work, influences of low-level visual features on fixation duration have only been found when visual information is degraded, with longer durations when global luminance is reduced (Loftus, 1985) or scenes are low-pass filtered (Mannan et al., 1995). These previous findings for degraded scenes may reflect decreased discriminability of information in scenes, leading to increased processing time, and as such are rather unrelated to the contribution of low-level visual information described in the present study. However, our finding that low-level edge information in scenes contributes to saccade timings is at odds with previous work that has found no relationship between low-level properties - namely contrast, which is highly correlated with edge information - and fixation durations in scene viewing (Einhäuser & König, 2003; Guo et al., 2006). The reason for this difference between our findings and those of Einhäuser and König (2003) and Guo et al. (2006) may arise from the

inclusion of other potentially correlated sources of information in our model, which may reveal relationships not found when these correlations are not accounted for (see Baddeley & Tatler, 2006, for a similar argument about the need to account for correlations between features to reveal better understanding of the influences of any feature). Alternatively, if the contribution of different sources of information reflects an evaluation of the expected benefit of Staying or Going, then the contribution of any particular source of information is likely to be quite sensitive to task demands (we will return to this point in Part V).

Semantic effects on fixation duration have been found in a variety of prior studies. Objects that are semantically inconsistent with the scene in which they are placed are fixated for longer (Bonitz & Gordon, 2008; Castelhana & Heaven, 2011; De Graef et al., 1990; Underwood et al., 2008). Our effect of prolonged fixation at semantically informative locations is therefore consistent with previous studies.

**Factors influencing  $\mu_{GO}$ .** Breaking down  $\mu_{GO}$  into factors related to visual processing at the target location  $\mu_2$  and factors related to getting the gaze to the target location  $\mu_{12}$  we can suggest the following contributing factors:

$$\mu_2 = gI_2 + hO_2 \quad (6)$$

where  $I$  is the semantic interest rating and  $O$  is the orientation or edge content at the target location (2); each weighted ( $g-h$ ) according to the LMM output in Table 1.

$$\mu_{12} = iC + jR_2 + kR_2^2 + lR_2.I_2 + mC.R_2 + nC.A \quad (7)$$

where  $C$  is the change in saccadic direction required to saccade to the next location, and  $R$  and  $R^2$  represent the linear and quadratic effects of the retinocentric position of the target location (i.e. outgoing saccade amplitude). Interactions between semantic interest at the target location and its retinal eccentricity,  $R_2.I_2$ , between change in saccade direction and outgoing saccade amplitude,  $C.R_2$ , and between change in saccade direction and

incoming saccade amplitude,  $C.A$ , can all be attributed to  $\mu_{12}$ . Weights  $i-n$  are derived from Table 1.

Our finding that higher-level semantic interest at a peripheral target location can influence saccade timings in scene viewing suggests that processing of potential target locations is sufficient to extract semantic information in peripheral vision. Similar proposals of semantic processing prior to foveal inspection have been suggested for scene viewing (Li et al., 2002; Underwood et al., 2008) and reading (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998).

Decision rate decreased (non-linearly) with increasing outgoing saccadic amplitude. Because outgoing saccade amplitude is itself a measure of target eccentricity, this effect is as would be predicted given retinal sampling limits, which inevitably result in visual information being less readily available as retinal eccentricity increases, thus leading to longer saccade latencies (Reddi et al., 2003). The interaction between outgoing saccade amplitude and semantic interest at the target location suggests that semantic processing was affected by target eccentricity in a way that was above and beyond the influence of eccentricity on other visual information sources.

The decision process outlined in this section and characterised in the LMM describes a single decision about whether to move the gaze to a particular location in the scene or maintain the current fixation. However, we are modelling the outcome of a decision process rather than the decision process itself. That is, we already know the chosen target location and so can look for features of that target location that might have contributed to the decision process. In complex scenes, the decision to move the eyes is not a simple evaluation of a single selected target location in peripheral vision. Rather, every location in the scene is a potential saccade target and must be evaluated. Thus, in our conceptualisation of the problem, the observed saccade reflects the winner of multiple Stay-or-Go evaluations carried out in parallel across candidate locations in the scene, each rising to threshold as evidence is accumulated over fixation time. In this way, the saccade



that is ultimately triggered reflects the winner of a race between many competing scene locations. This winner is the location for which the evidence that behavioural goals would best be served by moving to that location rather than maintaining fixation was fastest to reach the criterion threshold for triggering a saccade.

If this framing of the decision process is correct, then we should be able to predict the likely time taken to decide to saccade to any location in the scene given a known starting location. Assuming that the ultimate destination is that location at which the support for moving the eyes compared to maintaining fixation reaches a threshold criterion the fastest, then we should be able to use our understanding of the dynamics of saccade timing decisions to predict where we direct our saccades in the scene.

As promised earlier, we now turn from considerations of saccadic timing to considerations of where the saccades are directed; more specifically, whether our model of saccade timing is also able to provide a description of spatial selection in scene viewing. To date no single model has offered a unified description of spatial and temporal selection in scene viewing.

### **PART III: MODELLING *WHERE* WE MOVE OUR EYES**

No variant of a LATER-like model of latency or fixation duration has ever also been used to predict the endpoint of a saccade. Our goal in this paper was not only develop a novel LATER-like model of fixation duration during scene viewing, but to use that same model as a means to predict fixation locations. In Part III, we evaluate LATEST's ability to do exactly this. Before turning to this issue directly, some general comment on the modelling of fixation location, including a review of specific existing models, is needed. We begin this discussion by noting that a prevalent concept in trying to explain where saccades are directed in complex scenes is that the scene is represented in the brain as a spatial map, and that the choice of target is the result of competition between elements of that map. This in turn poses two major questions: 'What is being mapped?' and 'What is the

nature of the competitive process?’

### **What is being mapped?**

An obvious suggestion is that it is a map of conspicuity or salience, based on low-level image features, as pioneered by Itti, Koch and colleagues (Itti & Koch, 2000; Itti et al., 1998; Koch & Ullman, 1985). A wide range of models based on conspicuity have been proposed since Itti’s model, which vary in the manner in which features are defined and combined to produce the salience map (see Borji & Itti, 2013; Judd et al., 2012). But it is clear that maps also need to incorporate high-level contributions as well, for example in the form of task-based weighting of feature channels (Navalpakkam & Itti, 2005), expected object appearance (Kanan, Tong, Zhang, & Cottrell, 2009) and expected object location (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Torralba, Oliva, Castelhana, & Henderson, 2006). Models that include both image salience and such higher-level information tend to make better predictions for where saccades are directed when viewing scenes. While most contemporary models are based around conspicuity maps, some suggest that the underlying map might reflect proto-objects (Wischniewski, Belardinelli, Schneider, & Steil, 2010) or objects (Nuthmann & Henderson, 2010) rather than low-level conspicuity. In all cases, saccade targets are selected from spatial maps created from incoming visual information, which can be thought of collectively as priority maps for attention allocation. Neurophysiological evidence supports the notion of such priority maps that combine low- and high-level information (e.g., Fecteau & Munoz, 2006). Evidence for such maps is abundant and can be found for the superior colliculus (McPeck & 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).

Our proposed decision map bears some similarities with those described above for spatial selection: we propose a spatial map of activity that reflects priority from which decisions about saccades are made using a winner-takes-all process. However, we can

qualify the spatial frame of reference for our map as being fundamentally retinocentric: each decision process involves the calculation of the relative merit of moving compared to staying at the currently fixated location. Retinotopy is an increasingly common component of priority map models of fixation selection, given the need to account for peripheral acuity sampling limitations (e.g., Vincent, Troscianko, & Gilchrist, 2007; Wischniewski et al., 2010).

### **What is the nature of the competitive process?**

In previous work, it is generally assumed that selection involves some kind of winner-takes-all race between candidate areas of the map, but the implication that spatial selection of a target ought to be closely related to the time taken to decide to initiate the saccade has tended to be neglected. The key difference between our proposed decision map and the priority maps commonly proposed in the literature is that the map evolves dynamically over time within a fixation as evidence is gathered and evaluated. This temporal evolution provides a map that ought to be able to account not only for where we move our gaze to but also for when we move it. The manner in which the map evolves over time should reflect factors related to the local to the availability of evidence at each peripheral location but also related to the information in central vision.

The results of Part II show that decision times in scene viewing are influenced by visual information in central and peripheral vision and by less visually-based factors such as ordinal fixation number and saccade amplitudes and direction. If our Stay-or-Go theoretical foundation is appropriate for describing these decisions then not only should the factors identified in Part II also influence where observers fixate, but also the relative importance of different factors in decision timing should reveal the relative importance of those factors in spatial selection. In Part III we test this assertion by using the weights of the factors evaluated in the LMM presented in Part II to predict where observers looked when viewing the scenes.

## Method

**Computing LATEST spatiotemporal decision maps.** We can use the weights in the LMM to predict decision rates for any location in a scene. Given that decision rates derive from the balance between factors favouring moving to a new location (Go) and those favouring staying at the current location (Stay), decision rate estimation requires both a starting location and a target location. Thus for the images we used in our study, for a particular starting pixel location we can calculate the expected decision rate for deciding to move gaze to any other pixel in the image. If saccades result from a race between multiple Stay-or-Go evaluations being carried out in parallel throughout peripheral vision, the pixel with the highest predicted decision rate should be the one that would be looked at next given the selected starting position in the scene (with the obvious caveat that there is precision error from both the eye tracker and the saccadic targeting system).

In order to evaluate whether these decision maps calculated from the relative weights of factors influencing decision time do indeed predict where observers fixate, for each observed fixation in the dataset we calculated the expected decision rate for every pixel in the scene. We can then use these decision maps, derived from timing data, to compare these predictions with where our human observers actually looked next. If our framing of saccadic decisions in space and time as arising from the same underlying decision processes is appropriate, we should find good correspondence between these predictions and where subjects actually look; if not - if when and where we move our gaze are under separate, independent control - the correspondence is likely to be low.

For our calculation of predicted decision rate,  $\mu$ , for each fixation we can break down this calculation into the four components  $\mu_{0-3}$  as described in equations 4-7.  $\mu_0$  was estimated based upon the ordinal fixation number and the incoming (i.e. previous) saccade amplitude.  $\mu_1$  was estimated by looking up the observed fixation location in maps of edge content, salience (RARE2012) and semantic ratings, and taking the mean values within 0.5 degrees of fixation as described in the method for the previous section. We then calculated

$\mu_2$  and  $\mu_{12}$  for each pixel in the scene in turn.  $\mu_2$  was estimated in the same way as  $\mu_1$  but for the candidate target pixel rather than starting pixel.  $\mu_{12}$  was estimated by calculating the size and change in direction of the saccade that would be needed to bring gaze to the pixel under evaluation (i.e. the distance between it and the observed point of fixation, and the change in direction between the incoming and candidate outgoing saccade that would be required to bring the eye on to it). We also included calculations for the interactions between change in saccadic direction and incoming saccade amplitude, change in saccadic direction and the candidate outgoing saccade amplitude, and between the candidate outgoing saccade amplitude and the semantic interest at the candidate location. Examples of predicted decision rate maps for five successive observed human fixations can be seen in Figure 10.

**Evaluating the spatial selection in LATEST.** If spatial selection proceeds on a winner-takes-all basis from our predicted decision rate maps, the pixel in the image associated with the predicted maximum decision rate should correspond with the target of the next observed human saccade. However, given the inherent randomness in saccadic decision processes (Carpenter, 1981), together with instrument error from the eye tracker and precision error from the saccadic system, such predictions must necessarily be subject to a degree of stochastic variability. We therefore evaluated the spatial predictiveness of the model by creating binary maps indicating the n% pixels with highest predicted decision rate (Figure 10), and then considered the whether each observed saccade targeted a location that fell within the regions predicted in the binary map (as in Judd et al., 2012).

In order to quantify the ability of the model to account for spatial selection, we compared the proportion of observed human fixations falling within predicted regions of the scene to that expected by chance. This chance baseline for selecting predicted regions was calculated using control locations drawn from a distribution that encompasses image-independent viewing biases in observers, such as the tendency to look at the centre of a scene irrespective of the content of the displayed scene (e.g., Tatler, 2007). This

approach offers a more conservative, and arguably more suitable (Clarke & Tatler, 2014), evaluation of any spatial selection model because it allows an evaluation of the degree to which a model accounts for fixation selection above and beyond that which can be attributed to image-independent biases.

In order to measure model performance we evaluated the ability of the threshold maps to account for observed human behaviour over a range of thresholds. This allowed us to calculate the area under the Receiver Operating Characteristic (ROC) curve for the spatial predictions of our LATEST model by comparing hit rate to false alarm rate over varying thresholds (e.g., Judd et al., 2012). The area under the ROC curve (AUC) provides a powerful and commonly used measure of model performance (Borji & Itti, 2013; Tatler et al., 2005).

**Comparing LATEST to existing models of spatial selection.** We selected eight salience models that cover a range of different methods for describing visual information, are available as Matlab toolboxes and do not rely on parochial parameter settings for particular tasks. Specifically we evaluated spatial selection performance of (1) AIM, the information-maximization method proposed by Bruce and Tsotsos (2007); (2) the Adaptive Whitening Saliency Model (AWS) proposed by Garcia-Diaz, Leborán, Fdez-Vidal, and Pardo (2012); (3) the feature congestion measure of visual clutter from Rosenholtz et al. (2007); (4) Graph-based visual saliency (GBVS: Harel, Koch, & Perona, 2006); (5) Image Signature (Hou, Harel, & Koch, 2012); (6) RARE2012 (Riche et al., 2013); the salience toolbox (Walther & Koch, 2006), which is similar, but not identical, to the algorithm proposed by Itti et al. (1998); and (8) SUN: the Bayesian model for saliency using natural statistics developed by Zhang, Tong, Marks, Shan, and Cottrell (2008). Figure 11 shows the eight resultant salience maps for five images used in the present study.

To evaluate the performance of each model in accounting for human fixation behaviour we used the same thresholding procedure and comparison to control locations described for evaluating the predictions derived from our LATEST model. We were then

able to compare the ROC curves for the predictions from LATEST to those for each of the eight computational salience models of spatial selection in scene viewing.

## Results

Figure 12 provides an example of the proportion of fixations accounted for by LATEST, when the threshold of the predicted latency map includes the 20% of pixels with the fastest predicted decision rate. Performances of each of the eight tested salience models are also shown when the threshold includes the 20% of pixels with the highest salience. For comparison, the proportion of control samples falling within the above-threshold portions of the maps for each model are also shown. At this 20% threshold not only does LATEST account for a higher proportion of fixations than any other model (at 48.0%) but also has the second lowest proportion of false alarms (at 21.9%). Notably this false alarm rate is considerably lower for LATEST than the four salience models that predict the highest proportion of human fixations (RARE2012, GBVS, IS and AWS), and indeed the only model that has a lower false alarm rate, STB, also has the lowest hit rate.

Of course the example in Figure 12 is merely illustrative. A better test of any model is to vary the threshold of the binary prediction map and see how this affects the hit rate and false alarm rate for each model. The area under the ROC curve (AUC) provides a single summary metric that expresses the relationship between false alarm rate and hit rate. Table 2 shows that LATEST performed on a par with the best of the existing, tested models of spatial selection (for our dataset, the best overall account of spatial selection by a pre-existing salience model was for AWS). However, to better understand the performance of our model with respect to existing models, it is useful to plot the effect of varying the threshold on the ROC curve. Figure 13 (which plots the curves used to calculate the AUC values in Table 2) shows that the performance of LATEST relative to the other models and indeed relative to chance varies considerably over varying false alarms (thus varying thresholds in the binary predictions maps). Specifically, at low false

alarm rates, LATEST performs better than any of the evaluated existing models by a considerable margin. That is, when the models are used to predict only a small proportion of pixels as likely targets of the next saccade - thus evaluating the correspondence between fixation selection and the pixels that the models predict as being the most likely candidates for selection - LATEST performs considerably better than other models.

However, at high false alarm rates, when the threshold includes high proportions of pixels as predicted target locations, LATEST performs worse than other models and indeed for very high false alarm rates (for thresholds approaching 100% in the binary prediction maps) LATEST performs worse than our biased random baseline samples. It is important to consider the implications and source of the poor performance of LATEST at high false alarm rates. Essentially, these results likely arise from a small subset of fixations that the model is particularly bad at predicting. By applying a high threshold we can reveal the locations that the model predicts are the least likely to be targeted by the outgoing saccade (Figure 14, top).

The prediction maps shown in Figure 14 (top) suggest that the poor performance of the model at high thresholds might arise from its inability to predict a subset of outgoing saccades that reverse direction and have an amplitude of up to or slightly beyond the incoming saccade. To test whether the unusual shape of the ROC curve plotted in Figure 13 derives from the model's particular difficulty with this subset of saccades, we considered how the model performed when they are removed. We evaluated two approaches for removing the problematic saccades. First we used a simple heuristic of removing outgoing saccades that were within 20 angular degrees of a complete reversal in direction to that of the incoming saccade and that had an amplitude less than 110% of that of the incoming saccade. This simple approach removed 8.69% of the dataset and resulted in the improved model performance seen in Figure 14 (bottom left), corresponding to an AUC of 0.6873. Second, we used a 95% threshold (like the prediction maps depicted in Figure 14) to identify outgoing saccades that LATEST failed to predict at this threshold and then



excluded these saccades from all evaluations of the model (at all thresholds). This second approach removed 7.5% of the dataset and resulted in the ROC curves shown in Figure 14 (bottom right), corresponding to an AUC of 0.7354. For fairness, the eight salience models were evaluated again for each of these two methods, excluding the same saccades from their evaluations. Thus the ROC curves plotted for the salience models in Figure 14 are for model performances for datasets with the same saccades excluded as were excluded for the LATEST model evaluations. These plots support our suggestion that it is this small subset of saccades that poses particular problems for the spatial predictions of LATEST and without their inclusion the model performs very well at predicting the remaining saccades that our observers made when viewing the images in our study.

## Discussion

When compared to a range of existing saliency models, developed to explain spatial selection in scene viewing, our proposed model performed well. This is particularly remarkable given that the weights for information that contributed to the decision maps in our model were derived from a linear mixed model describing factors that influence fixation duration in scene viewing. The fact that a model describing fixation duration offers a powerful description of spatial selection has two key implications for our understanding of saccade generation in scene viewing. First, it provides a challenge to existing frameworks that have proposed separate mechanisms for selection in space and time in scene viewing (e.g., Findlay & Walker, 1999; Nuthmann et al., 2010). We propose that a single underlying decision mechanism can explain both where we look and when we move our gaze, providing a more parsimonious account of saccadic selections in space and time than previous models. Second, it confirms that our theoretical proposal that saccade timing decisions reflect Stay-or-Go evaluations, with multiple racing decision processes across scenes, is a suitable way to conceptualise saccadic decisions in space and time.

We did find, however, a small subset of fixations that LATEST struggled to predict.

These saccades may highlight a limitation of the manner in which we have implemented our model, such as an inappropriate account of effects of changes in direction (which we modelled as a linear effect, interactive with incoming and outgoing saccade amplitude). However, allowing higher order (quadratic and cubic) effects of change in saccadic direction did not improve the fit of the model. Alternatively, our findings might suggest that decisions to move the eyes to these locations are governed slightly differently from how we have implemented our model. One possibility is that these locations that fall between the current and previous fixation have an increased prior or starting activation level. An increase in the starting activation level at these locations would mean that Stay-or-Go decisions reached the threshold likelihood ratio for deciding to move the eyes quicker than expected for a predicted decision rate, this increasing the probability that these locations would reach threshold before others and be targeted by the next saccade. An increase in the prior at these locations might arise from spillover of processing from the previous fixation.

#### **PART IV: EARLY SACCADES AND THE MAVERICK UNIT**

So far in this paper we have focussed our modelling efforts on the main population of saccades, excluding saccades that terminate the sub-population of very short duration fixations. Saccades after very short fixations are often referred to as early saccades because they are capable of terminating the fixation earlier than would have happened based on accumulating evidence for a decision by the main decision unit.

We have focussed on the main unit saccades because in LATEST, the assumption is that it is only the main unit that uses visual information for evaluating decisions about when to move gaze. Very short duration fixations arise because of a maverick decision unit that races with the main unit, but is not based upon detailed visual processing; rather these decision times arise from a signal that rises to threshold with a rate drawn randomly from a distribution with a mean of zero but high variance; but the fact that they are

activated at all may well depend on the existence of some kind of crude visual input at that location complicating their interpretation. Nearly all our knowledge of these early saccades has come from studying responses to single targets in evoked tasks, which cannot provide enlightenment on this point. Here in Part IV we offer a first exploration of this set of early saccades in complex scene viewing.

If LATEST provides an appropriate means to characterise all saccades, we can use it to make and test predictions about the very fast population of saccades. First and most importantly, decision rates for these saccades should not be under the influence of visual information processing at fixation or at the intended target location. If this is the case then we should also expect that where these saccades land should also differ from saccades generated by the main decision unit.

In order to characterise temporal and spatial selection for saccades following very short duration fixations we used the same procedure outlined for the main unit and illustrated in Appendix Figure A3. To ensure that the selected fixations were very unlikely to have been generated by the main decision unit, we set a criterion probability of 80% that the fixation duration was terminated by a saccade from the maverick LATER unit. This process allowed us to identify 16,016 fixations for subsequent analyses.

We ran a linear mixed effects model of the same form as that used to analyse our main population of fixations, with the same fixed and random effects being used to predict decision rate. As before we first ran a model using all fixed effects and then reduced the model in step-wise fashion until we had removed as many fixed effects as possible without significantly changing the overall fit of the model. Table 3 shows the outcome of this modelling process for both the full and reduced version of the model.

It is clear from Table 3 that fixed effects relating to visual information in the scenes (edges, salience and semantics) both at fixation and target locations had no significant influence on decision rate. This result is entirely consistent with our assertion that these short fixations arise from processes that are unrelated to visual information processing at

the fovea or peripheral target location. Decision rate was, however, influenced by fixed effects related to the amplitudes and directions of saccades around the very short duration fixations and to ordinal fixation number (Figure 15). For ordinal fixation number, incoming saccade amplitude and outgoing saccade amplitude, the effects on decision rate are essentially opposite to those for the main unit (see Figure 7). This was not the case for the effect of change in direction between the incoming and outgoing saccades, where the same direction of effect can be seen for the early saccades as for the main population of saccades.

In order to consider where early unit saccades target, we first compared spatial distributions of these saccade endpoints to the endpoints of saccades generated by the main unit. Logically, it might be expected that if very short duration fixations arise because they are terminated by randomly generated saccades from the maverick decision unit, the distribution of end points of these saccades should itself be random and thus less clustered than the end points of saccades generated by the main decision unit. This was not what we observed (Figure 16) and indeed the entropy of the probability distributions for early saccades was less than that for main unit saccades: creating probability distributions using Gaussians with full width at half maximum of two degrees, entropy was significantly lower for the early saccades than for the main saccades,  $z = 4.79$ ,  $p < .001$ , suggesting more clustering (or at least peakier distributions) for saccades generated by the early decision unit than by the main decision unit.

Not only were distributions of saccades generated by the early unit more clustered, but the locations targeted were on average of higher semantic informativeness than main unit saccades. A GLMM to predict whether a location was targeted by an early unit saccade or main unit saccade showed effects of semantic informativeness,  $t = 3.5$ ,  $p < .001$ , salience,  $t = -2.57$ ,  $p = .010$ , and edges,  $t = -2.16$ ,  $p = .030$ , with early saccades targeting regions of higher semantic informativeness but lower salience and edge content than main unit saccades. This result is contradictory to our expectations and suggests that there is more to the population of very short duration fixations than merely random saccade

generation.

One possible explanation of the observed spatial clustering in the endpoints of saccades following very short fixations is that some of them are correction saccades, bringing the fovea to bear upon the target initially intended, but missed, by the preceding saccade. Correction saccades are often observed in laboratory-based paradigms that require the eye to be directed to small peripheral targets. In such paradigms the eyes typically land short of the intended target, and a small corrective saccade ensues to bring the fovea to bear upon the target (Abrams, Meyer, & Kornblum, 1989; Becker, 1972, 1991). These corrective saccades tend to be small, with initial saccades falling about 10% short of the distance required to move to the peripheral target, and to be made in the same direction as the initial saccade. If the observed early saccades in our dataset are corrective saccades then we should be able to identify these from their amplitudes (they should be very small) and the change in direction between incoming and outgoing saccades (they should continue in the same direction).

The distributions of how much the direction of a saccade deviates from the saccade immediately preceding it differed for early and main saccades,  $z = 25.30$ ,  $p < .001$ , with early saccades more frequently continuing in the direction of the incoming saccade than was the case for the main population of saccades (Figure 17a). Small ( $< 2$  degrees) and large ( $> 4$  degrees) amplitude early saccades differed in their distributions of change in direction,  $z = 20.55$ ,  $p < .001$ , with small amplitude outgoing early saccades being more likely to continue in the same direction ( $+/- 22.5$  degrees) as the incoming saccade (Figure 17b). Furthermore, early saccades that did continue in the same direction as the previous saccade were more frequently of smaller amplitude than those that did not continue in the same direction,  $z = 20.03$ ,  $p < .001$  (Figure 17c).

Collectively, these observations are consistent with the possibility that there is a subset of saccades within the population that we are referring to as ‘early saccades’ that conform to the characteristics that would be expected for corrective saccades. This subset

is likely to be quite small: 3858 early saccades had amplitudes of less than 2 degrees, of which 1295 were under 1 degree in amplitude; expressed as a percentage of the incoming saccades, 1299 saccades were of an amplitude less than 20% that of the incoming saccade, or which 330 were of an amplitude of less than 10% of the incoming saccade; in terms of the change in direction between incoming and outgoing saccades, 3854 outgoing early unit saccades continued in approximately the same direction as the incoming saccade ( $\pm 22.5$  degrees). Combining criteria of amplitude ( $< 2$  degrees) and direction (within 22.5 degrees of the incoming saccade) identified 1219 of the 16,016 early saccades.

Corrective saccades should be under different control to other early saccades. Specifically, if corrective saccades function to bring the eye to the intended target of the incoming saccade, whereas other early saccades are more randomly directed, these two sets of early saccades should target locations that are quantifiably different in content. One way to test this is to see whether the target locations for corrective and non-corrective early saccades differ in terms of their edge, salience and semantic content. A GLMM to predict type of early saccade (corrective vs. not-corrective) showed effects of semantics,  $t = 9.55$ ,  $p < .001$ , edge information,  $t = 4.31$ ,  $p < .001$ , and salience,  $t = -1.98$ ,  $p = .048$ , at the target location. These results show that corrective saccades targeted locations of higher semantic and edge information but lower salience than did other early saccades. Not only were corrective saccades differently distributed with respect to visual information in the scene, but also their timing was influenced by different factors. We ran LMMs to predict decision rate for saccades decision rate for corrective and non-corrective early saccades using the same fixed effects as in previous models in this paper (Table 4). It should be noted that the selection criteria for differentiating corrective and non-corrective saccades limit the range of values in fixed effects describing outgoing saccade amplitude and change in direction. We found that the timing of corrective saccades was influenced only by an interaction between semantic information at the target location and the eccentricity of that target location. For peripheral targets of low semantic interest, there was little influence of

retinal eccentricity on saccade timing; however, for peripheral targets of high semantic interest, saccades were initiated faster as eccentricity increased from 0.5 to 2 degrees (Figure 18). In contrast, for non-corrective early saccades, timing was influenced by fixation number, change in direction and incoming saccade amplitude, but not any factors related to visual information in the scene.

Taken together, and using our rather crude way of identifying candidate corrective saccades, we can suggest that small amplitude saccades that continue in the same direction as the incoming saccade (our candidate corrective saccades) differed from other early saccades in terms of the factors that influence their timing and the visual information that they selected in the scene. Corrective saccade timing depended on semantic information at the intended peripheral location and these saccades targeted locations with higher semantic and edge content than other early saccades.

After identifying the likely presence of corrective saccades within the population of early saccades, we can reconsider the question of whether spatial selection in this population is as would be expected for a maverick decision process. Above, we reported that early saccades targeted regions of higher semantic interest, but lower salience and edge content. If we remove the candidate corrective saccades, the remaining early saccades do not target locations of higher semantic informativeness than main unit saccades: indeed, a GLMM showed that non-corrective early saccades and main unit saccades targeted locations that did not differ in semantic informativeness,  $t = 0.40$ ,  $p = .690$ , or salience,  $t = -1.87$ ,  $p = .062$ ; however, non-corrective early saccades did target locations of lower edge content than main unit saccades,  $t = -3.20$ ,  $p = .001$ .

The destinations of saccades are not the only potential diagnostic marker of whether these early saccades are generated from a maverick decision signal. In evoked tasks, whether or not the maverick unit is triggered at all depends upon the information processing demands at the fovea: with more frequent early saccades in gap paradigms where there is no central processing demand at target onset compared to step tasks where

processing demand continues until target appearance (Story & Carpenter, 2009). It is therefore likely to be the case that factors that contribute to the Stay component of our Stay-or-Go evaluation should reduce the frequency of early saccades. If this is the case then we should find that early saccades in our dataset were launched from locations with strong Stay signals than main unit saccades. This is what we found: non-corrective saccades were launched from locations of lower semantic informativeness,  $t = 17.14$ ,  $p < .001$ , salience,  $t = 4.83$ ,  $p < .001$ , and edge content,  $t = 4.93$ ,  $p < .001$ , than main unit saccades.

Our data suggest that once we have removed the likely corrective saccades from our population of saccades following very short fixation durations, the remaining early saccades are consistent in their timing and spatial selection with what would be expected for a maverick decision process.

## **PART V: GENERAL DISCUSSION**

We have shown that fixation durations when viewing images of natural scenes can be well described as reflecting the time taken to accumulate sufficient evidence regarding visual information in the scene in order to support a decision to move the eyes. More precisely, the decision to move the eyes appears to be the result of a comparison between competing Stay and Go hypotheses: each unit evaluates the evidence in favour of moving to a peripheral location relative to staying at the currently fixated location. This framing is supported by our finding that visual information both at foveation and at the intended peripheral location contribute to decision time, but in opposite directions: increased visual information at fixation prolongs fixation time whereas at the peripheral target it reduces it.

We used LATER as our underlying decision model because it offers a simple but powerful decision mechanism that describes decisions as arising from the likelihood ratio of two competing hypotheses. As is common in LATER modelling, in addition to this ‘rational’ decision process that entails scrutiny and processing of visual information in the scene, we included another that produces few but unusually short fixation durations (early



saccades: a prominent feature both of reading and nystagmus) that is likely to arise (mainly) from a maverick decision process that is relatively unaffected by visual stimulation.

LATER was developed as a two-parameter model of the distribution of reaction times in a simple RT task and relates this to what would be expected of an ideal Bayesian process. Subsequently, it was shown that such units, acting in parallel and corresponding to different peripheral targets, can explain behaviour in choice tasks (for a review see Noorani & Carpenter, 2016), and other more complex ones such as antisaccade tasks and, in at least one preliminary investigation, fixation durations in spontaneous visual scanning (Roos et al., 2008). While this work gives credence to the idea that LATER-like decisions are appropriate for understanding saccade timings during scene viewing, it is important to note that LATEST and LATER differ considerably. LATER is strictly a model of (saccade) latency which has never been used to predict saccade end points, employs a decision unit that only evaluates information at peripheral locations, and has never been used to characterise the contribution of different forms of information in scenes (or for that matter the oculomotor variables we explore) on decision rate. In stark contrast, LATEST is able to explain both when and where people look in scenes, frames these decisions as evaluations between hypotheses derived from information in central vision (Stay) and information at a peripheral location (Go), and considers a wide range of visual, oculomotor, and cognitive factors that are known to influence gaze control during scene viewing.

### **LATEST: a single model for saccadic decisions in space and time**

Conceptualising the underlying decisions in our model as arising from Stay-or-Go evaluations makes the strong prediction that such a decision framework should not only describe *when* the eyes move but also *where*. More specifically, each location in peripheral space would carry out its own Stay-or-Go evaluation: the location in space associated with the fastest log odds for this evaluation should be the location targeted by the next saccade.

Thus factors that modulate saccade timing should modulate spatial selection, and the model we have proposed to account for saccade timing should also account for where people look. Because the decision processes that underpin LATEST are intrinsically stochastic, with ‘deliberate’ built-in randomness of timing, the predictions of the model must necessarily be of average behaviour rather than being exact for every single saccade.

Our findings support the theoretical assertions that underlie LATEST: a spatial selection model based on the outcome of our decision rate model offers a good account of where people look during viewing. Saccades target locations that fall within regions associated with the fastest predicted Stay-or-Go decision processes. Overall, LATEST performed at a level equivalent to the best of the existing spatial selection models that we evaluated, which were drawn from among the best performing contemporary models of fixation selection. When we set the threshold on our binary prediction maps to predict only the most likely locations to be fixated, LATEST far outperformed all tested spatial selection models (outperforming these models for thresholds up to 35% of the most likely locations in the image). That the overall performance of LATEST was on a par with the best of the evaluated spatial selection models rather than beyond them was because LATEST performed less well at accounting for a small subset of saccades that brought the eye to locations between the current and previous location that were better predicted by other models. When this small subset of fixations was removed, LATEST far outperformed all other models we evaluated.

LATEST, therefore, is the first account of fixation selection that explains where we look and when we move our eyes as arising from the same underlying process. All previous models have treated the decisions about when to move the eyes and where to look as separate processes arising from different underlying mechanisms.

### **Early saccades: mavericks or more?**

While the main population of fixation durations had decision rates that were modulated by visual factors at the current and intended target locations, this was not the case for the small population of ‘early’ saccades mentioned earlier, that can be attributed to a maverick decision unit that races against the main decision process. This maverick unit draws randomly from a distribution of rates of rise with large standard deviation and a mean of zero. As such, the maverick process rarely wins the race, but when it does so, it triggers a saccade after an unusually short period of time. If the short duration fixations observed in our scene viewing data arose from such a maverick decision process, then their timing might be expected to be unrelated to visual information at fixation or at the targeted peripheral location. However, although their timing is extremely random, the fact that they are evoked at all in respect of a particular location is undoubtedly stimulus-related, at least in the case of evoked saccades. Examination of the spatial distribution of locations targeted by these early saccades suggested that some of these saccades were more likely to be corrective, bringing the eyes to locations of higher semantic interest, rather than random. Separating out early saccades that look like corrections from the rest showed that the timing and spatial selection of our putative corrections correlated with visual information at the targeted location, whereas timing and spatial selection for the remaining early saccades did not. Thus we suggest that our observed population of early saccades comprised at least two different types of saccades: a minority of saccades that were consistent with what would be expected of a corrective saccade (see Abrams et al., 1989; Becker, 1972, 1991) and a majority of saccades that were consistent with what would be expected for a maverick saccade generation process.

Explaining very short duration fixations in this way sets LATEST aside from other models that have attempted to explain this aspect of viewing behaviour, which have required the inclusion of saccade programming time, part of which is uninterruptible (Becker & Juergens, 1979; Nuthmann et al., 2010). LATEST avoids the need to include a

programming phase, a phase that seems unnecessary given neurophysiological evidence (Büttner et al., 1977; Robinson, 1972; Sparks, 1986; Sylvestre & Cullen, 1999).

Having a maverick saccadic decision process that competes with decisions based on on-line visual processing of the scene seems like a somewhat unintuitive component of natural scene viewing. This decision process will necessarily interrupt the main process that underlies decisions about when and where to move the eyes. Such interruptions to scene processing and inspection might be considered as unhelpful for scene exploration and understanding. However, randomness in eye movement behaviour need not be disruptive and indeed may be advantageous (Carpenter, 1999), and it is important to recollect - as mentioned earlier - that it is not just the early saccades that are random: the main decision process has randomness intrinsically built into it. In a competitive situation, the resulting randomness of timing will necessarily cause random selection of the final target. In the context of scene viewing, this will serve to bring the eyes to locations that would otherwise be less likely to be targeted by the main mechanisms controlling saccadic decisions, which may turn out to be unexpectedly useful. A maverick process will add to this fundamentally stochastic mechanism for more venturesome scene exploration.

### **The language of decisions in LATEST**

In this paper we have described the relative weights of a range of underlying factors in saccadic decision time and shown that applying these weights to predict latency to locations in scenes provides an effective model of spatial selection in scene viewing. However, we do not propose that the set of weights identified in the present work are a fixed set of weights for describing saccadic decisions in other situations. That is, we do not expect these particular weights to explain saccadic decisions in other tasks. Rather, we suggest that the underlying theoretical proposal of posing saccadic decisions in space and time as Stay-or-Go decision processes to evaluate the merit of moving to a new location relative to maintaining fixation on the current foveal target are appropriate and

generalisable. However, the relative importance of different sources of information in evaluating these decisions is likely to depend upon what task the viewer is asked to perform.

In this way, when conceptualising the peaks and troughs in the LATEST decision maps they should not be considered as necessarily reflecting any particular form of visual information in the scene. Rather these peaks arise from the evaluation of the Stay-or-Go decision at each location, and will reflect the visual information that is relevant to the demands of the current task. By framing each decision as a Stay-or-Go evaluation, and by acknowledging that the information that contributes to these evaluations will vary depending upon the task of the observer, our proposed decision maps can be seen as fundamentally goal-related, yet modulated at the same time by what is at the current point of fixation. If the content of the map reflects the evidence in favour of anticipated behavioural benefit for moving to each location, relative to maintaining fixation, the map we describe is very similar to a map of anticipated behavioural reward. The possibility that target selection might arise from such a map has been receiving considerable interest in the field (see Tatler et al., 2011, for a review and discussion).

The eye movement circuitry is sensitive to reward (Dorris & Glimcher, 2004; Glimcher, 2003; Glimcher & Fehr, 2013; Platt & Glimcher, 1999; Stuphorn & Schall, 2006; Stuphorn, Taylor, & Schall, 2000; Sugrue, Corrado, & Newsome, 2004). Furthermore, external manipulations of reward (e.g., monetary reward) influence manual and oculomotor behaviour in a manner that appears to maximise reward (e.g., Seydell, McCann, Trommershäuser, & Knill, 2008; Stritzke, Trommershäuser, & Gegenfurtner, 2009; Trommershäuser, Maloney, & Landy, 2003). Given that behavioural goals require visual information for their completion, visual information in itself is a behavioural reward, and in evoked tasks distributions of saccadic latencies shift in manners that suggest that the underlying evaluation reflects the anticipated reward associated with saccading to a target (Bray & Carpenter, 2015). Navalpakkam, Koch, Rangel, and Perona (2010) showed that visual search behaviour can be well explained by a Bayesian optimal observer that accounts

for reward and stimulus detectability. Reward-based models of gaze selection have been proposed (Ballard & Hayhoe, 2009; Rothkopf & Ballard, 2009; Rothkopf, Ballard, & Hayhoe, 2007) and have been used successfully to model goal directed behaviour in virtual reality (Sprague, Ballard, & Robinson, 2007).

Where our suggestion departs from the reward-based models suggested by Ballard and colleagues is that they suggest that anticipated reward is calculated across all sub-tasks (which are tied to spatial locations) and the winner receives fixation. In contrast, we suggest that each location is associated with an independent decision about whether it will benefit the viewer to move their eyes to that location rather than maintain fixation at the current location. Each of these independent decision processes then races to threshold and the winner receives the next fixation. In many ways the end result will be very similar in these two accounts, but the underlying decision process is very different. A further departure is that in the model suggested by Sprague et al. (2007), decisions are based on a snapshot evaluation of anticipated reward made every 300 ms whereas in LATEST the temporal evolution of the supporting evidence in favour of moving the eyes to each location is fundamental to the model and is what drives the eventual selection in space and time.

Any decision map that is based on anticipated behavioural benefit for making a particular eye movement has the advantage that the evidence that contributes to the decisions need not be fixed, but can vary depending upon the task demands. Indeed, we make no assumptions as to the nature of the evidence that contributes to the Stay and Go hypotheses that are evaluated for each decision under LATEST. The evidence can therefore take the form of anything from basic low-level features, through intermediate-level structures such as objects, to high-level semantics and goal-relevance. A key aim for future work is to explore what information may contribute evidence to the evaluation of saccadic decisions across variations to task and stimuli.

## Next steps for LATEST

LATEST is a first step toward better understanding the decision processes that underlie spatiotemporal selection in scene viewing, and future work will need to consider a number of issues.

Our linear mixed models necessarily use a limited range of factors. Further developments of the model should consider other informational and strategic factors in scene viewing. A particular candidate for improvement is to consider a greater range of descriptions of visual information in scenes. Our current implementation of semantic interest is restricted to data from observers performing a rating experiment, partly because of the context of a free viewing task. A better-defined viewing task might offer a more objective measure of higher level information in scenes: for example, when searching for people, computational people detectors offer an appropriate description of task-relevant high level information in scenes (e.g., Ehinger et al., 2009; Torralba et al., 2006). Further levels of description might also benefit our ability to describe fixation duration. A particular omission from the present model might be the lack of object-level description of scene content. It is increasingly clear that object-level descriptions are important components of models of spatial selection (e.g., Einhäuser et al., 2008; Wischniewski et al., 2010).

Future work should consider the level of description of scene content at which Stay-or-Go evaluations are organized and calculated. At present, we calculate expected decision rate for every pixel in the image. Basing a model on the arbitrary and artificial unit of a pixel in an image is of course unrealistic. But this raises the question about what spatial granularity is appropriate for the multiple racing Stay-or-Go decision processes proposed in LATEST. In line with recent suggestions that it may be more appropriate to model spatial selection as being organised around objects rather than pixels in models of eye guidance in scene viewing (Nuthmann & Henderson, 2010), we could propose that the LATEST Stay-or-Go evaluations are calculated at each object in the scene. However, there

are some advantages to a framework that is blind to such high-level segmentation in scenes: specifically, objects themselves may not always be the unit around which selection is organised. For example, in some circumstances, an optimal strategy for information sampling is to fixate locations that fall between objects. Such strategies are seen in expert chess players viewing chess boards (Reingold & Charness, 2005), search behaviour (Zelinsky, Rao, Hayhoe, & Ballard, 1997) and can be important strategic components of models of search (Zelinsky, 2012). Thus while using pixels as the unit of calculation is entirely arbitrary, describing the locations at which Stay-or-Go decisions are evaluated in a manner that is blind higher-level scene structure may be useful.

As a model of screen-based scene viewing, LATEST could be improved by taking better account of known constraints that operate when viewing images on screens. It is important to remember that we constructed our model from data gathered while people look at images on a computer screen and as such what we describe is a model of this situation not of all viewing settings. Screen-based viewing is necessarily bounded by the frame of the monitor on which the scene is displayed and this will impart particular constraints upon inspection behaviour. Not only does screen-based viewing involve a marked bias for observers to fixate near the middle of the scene irrespective of its content (Tatler, 2007), but also fixations are (unsurprisingly) rarely made to locations outside the screen (0.35% of fixations in our dataset). These boundary constraints on viewing mean that the likelihood of a saccade being followed by another in the same direction will depend upon where in the scene the saccade lands: saccades that land close to an edge of the scene are necessarily less likely to be followed by continuations in the same direction. Taking into account where on the screen saccades are launched from when modelling image-independent biases in scene viewing greatly improves the ability to describe eye movement data (Clarke et al., in prep). In its present form, LATEST accounts for information content at the saccade launch site, but is ignorant of these screen-position-contingent viewing constraints and will thus underestimate the probability of changing direction after a saccade that



brings the eye close to a screen boundary. It is likely that taking better account of these viewing constraints would improve the fit of the model. However, the aim of the present work was not to produce a model that is tuned to best describe screen based viewing. Rather the aim was to consider whether an account of saccade timing will also provide an account of spatial selection in order to evaluate the suitability of LATER-like decision processes for explaining both when and where we move the eyes.

## CONCLUSION

In this paper we proposed a new theoretical framework for explaining both when and where people move their eyes when viewing images of real world scenes. It is based on a Stay-or-Go decision process that evaluates the relative merits of moving to a new location or maintaining the current fixation. We implemented and tested this framework by developing an empirically-based model of gaze control called LATEST. This model turned out to predict spatial selection better than a range of previous models. Our data therefore suggest there is no need to suppose separate mechanisms that determine where the eyes will travel and when they will travel there. Instead, saccadic decisions both in space and also in time can be explained as arising from a common underlying Stay-or-Go decision process.

## Footnotes

<sup>1</sup>In the SR Research algorithm, blinks terminate fixation events in the dataset. For all analyses we considered the two fixation events that fall either side of a blink as separate fixations

<sup>2</sup>Fixations were excluded from any trial on which the pre-trial calibration check yielded an error in gaze position greater than 1 deg (3803 fixations) or where calculated fixation duration was less than 10 ms (5 fixations). To ensure the reliability of the fixation events identified by the SR Research algorithm we employed precision criteria for the sample-to-sample characteristics of samples recorded within fixation events. Here we used two precision measures: root mean squared error (RMS) and bivariate contour ellipse area (BCEA). The former measure considers only the between-sample distances for each pair of samples within a fixation. The latter fits an ellipse around all samples from a fixation event such that 63.2% of samples are contained within it. For our data we used a simplification of BCEA in which the two dimensions of the fitted ellipse are averaged to provide a unidimensional measure of the dispersion of samples within each fixation event,  $r(\text{BCEA})$ . For a consideration of these and other metrics for assessing eye tracking data quality see Blignaut and Beelders (2012). We employed a criterion RMS of 0.03 degrees and a criterion  $r(\text{BEAC})$  of 0.15 degrees. This procedure excluded 6,199 fixations from subsequent analyses

<sup>3</sup>The `lmer()` function returns  $t$ -values but no associated  $p$ -values. Associated  $p$ -values were calculated by creating models that selectively left out each factor of interest and comparing these using the `anova()` function in R.

<sup>4</sup>This model was chosen by first selecting four candidate salience models that have been evaluated as part of the MIT saliency benchmark project (Bylinskii et al., n.d.) and have freely available Matlab code for their implementation. Models were selected based upon three desirable criteria: (1) that they are based on bottom-up computations, rather than including task-based parameters that describe higher-level sources of information for particular tasks (such as the person detector in Ehinger et al., 2009); (2) that they do not include a central weighting to improve their fit to human data, due to the prominent tendency for humans to fixate the centre of the screen irrespective of displayed content (Tatler, 2007), and (3) that they involved some combination of features and computation with respect to context in order to make the resultant maps somewhat sparse and qualitatively different from feature maps such as those we used for edge information. The four candidate models evaluated were AWS (Garcia-Diaz et al., 2012), Image Signature (Hou et al., 2012), RARE2012 (Riche et al., 2013) and the salience toolbox (Walther & Koch, 2006). We ran the LMM described below for predicting decision rate four times, each time with a different one of the four candidate salience models providing the salience description for the model. RARE2012 was selected as it provided the LMM with the best overall fit to the data (assessed via `anova()` model comparisons and comparisons of

AIC). Thus we use RARE2012 as our description of intermediate-level visual information in scenes for the analyses that follow.

## References

- Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and Accuracy of Saccadic Eye Movements: Characteristics of Impulse Variability in the Oculomotor System. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 529–543.
- Anderson, A. J., Yadav, H., & Carpenter, R. H. S. (2008). Directional prediction by the saccadic system. *Current Biology*, *18*(8), 614–618.
- Andrews, T. J. & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, *39*(17), 2947–2953.
- Baddeley, R. J. & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research*, *46*(18), 2824–2833.
- Ballard, D. H. & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, *17*(6-7), 1185–1204.
- Basso, M. A. & Wurtz, R. H. (1998). Modulation of neuronal activity in superior colliculus by changes in target probability. *The Journal of Neuroscience*, *18*(18), 7519–7534.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: linear mixed-effects models using eigen and s4*. R package version 1.1-6. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 20–30.
- 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 and visual dysfunction*. (Vol. 8, pp. 95–137). Basingstoke: Macmillan.

- Becker, W. & Juergens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19(9), 967–983.
- Becker, W., Hoehne, O., Iwase, K., & Kornhuber, H. H. (1972). Bereitschaftspotential, prämotorische positivierung und andere hirnpotentiale bei sakkadischen augenbewegungen. *Vision Research*, 12(3), 421–436.
- Blignaut, P. & Beelders, T. (2012). The precision of eye-trackers: a case for a new measure. In *Proceedings of the symposium on eye tracking research and applications* (pp. 289–292). ACM.
- Bonitz, V. S. & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, 129(2), 255–263.
- Borji, A. & Itti, L. (2013). State-of-the-Art in Visual Attention Modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1), 185–207.
- Bray, T. & Carpenter, R. (2015). Saccadic foraging: reduced reaction time to informative targets. *European Journal of Neuroscience*, 41(7), 908–913.
- Brockmole, J. R. & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, 13(1), 99–108.
- Bruce, N. & Tsotsos, J. (2007). Attention based on information maximization. *Journal of Vision*, 7(9), 950–950.
- Buswell, G. T. (1935). *How people look at pictures: a study of the psychology and perception in art*. University of Chicago Press.
- Büttner, U., Hepp, K., & Henn, V. (1977). Neurons in the rostral mesencephalie and paramedian reticular formation generating fast eye movements. In R. Baker & A. Berthoz (Eds.), *Control of gaze by brain stem neurons* (pp. 309–318). Amsterdam: Elsevier.
- Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. (n.d.). Mit saliency benchmark. <http://saliency.mit.edu/>.

- Carpenter, R. H. S. (1981). Oculomotor procrastination. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements* (pp. 237–246). Hillsdale: Lawrence Erlbaum: Eye Movements: Cognition and Visual Perception.
- Carpenter, R. H. S. (1994). Spic: a pc-based system for rapid measurement of saccadic responses. *Journal of Physiology (Proceedings)*, *480*, 4P.
- Carpenter, R. H. S. (2001). Express saccades: is bimodality a result of the order of stimulus presentation? *Vision Research*, *41*(9), 1145–1151.
- Carpenter, R. H. S. (2012). Analysing the detail of saccadic reaction time distributions. *Biocybernetics and Biomedical Engineering*, *32*(2), 49–63.
- Carpenter, R. H. S. & McDonald, S. A. (2007). LATER predicts saccade latency distributions in reading. *Experimental Brain Research*, *177*(2), 176–183.
- Carpenter, R. H. S. & Williams, M. L. L. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, *377*, 59–62.
- Carpenter, R. (1999). A neural mechanism that randomises behaviour. *Journal of Consciousness Studies*, *6*(1), 13–13.
- Carpenter, R., Reddi, B., & Anderson, A. (2009). A simple two-stage model predicts response time distributions. *The Journal of Physiology*, *587*(16), 4051–4062.
- Castelhano, M. S. & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin and Review*, *18*(5), 890–896.
- Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. (2001). The perceptual aspect of skilled performance in chess: evidence from eye movements. *Memory and Cognition*, *29*(8), 1146–1152.
- Clarke, A. D. & Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision Research*, *102*, 41–51.
- Cohen, M. E. & Ross, L. E. (1977). Saccade latency in children and adults: Effects of warning interval and target eccentricity. *Journal of Experimental Child Psychology*.

- De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317–329.
- Dorris, M. C. & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, *44*(2), 365–378.
- Edwards, A. W. F. (1992). *Likelihood*. Johns Hopkins University Press.
- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: a combined source model of eye guidance. *Visual Cognition*, *17*(6-7), 945–978.
- Einhäuser, W. & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*(5), 1089–1097.
- Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision*, *8*(14).
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). Swift: a dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777.
- Fecteau, J. & Munoz, D. (2006). Saliency, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390.
- Findlay, J. M. & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661–674.
- Garcia-Diaz, A., Leborán, V., Fdez-Vidal, X. R., & Pardo, X. M. (2012). On the relationship between optical variability, visual saliency, and eye fixations: a computational approach. *Journal of Vision*, *12*(6), 17.
- Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision making. *Annual Review of Neuroscience*, *26*(1), 133–179.
- Glimcher, P. W. & Fehr, E. (2013). *Neuroeconomics: decision making and the brain*. Academic Press.

- Goffart, L., Hafed, Z. M., & Krauzlis, R. J. (2012). Visual fixation as equilibrium: evidence from superior colliculus inactivation. *The Journal of Neuroscience*, *32*(31), 10627–10636.
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*(6666), 481–484.
- Guo, K., Mahmoodi, S., Robertson, R. G., & Young, M. P. (2006). Longer fixation duration while viewing face images. *Experimental Brain Research*, *171*(1), 91–98.
- Hanes, D. P. & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427–430.
- Hanes, D. P. & Wurtz, R. H. (2001). Interaction of the frontal eye field and superior colliculus for saccade generation. *Journal of Neurophysiology*, *85*(2), 804–815.
- Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. In *Advances in neural information processing systems* (pp. 545–552).
- Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision*, *9*(1), 24–24.
- Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228.
- Hohenstein, S. & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 166.
- Hou, X., Harel, J., & Koch, C. (2012). Image signature: highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(1), 194–201.
- Huk, A. C. & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *The Journal of Neuroscience*, *25*(45), 10420–10436.



- Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489–1506.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *20*(11), 1254–1259.
- Judd, T., Durand, F., & Torralba, A. (2012). *A benchmark of computational models of saliency to predict human fixations*.
- Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*(6-7), 979–1003.
- Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227.
- Kolmogorov, A. (1941). Confidence limits for an unknown distribution function. *Annals of Mathematical Statistics*, *23*, 525–540.
- Kornhuber, H. H. & Deecke, L. (1965). Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: bereitchaftspotential und refferente potentiale. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, *284*(1), 1–17.
- Krauzlis, R. J., Dill, N., & Kornyló, K. (2002). Activity in the primate rostral superior colliculus during the "gap effect" for pursuit and saccades. *Annals of the New York Academy of Sciences*, *956*(1), 409–413.
- Krauzlis, R. J., Liston, D., & Carello, C. D. (2004). Target selection and the superior colliculus: goals, choices and hypotheses. *Vision Research*, *44*(12), 1445–1451.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16.

- Loftus, G. R. (1985). Picture Perception. Effects of Luminance on Available Information and Information-Extraction Rate. *Journal of Experimental Psychology: General*, *114*(3), 342–356.
- Lynch, J., Mountcastle, V., Talbot, W., & Yin, T. (1977). Parietal lobe mechanisms for directed visual attention. *Journal of Neurophysiology*, *40*(2), 362–389.
- Mackay, M., Cerf, M., & Koch, C. (2012). Evidence for two distinct mechanisms directing gaze in natural scenes. *Journal of Vision*, *12*(4), 9, 1–12.
- Mannan, S., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, *9*(3), 363–386.
- 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*(6), 1241–1250.
- McDonald, S. A., Carpenter, R. H. S., & Shillcock, R. C. (2005). An anatomically constrained, stochastic model of eye movement control in reading. *Psychological Review*, *112*(4), 814–840.
- McPeck, 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*(4), 1805–1815.
- Munoz, D. P. & Istvan, P. J. (1998). Lateral inhibitory interactions in the intermediate layers of the monkey superior colliculus. *Journal of Neurophysiology*, *79*(3), 1193–1209.
- Munoz, D. P. & Wurtz, R. H. (1995). Saccade-related activity in monkey superior colliculus. i. characteristics of burst and buildup cells. *Journal of Neurophysiology*, *73*(6), 2313–2333.
- Navalpakkam, V. & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 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*, *107*(11), 5232–5237.
- Noorani, I. & Carpenter, R. H. S. (2011). Full reaction time distributions reveal the complexity of neural decision-making. *European Journal of Neuroscience*, *33*(11), 1948–1951.
- Noorani, I. & Carpenter, R. (2016). The later model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, *64*, 229–251.
- Nummela, S. U. & Krauzlis, R. J. (2011). Superior colliculus inactivation alters the weighted integration of visual stimuli. *The Journal of Neuroscience*, *31*(22), 8059–8066.
- Nuthmann, A. & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8).
- Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). Crisp: a computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405.
- Pelz, J. B. & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*(25-26), 3587–3596.
- Platt, M. L. & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*(6741), 233–238.
- R Core Team. (2014). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Ratcliff, R. (2001). Putting noise into neurophysiological models of simple decision making. *Nature Neuroscience*, *4*(4), 336–336.
- Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422.

- Rayner, K., Castelano, M. S., & Yang, J. (2009). Eye Movements When Looking at Unusual/Weird Scenes: Are There Cultural Differences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 254–259.
- Reddi, B. A. J. & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, *3*(8), 827–830.
- Reddi, B., Asrress, K., & Carpenter, R. (2003). Accuracy, information, and response time in a saccadic decision task. *Journal of Neurophysiology*, *90*(5), 3538–3546.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125.
- Reingold, E. M. & Charness, N. (2005). Perception in chess: evidence from eye movements. *Cognitive Processes in Eye Guidance*, 325–354.
- Riche, N., Mancas, M., Duvinage, M., Mibulumukini, M., Gosselin, B., & Dutoit, T. (2013). Rare2012: a multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, *28*(6), 642–658.
- Robinson, D. (1972). Eye movements evoked by collicular stimulation in the alert monkey. *Vision Research*, *12*(11), 1795–1808.
- Robinson, D. L. & Petersen, S. E. (1992). The pulvinar and visual salience. *Trends in Neurosciences*, *15*(4), 127–132.
- Roos, J. C. P., Calandrini, D. M., & Carpenter, R. H. S. (2008). A single mechanism for the timing of spontaneous and evoked saccades. *Experimental Brain Research*, *187*(2), 283–293.
- Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, *39*(19), 3157–3163.
- Rosenholtz, R. (2001). Search asymmetries? What search asymmetries? *Perception and Psychophysics*, *63*(3), 476–489.
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17.1–22.

- Rothkopf, C. A. & Ballard, D. H. (2009). Image statistics at the point of gaze during human navigation. *Visual Neuroscience*, *26*(01), 81–92.
- Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14), 16.
- RStudio Team. (2015). *Rstudio: integrated development environment for r*. RStudio, Inc. Boston, MA. Retrieved from <http://www.rstudio.com/>
- Schall, J. D. & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience*, *22*(1), 241–259.
- Schütz, A. C., Trommershäuser, J., & Gegenfurtner, K. R. (2012). Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Science*, *109*(19), 7547–7552.
- Seydell, A., McCann, B. C., Trommershäuser, J., & Knill, D. C. (2008). Learning stochastic reward distributions in a speeded pointing task. *The Journal of Neuroscience*, *28*(17), 4356–4367.
- Shadlen, M. N. & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936.
- Smith, T. J. & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6-7), 1083–1108.
- Sparks, D. L. (1986). Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus. *Physiological Reviews*, *66*(1), 118–171.
- Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes. *Journal of Vision*, *15*(2), 12, 1–21.
- Sprague, N., Ballard, D., & Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Transactions on Applied Perception (TAP)*, *4*(2), 11.

- Story, G. W. & Carpenter, R. (2009). Dual later-unit model predicts saccadic reaction time distributions in gap, step and appearance tasks. *Experimental Brain Research*, *193*(2), 287–296.
- Stritzke, M., Trommershäuser, J., & Gegenfurtner, K. R. (2009). Effects of salience and reward information during saccadic decisions under risk. *JOSA A*, *26*(11), B1–B13.
- Stuphorn, V. & Schall, J. D. (2006). Executive control of countermanding saccades by the supplementary eye field. *Nature Neuroscience*, *9*(7), 925–931.
- Stuphorn, V., Taylor, T. L., & Schall, J. D. (2000). Performance monitoring by the supplementary eye field. *Nature*, *408*(6814), 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*(5678), 1782–1787.
- Sylvestre, P. A. & Cullen, K. E. (1999). Quantitative analysis of abducens neuron discharge dynamics during saccadic and slow eye movements. *Journal of Neurophysiology*, *82*(5), 2612–2632.
- Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., & Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Experimental Brain Research*, *142*(2), 284–291.
- 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).
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, *45*(5), 643–659.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision*, *11*(5), 5.
- Tatler, B. W. & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*.

- Tatler, B. W. & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6-7), 1029–1054.
- Thompson, K. G. & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, *147*, 249–262.
- Thompson, K. G., Bichot, N. P., & Schall, J. D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *Journal of Neurophysiology*, *77*(2), 1046–1050.
- Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology*, *76*(6), 4040–4055.
- Torralba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, *113*(4), 766.
- Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and the selection of rapid, goal-directed movements. *JOSA A*, *20*(7), 1419–1433.
- Underwood, G., Humphreys, L., & Cross, E. (2007). Congruency, saliency and gist in the inspection of objects in natural 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. 563–579). Oxford: Elsevier.
- Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition*, *17*(1), 159–170.
- Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494.

- Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 145–156.
- Vincent, B. T., Troscianko, T., & Gilchrist, I. D. (2007). Investigating a space-variant weighted salience account of visual selection. *Vision Research*, 47(13), 1809–1820.
- Walther, D. & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19(9), 1395–1407.
- Watanabe, K., Lauwereyns, J., & Hikosaka, O. (2003). Neural correlates of rewarded and unrewarded eye movements in the primate caudate nucleus. *The Journal of Neuroscience*, 23(31), 10052–10057.
- Wischnewski, M., Belardinelli, A., Schneider, W. X., & Steil, J. J. (2010). Where to look next? combining static and dynamic proto-objects in a tva-based model of visual attention. *Cognitive Computation*, 2(4), 326–343.
- Yang, T. & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, 447(7148), 1075–1080.
- Zelinsky, G. J. (2012). Tam: explaining off-object fixations and central fixation tendencies as effects of population averaging during search. *Visual Cognition*, 20(4-5), 515–545.
- Zelinsky, G. J., Rao, R. P., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, 448–453.
- Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). Sun: a bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7), 32.



Table 1

*Output of LMM models to predict decision rate in the main population of saccades. Results are shown for the full model containing all evaluated fixed effects and for the most reduced version of the model.*

Fixed Effects	Full model				Minimal model			
	Estimate	SE	<i>t</i>		Estimate	SE	<i>t</i>	
(Intercept)	3.2924	0.049	67.03	***	3.2933	0.049	67.03	***
Ordinal fixation number	-0.0368	0.004	-10.13	***	-0.0367	0.004	-10.12	***
Change in direction between saccades	-0.0860	0.004	-23.79	***	-0.0860	0.004	-23.82	***
Incoming saccade amplitude (linear)	0.0782	0.010	7.62	***	0.0777	0.010	7.59	***
Incoming saccade amplitude (quadratic)	-0.0445	0.010	-4.38	***	-0.0437	0.010	-4.31	***
Outgoing saccade amplitude (linear)	0.0346	0.011	3.26	**	0.0322	0.010	3.16	***
Outgoing saccade amplitude (quadratic)	-0.0411	0.011	-3.67	***	-0.0379	0.010	-3.66	***
Change in direction × incoming saccade amplitude (linear)	-0.0342	0.010	-3.37	***	-0.0189	0.004	-5.40	***
Change in direction × incoming saccade amplitude (quadratic)	0.0163	0.010	1.61	n.s.	—	—	—	
Change in direction × outgoing saccade amplitude (linear)	0.0223	0.010	2.17	*	0.0248	0.004	6.50	***
Change in direction × outgoing saccade amplitude (quadratic)	0.0033	0.011	0.31	n.s.	—	—	—	
Edge information at fixation	-0.0154	0.004	-3.73	***	-0.0147	0.004	-3.58	***
Edge information at target location	0.0110	0.004	2.70	**	-0.0147	0.004	2.76	**
Edge information at target location × outgoing saccade amplitude	-0.0052	0.004	-1.38	n.s.	—	—	—	
Saliency at fixation	-0.0165	0.004	-3.73	***	-0.0145	0.004	-3.70	***
Saliency at target location	0.0003	0.004	0.07	n.s.	—	—	—	
Saliency at target location × outgoing saccade amplitude	-0.0052	0.004	-1.36	n.s.	—	—	—	
Semantic interest at fixation	-0.0653	0.004	-16.03	***	-0.0659	0.004	-16.22	***
Semantic interest at target location	0.0288	0.004	7.43	***	0.0286	0.004	7.43	***
Semantic interest at target location × outgoing saccade amplitude	-0.0047	0.004	-1.25	n.s.	-0.0071	0.004	-1.96	*
Random effects Variance								
Subjects	0.163				0.164			
Scenes	0.003				0.003			
Log-likelihood	-67738.507				-67721.281			
Deviance	135477.013				135442.562			
AIC	135523.013				135478.562			
BIC	135728.202				135639.145			
N	55341				55341			

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Table 2

*AUC values for LATEST and each of the eight evaluated salience models.*

	AUC
LATEST	0.6205
AIM	0.6012
AWS	0.6267
FC	0.5889
GBVS	0.5546
IS	0.6170
RARE	0.6152
STB	0.5212
SUN	0.6015

Table 3

*Output of LMM models to predict decision rate in the early population of saccades. Results are shown for the full model containing all evaluated fixed effects and for the most reduced version of the model.*

Fixed Effects	Full model				Minimal model			
	Estimate	SE	<i>t</i>		Estimate	SE	<i>t</i>	
(Intercept)	8.4991	0.151	56.23	***	8.4927	0.151	56.17	***
Ordinal fixation number	0.0963	0.031	3.14	***	0.1076	0.030	3.59	***
Change in direction between saccades	-0.1530	0.031	-4.91	***	-0.1558	0.031	-5.02	***
Incoming saccade amplitude (linear)	-0.5902	0.085	-6.91	***	-0.5696	0.084	-6.80	***
Incoming saccade amplitude (quadratic)	0.3908	0.084	4.63	***	0.3744	0.084	4.47	***
Outgoing saccade amplitude (linear)	-0.1727	0.095	-1.82	n.s.	-0.1325	0.093	-1.43	n.s.
Outgoing saccade amplitude (quadratic)	0.2197	0.100	2.20	***	0.1818	0.098	1.85	n.s.
Change in direction × incoming saccade amplitude (linear)	-0.1273	0.082	-1.55	n.s.	-0.1285	0.082	-1.56	n.s.
Change in direction × incoming saccade amplitude (quadratic)	0.1950	0.082	2.37	*	0.1953	0.082	2.37	*
Change in direction × outgoing saccade amplitude (linear)	-0.2091	0.088	-2.36	*	-0.2156	0.088	-2.44	*
Change in direction × outgoing saccade amplitude (quadratic)	0.2624	0.090	2.90	**	0.2741	0.090	3.04	**
Edge information at fixation	-0.0060	0.032	-0.18	n.s.	—	—	—	
Edge information at target location	-0.0011	0.032	-0.04	n.s.	—	—	—	
Edge information at target location × outgoing saccade amplitude	0.0206	0.032	0.65	n.s.	—	—	—	
Saliency at fixation	-0.0028	0.035	-0.08	n.s.	—	—	—	
Saliency at target location	0.0133	0.034	0.40	n.s.	—	—	—	
Saliency at target location × outgoing saccade amplitude	0.0015	0.033	0.05	n.s.	—	—	—	
Semantic interest at fixation	-0.0385	0.034	-1.14	n.s.	—	—	—	
Semantic interest at target location	-0.0283	0.033	-0.87	n.s.	—	—	—	
Semantic interest at target location × outgoing saccade amplitude	0.0270	0.032	0.84	n.s.	—	—	—	
Random effects Variance								
Subjects	1.526				1.529			
Scenes	< 0.001				< 0.001			
Log-likelihood	-43916.929				-43896.767			
Deviance	87833.857				87793.534			
AIC	87879.857				87821.534			
BIC	88056.528				87929.073			
N	16016				16016			

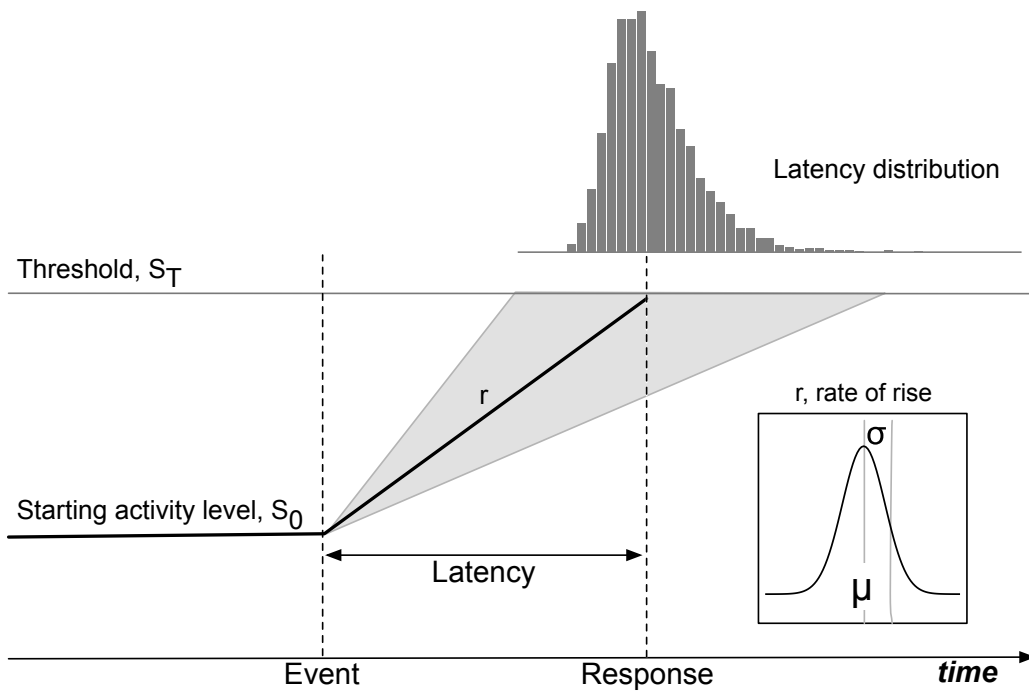
\* p<.05, \*\* p<.01, \*\*\* p<.001

Table 4

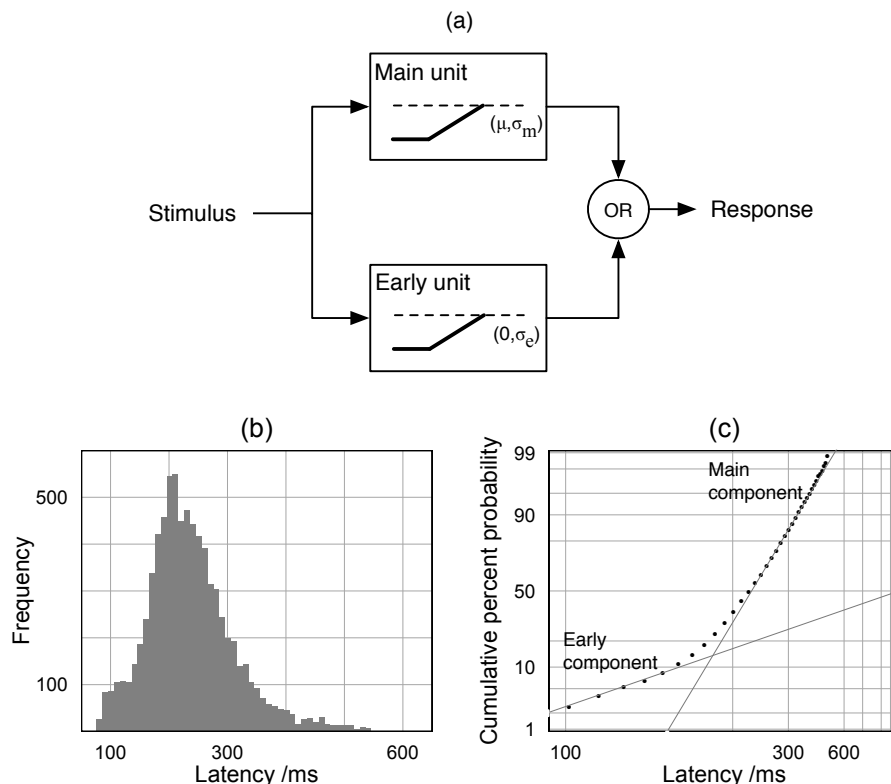
*Output of LMM models to predict decision rate in the early population of saccades, for those saccades identified as likely to be corrective and those that are unlikely to be corrective saccades.*

Fixed Effects	'Corrective' saccades				Other early saccades			
	Estimate	SE	<i>t</i>		Estimate	SE	<i>t</i>	
(Intercept)	8.3521	0.201	41.52	***	8.4019	0.144	58.53	***
Ordinal fixation number	-0.0899	0.108	-0.83	n.s.	0.1202	0.037	3.24	***
Change in direction between saccades	-0.0765	0.104	-0.74	n.s.	-0.1137	0.037	-3.04	***
Incoming saccade amplitude (linear)	-0.5547	0.323	-1.72	n.s.	-0.6412	0.099	-6.50	***
Incoming saccade amplitude (quadratic)	0.3363	0.319	1.05	n.s.	0.5053	0.097	5.23	***
Outgoing saccade amplitude (linear)	-0.3808	0.691	-0.55	n.s.	-0.1025	0.135	-0.76	n.s.
Outgoing saccade amplitude (quadratic)	0.4004	0.692	0.58	n.s.	0.2134	0.139	1.53	n.s.
Change in direction × incoming saccade amplitude (linear)	-0.2442	0.298	-0.82	n.s.	-0.0708	0.094	-0.75	n.s.
Change in direction × incoming saccade amplitude (quadratic)	0.0967	0.284	0.34	n.s.	0.1100	0.093	1.18	n.s.
Change in direction × outgoing saccade amplitude (linear)	0.0271	0.695	0.04	n.s.	0.0764	0.133	0.58	n.s.
Change in direction × outgoing saccade amplitude (quadratic)	-0.0185	0.691	-0.03	n.s.	0.0706	0.138	0.51	n.s.
Edge information at fixation	-0.0650	0.127	-0.51	n.s.	0.0563	0.038	1.47	n.s.
Edge information at target location	0.0116	0.125	0.09	n.s.	-0.0094	0.038	-0.24	n.s.
Edge information at target location × outgoing saccade amplitude	-0.0231	0.114	-0.20	n.s.	0.0219	0.038	0.58	n.s.
Saliency at fixation	-0.0435	0.192	-0.23	n.s.	-0.0204	0.039	-0.52	n.s.
Saliency at target location	-0.1299	0.187	-0.70	n.s.	0.0423	0.039	1.08	n.s.
Saliency at target location × outgoing saccade amplitude	-0.0696	0.106	-0.66	n.s.	-0.0526	0.039	-1.34	n.s.
Semantic interest at fixation	-0.0676	0.172	-0.39	n.s.	-0.0206	0.039	-0.53	n.s.
Semantic interest at target location	0.0153	0.160	0.10	n.s.	-0.0019	0.038	-0.05	n.s.
Semantic interest at target location × outgoing saccade amplitude	-0.2468	0.111	-2.22	*	0.0215	0.038	0.57	n.s.
Random effects Variance								
Subjects	1.546				1.338			
Scenes	0.247				< 0.001			
Log-likelihood	-3326.821				-25481.16			
Deviance	6653.641				50962.31			
AIC	6699.641				51008.31			
BIC	6817.074				51173.03			
N	1219				9523			

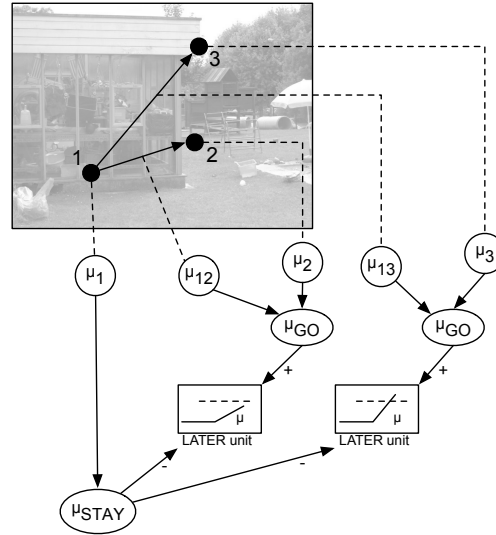
\* p<.05, \*\* p<.01, \*\*\* p<.001



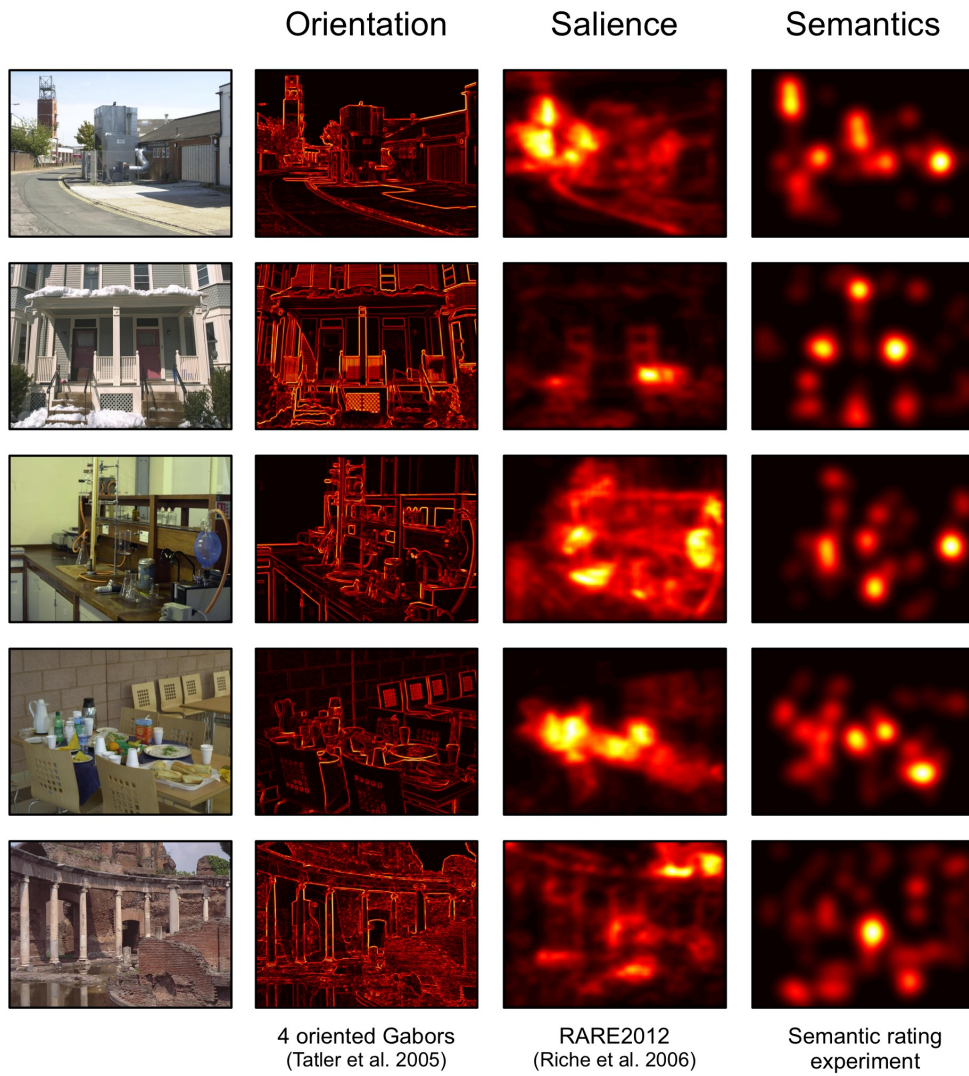
*Figure 1.* Accumulating evidence for saccadic decisions in LATER. After an external event, such as the onset of a stimulus or start of a fixation, the log odds ( $\log Q'$ ) will rise linearly as support ( $\log L$ ) for hypothesis  $H_1$  relative to  $H_2$  accumulates, rising from the starting level,  $S_0$ , which denotes prior expectation, with a rate of rise,  $r$ . The rate of rise varies from decision to decision with a mean of  $\mu$  and standard deviation of  $\sigma$ . A response (for example a saccade) is triggered when  $\log Q'$  reaches a criterion threshold,  $S_T$ , for accepting hypothesis  $H_1$ .



*Figure 2.* (a) The basic architecture for producing two populations of decision times. Two LATER units race to threshold - the main unit conducting sensible evaluations of evidence, the early unit providing a highly variable maverick decision signal - with the winning unit generating a response. (b) Simulated data ( $N = 8000$ ) generated by two competing LATER decision units. For the rates of rise parameters were set to  $\mu = 4$ ,  $\sigma = 1$  for the main unit and  $\mu = 0$ ,  $\sigma = 5$  for the early unit. The histogram shows an over-representation of very short latency simulated responses. (c) The same data as shown in (b), plotted on reciprobital axes. The responses generated by the early unit are easier to identify in this plot. The data points clearly lie along two straight lines (plotted in grey), corresponding to the two underlying response generators.

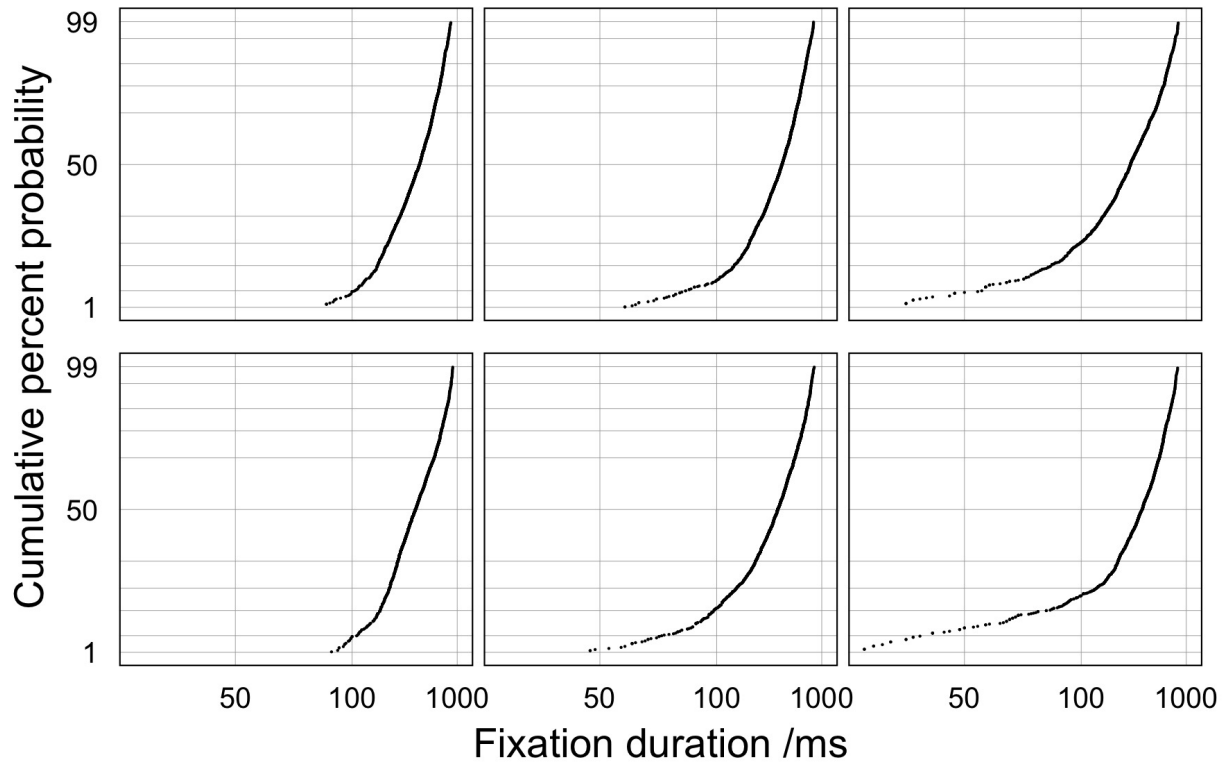


*Figure 3.* Schematic of evaluating potential target locations in LATEST. Here separate Stay-or-Go evaluations race against each other for two peripheral locations (2,3). For each evaluation the Go signal comprises information at the peripheral location ( $\mu_2$ ,  $\mu_3$ ) along with factors associated with the saccade that is required to move to that location ( $\mu_{12}$ ,  $\mu_{13}$ ). The Stay signal for each evaluation derives from the information present at the currently-fixated location ( $\mu_1$ ). The decision signal will rise at different rates in each decision unit, with the saccade being triggered to the location at which the decision signal reaches threshold first. In this case the unit associated with location 3 reaches threshold first and the outcome of this schematic example would be a saccade to location 3.



*Figure 4.* Maps of orientation, saliency and semantic ratings for five images used in the present study.





*Figure 5.* Reciprobbit plots of the distribution of observed fixation durations for six of our 70 participants. These six participants illustrate that participants varied considerably. All, however, show some over-representation of very fast saccades, seen by the leftward kink in the distribution toward very short durations. Data are shown for participants 8, 6, 34, 55, 13 and 60 from top left to bottom right.

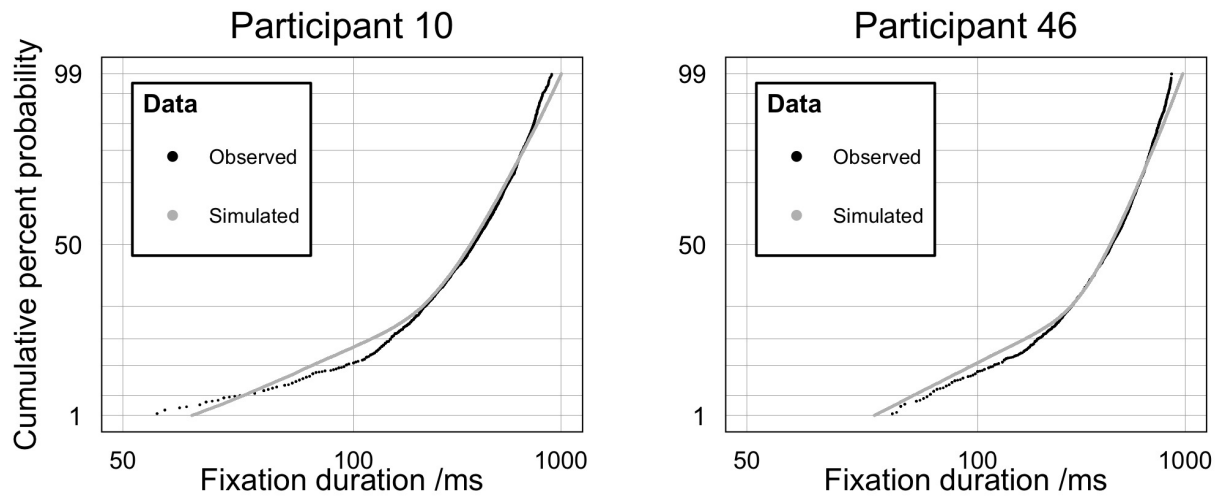


Figure 6. Example plots of observed fixation durations together with simulated fixation durations. Simulations were based upon parameter estimates derived from SPIC.

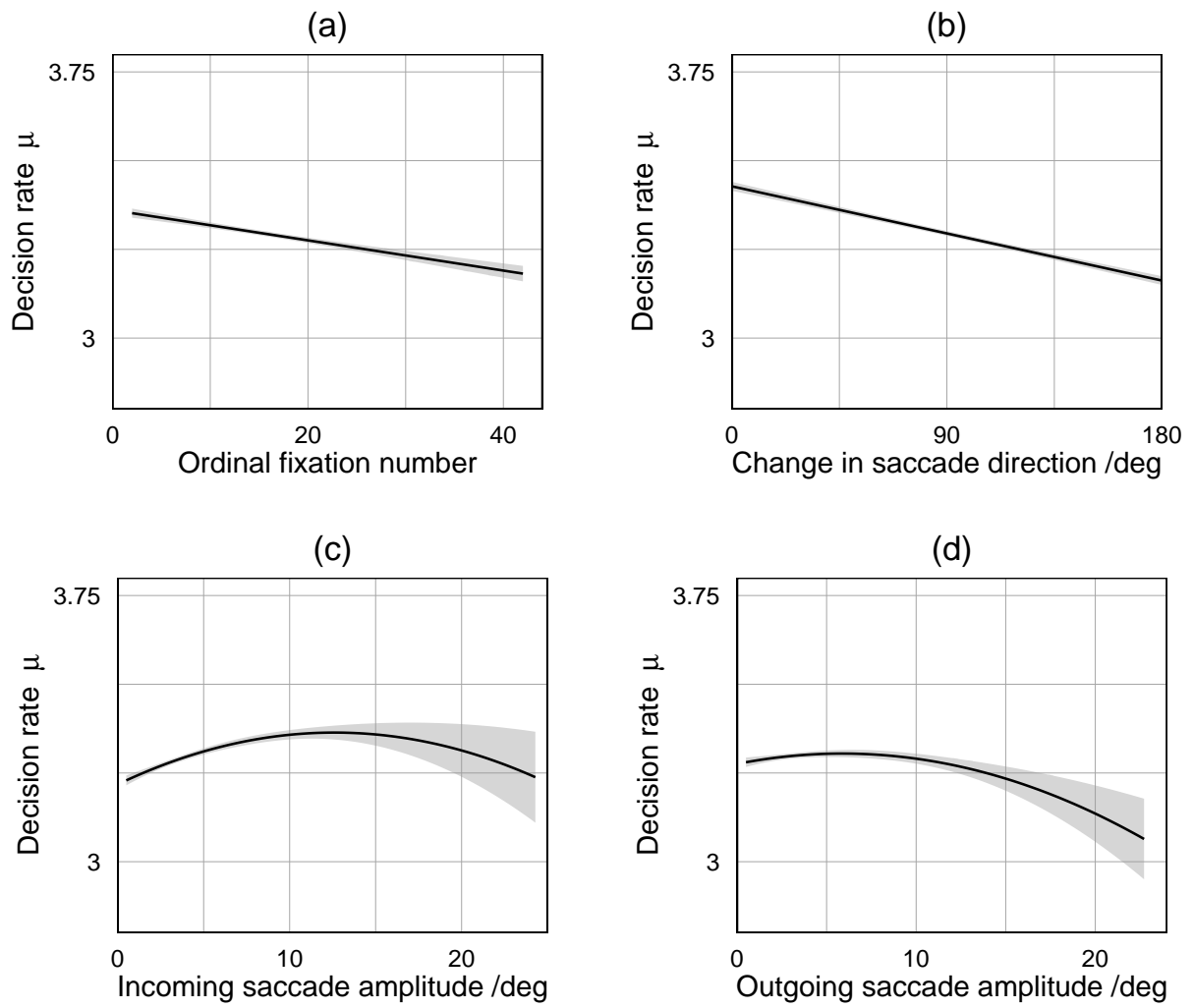


Figure 7. Partial effects for (a) ordinal fixation number, (b) change in saccade direction, (c) incoming saccade amplitude, (d) outgoing saccade amplitude on decision rates in scene viewing. Shaded regions indicate 95% confidence intervals.

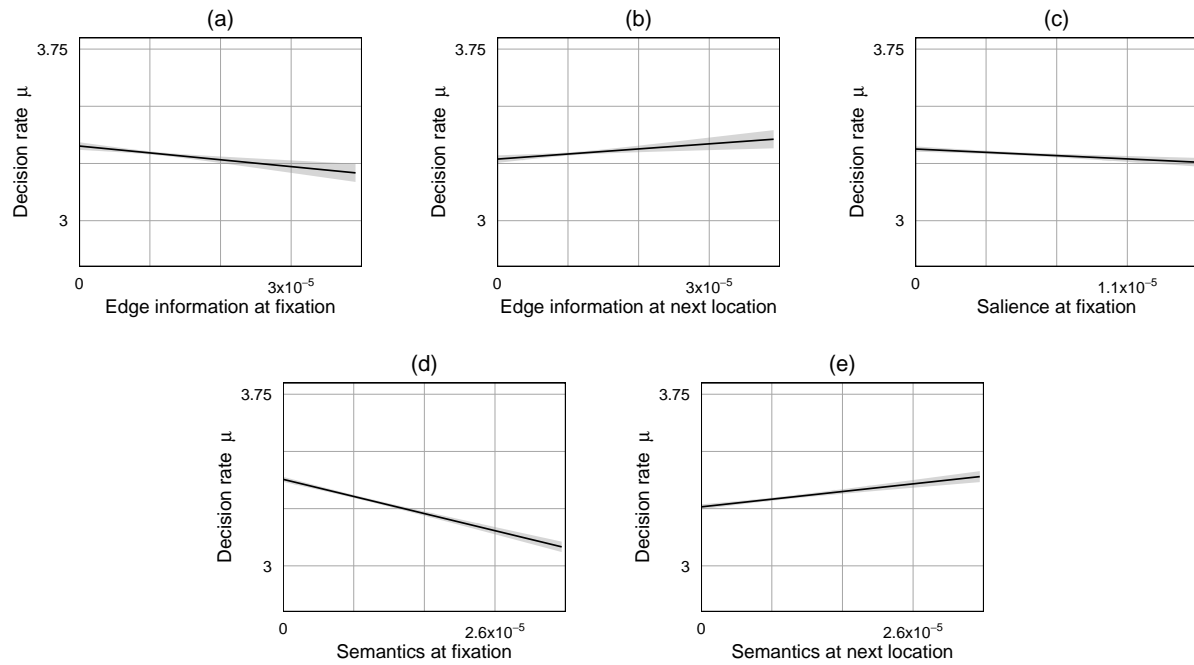
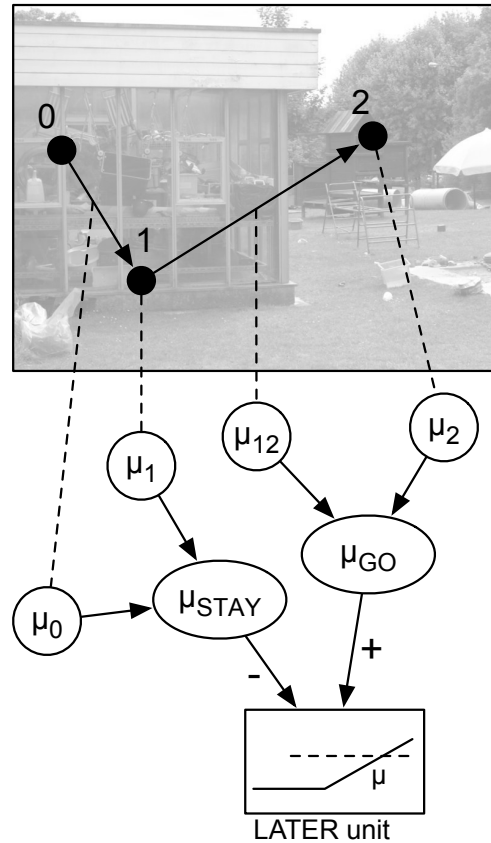
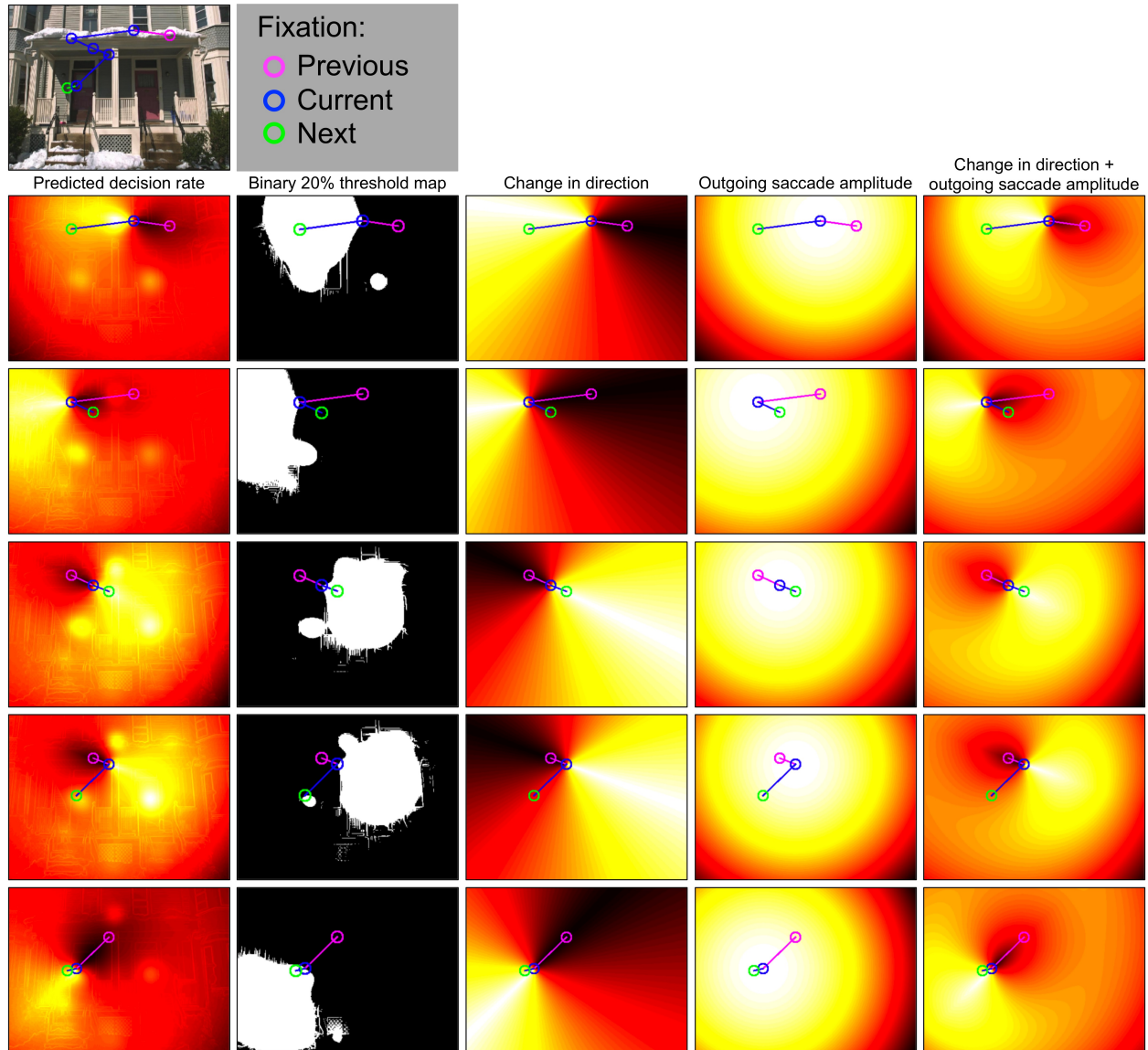


Figure 8. Partial effects for each of the visual predictors on decision rate in scene viewing. Shaded regions indicate 95% confidence intervals.



*Figure 9.* Revised schematic of a saccadic decision. Here location 0 is the previous fixation, 1 is the current fixation and 2 is the next fixation location. In addition to evaluating information from the current location,  $\mu_1$ , next location,  $\mu_2$ , and the retinocentric position of the next location,  $\mu_{12}$ , factors unrelated to these locations,  $\mu_0$ , also appear to contribute to saccade timing. At least in part  $\mu_0$  involves factors related to the incoming saccade.



*Figure 10.* Predicted decision maps for five sequential fixations (shown in top left panel) made by one participant in the present study. The left column shows the predicted decision rate maps for each fixation. The second columns shows binary thresholded maps, showing the pixels with the 20% highest predicted decision rates. The three columns that follow show decision rates predicted only on the basis of change in saccadic direction, outgoing saccade amplitude and the interaction between these two factors. In each panel we plot the current fixation location, the previous fixation location and next observed fixation location. Using the 20% thresholded prediction maps, our model accounts for four of the five fixations depicted here.

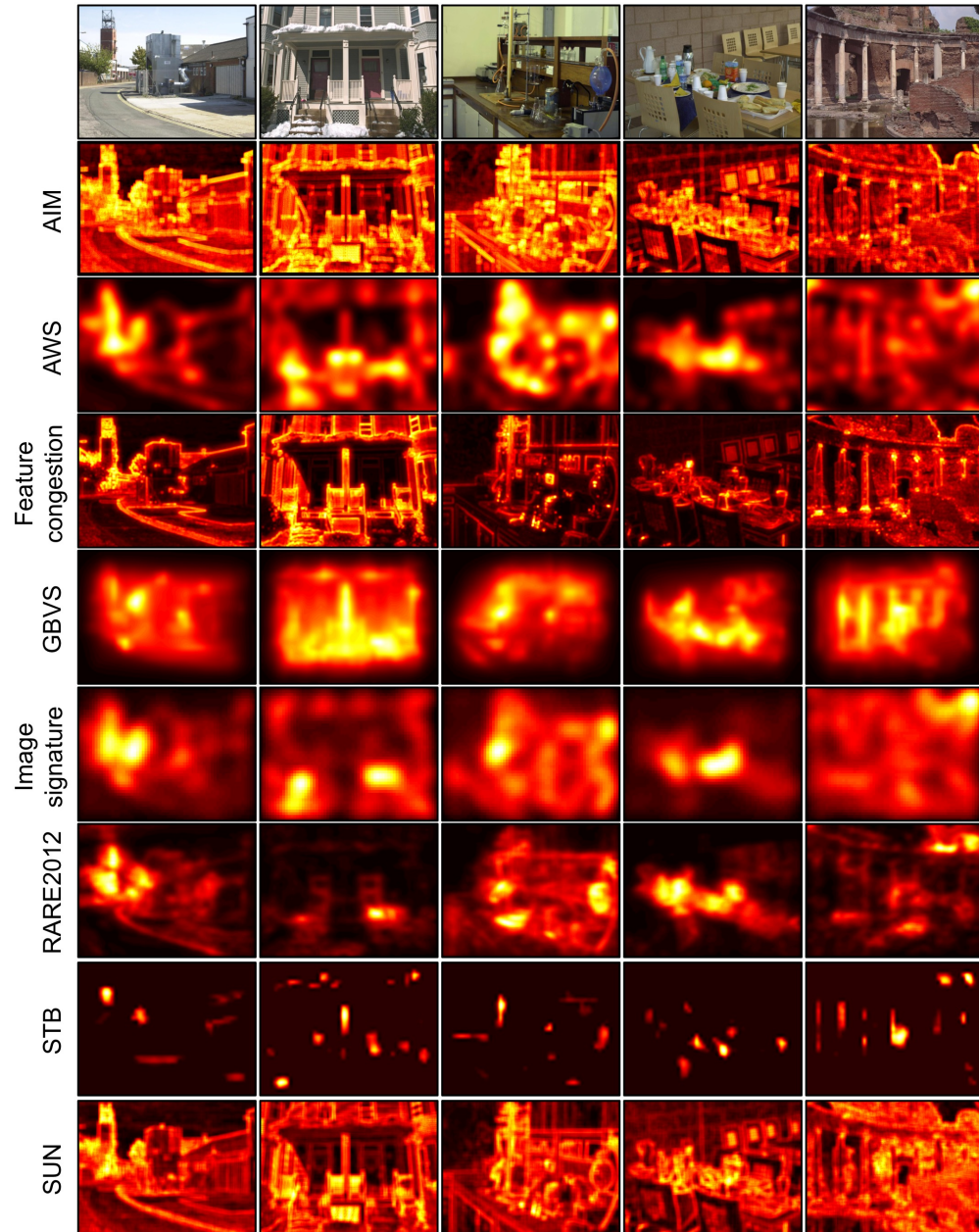
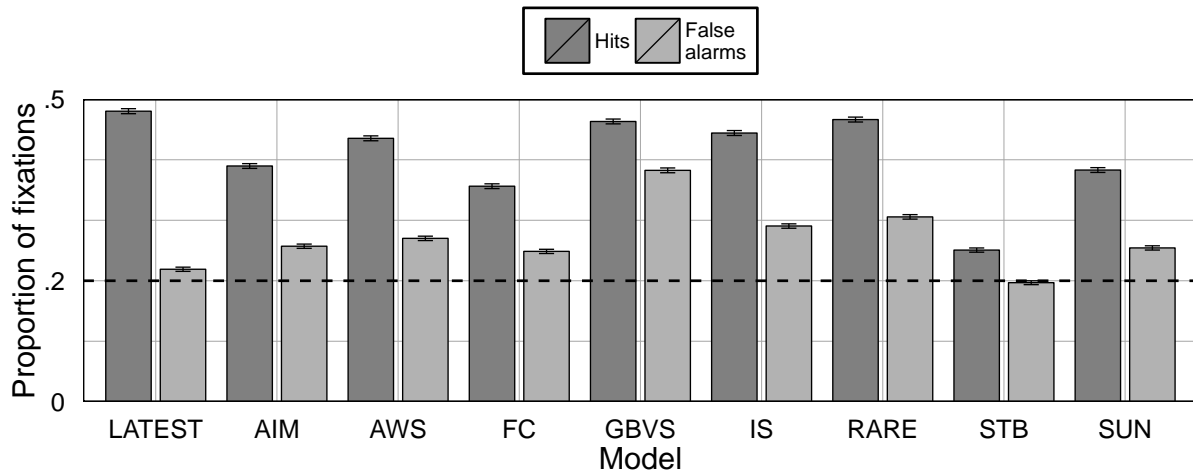


Figure 11. Saliency maps for five example images used in the present study.



*Figure 12.* Model performance (when spatial maps for each model were created with a threshold of 20%) for our proposed spatiotemporal decision model, LATEST, and for eight tested salience models. Dark grey bars show the proportion of observed fixations falling with regions predicted by each model (i.e. hits). Light grey bars show the proportion of control locations falling with regions predicted by each model (i.e. false alarms). Error bars show bootstrapped 95% confidence intervals (10,000 resamples).



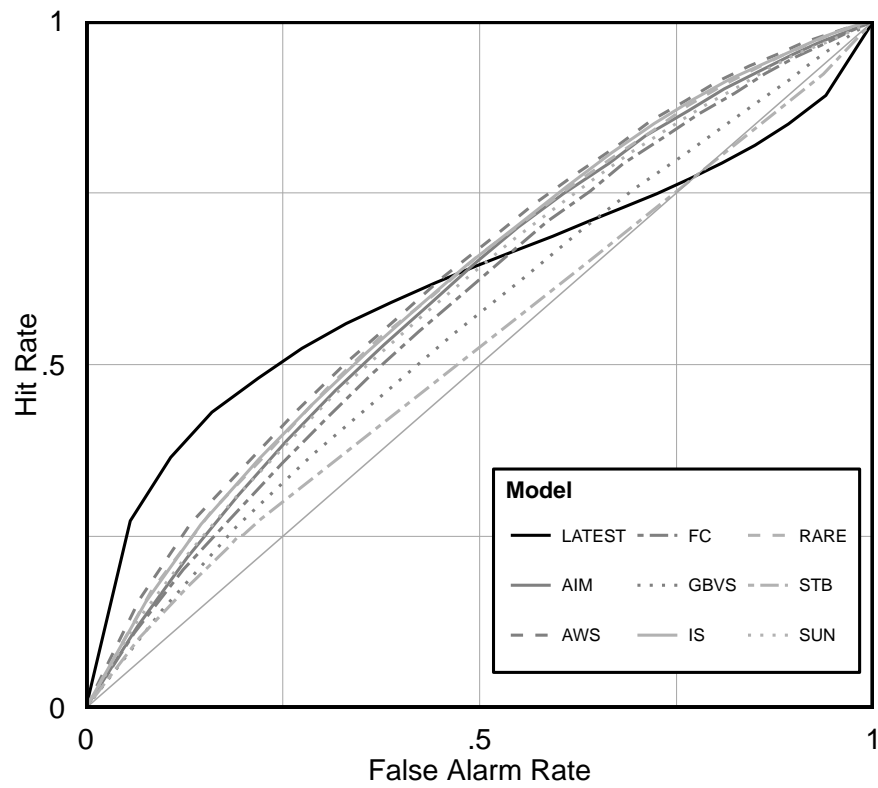
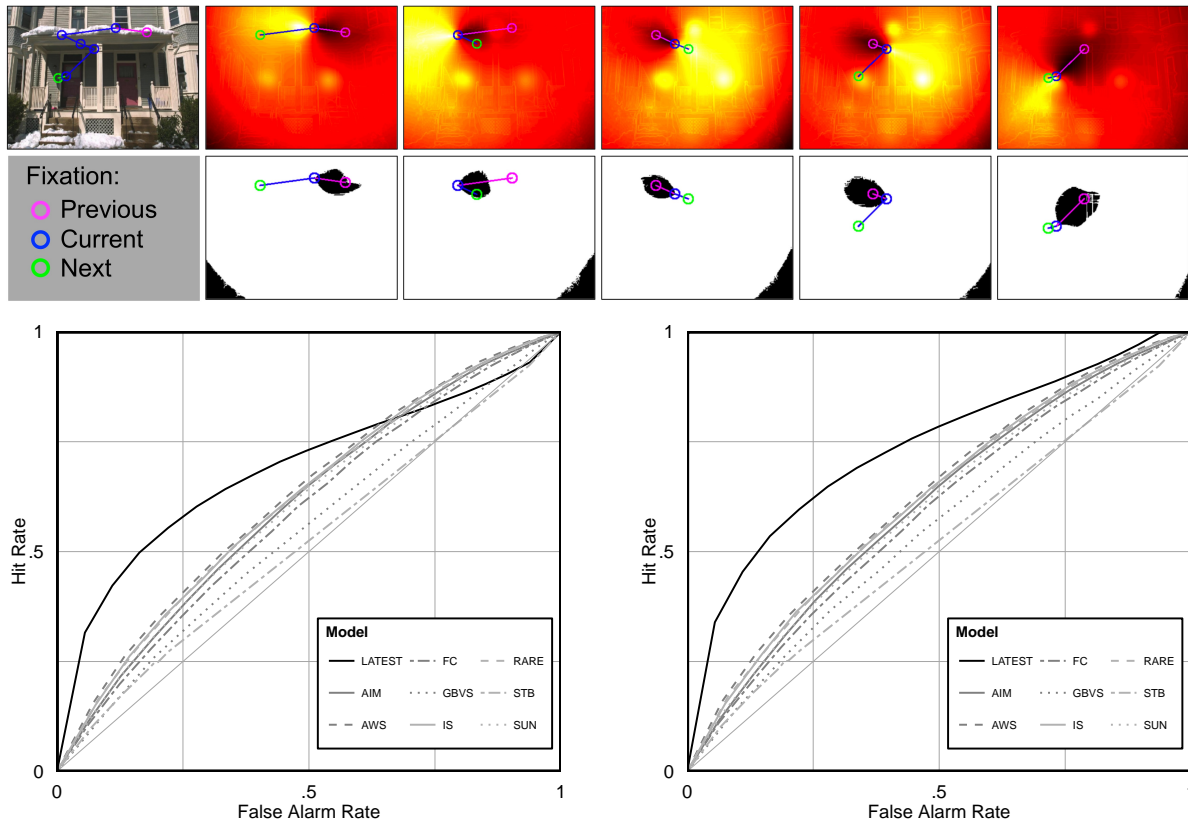


Figure 13. ROC curves for our proposed LATEST model and all eight evaluated saliency models, with false alarm rate calculated using a biased random baseline.



*Figure 14.* The top rows show 95% threshold binary prediction maps for five fixations. These plots suggest that the model considers saccades that reverse direction and land between the current and previous fixation location to be associated with particularly low decision rates and thus under a single framework for predicting spatial selection from temporal accumulation of evidence, these will be locations that the model predicts as highly unlikely to be selected. The bottom left panel shows ROC curves for all models after removing saccades that reverse direction and land between or close to the previous fixation (see text for exclusion criteria). The bottom right panel shows ROC curves for all models after removing those saccades that landed in the 5% least likely pixels in the predicted spatial selection maps generated from LATEST.

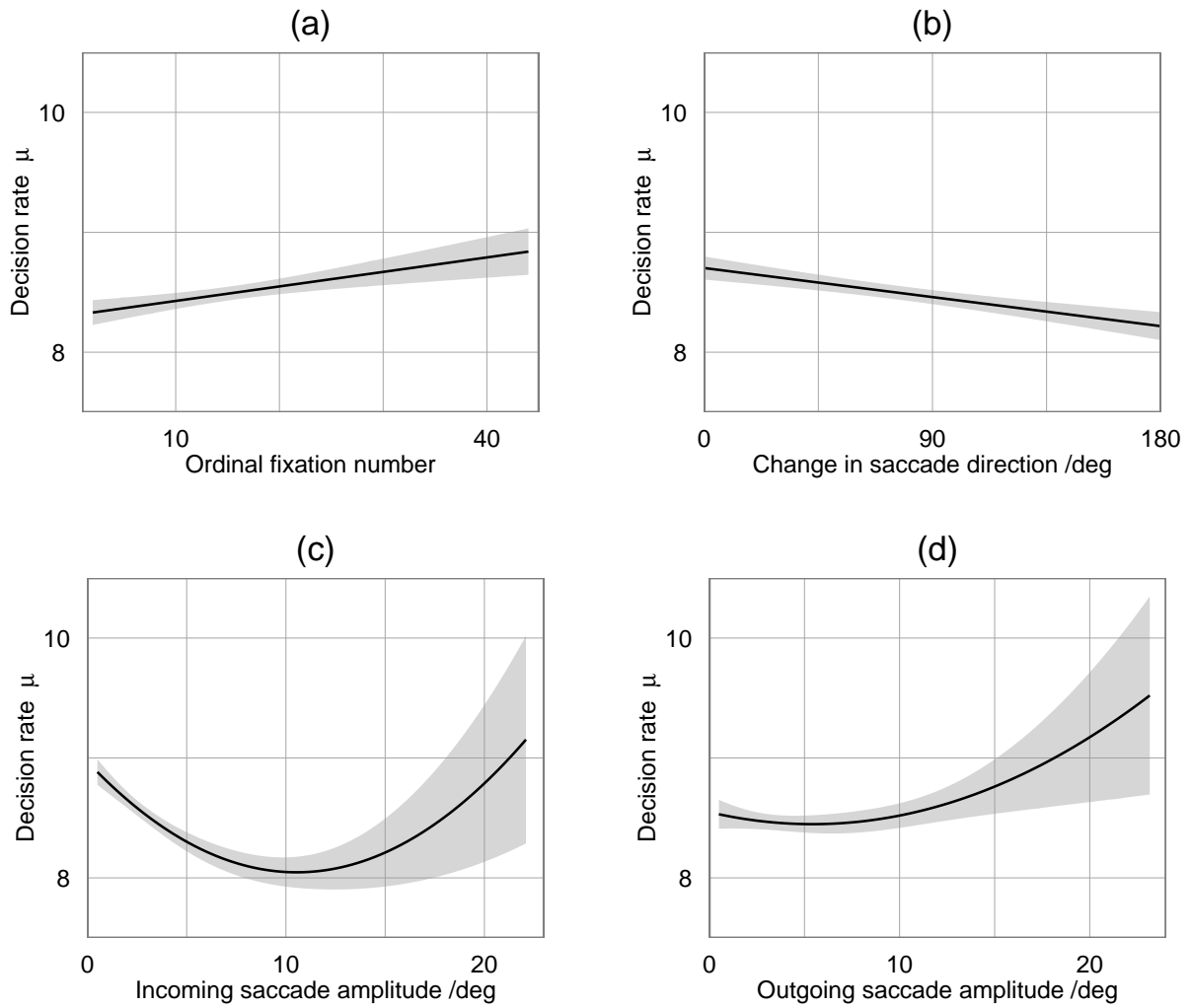
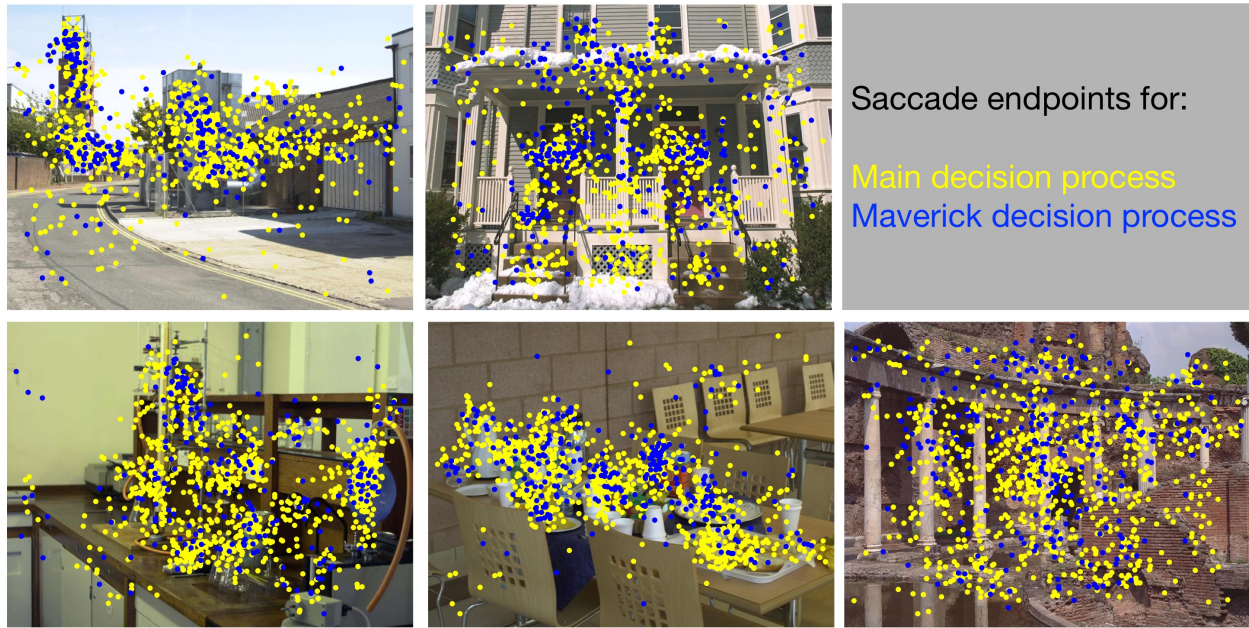
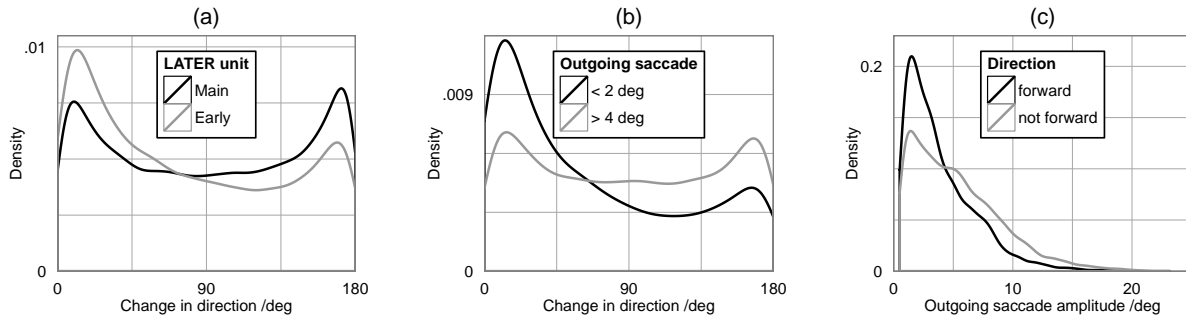


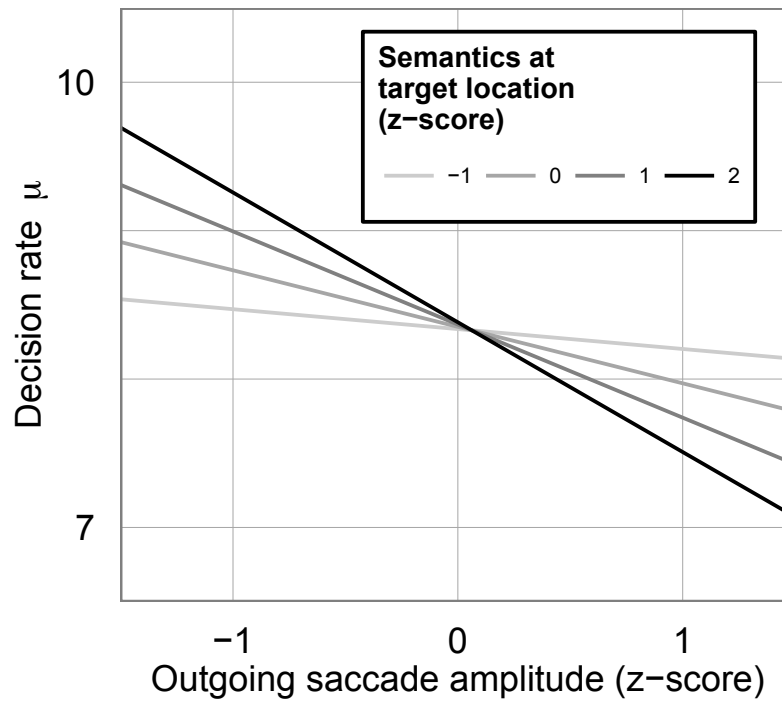
Figure 15. Partial effects for (a) ordinal fixation number, (b) change in saccade direction, (c) previous saccade amplitude, (d) next saccade amplitude on decision rates in scene viewing for the sub-population of very short duration fixations. Shaded regions indicate 95% confidence intervals.



*Figure 16.* Distributions of endpoints for saccades likely to have been generated by the main decision unit (yellow dots) or the maverick, early decision unit (blue dots) for five example images.

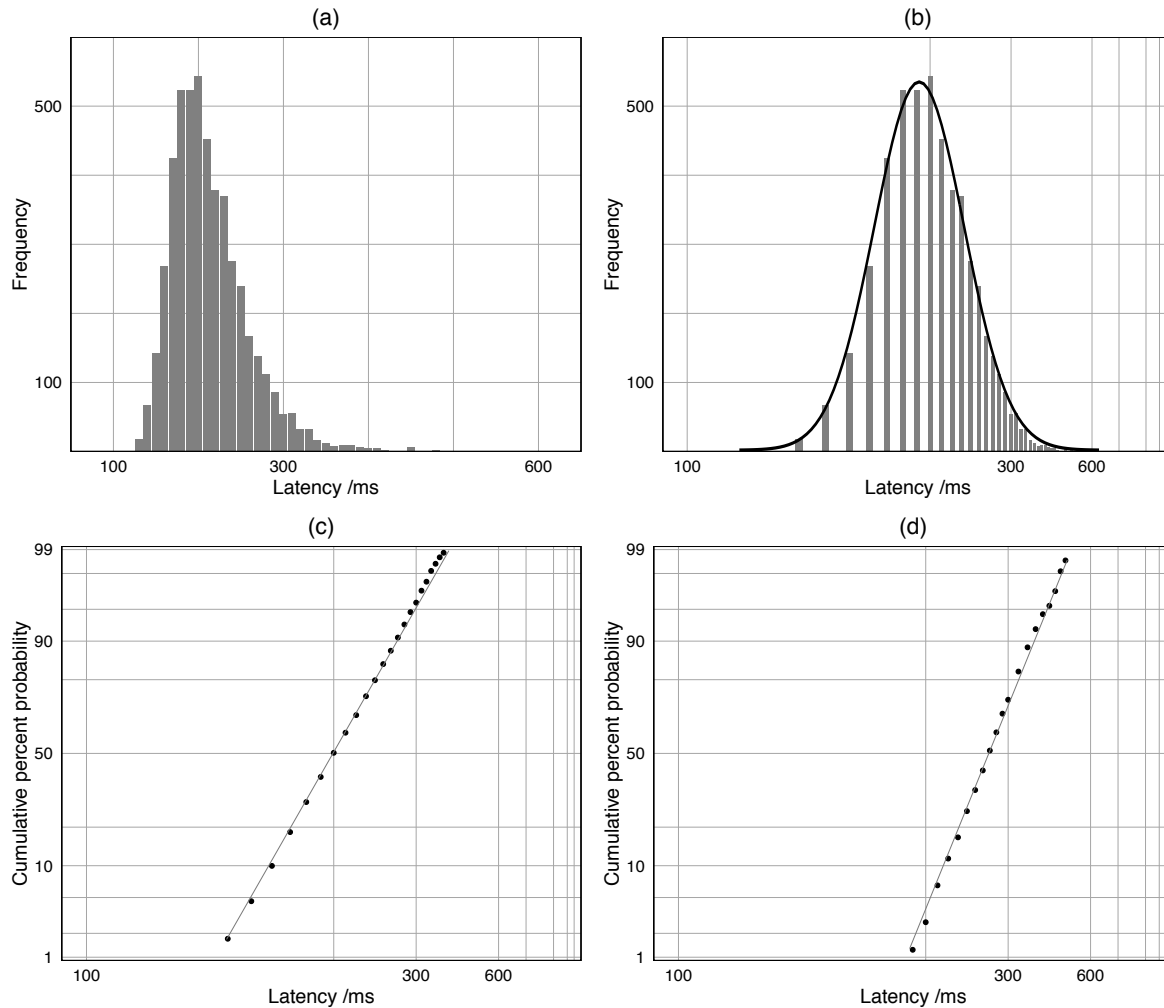


*Figure 17.* (a) probability density function for the angular change in direction between incoming and outgoing saccades for those likely to have been generated by the early unit and main unit. (b) Change in direction between incoming and outgoing saccades for fixations likely to have been associated with the early unit, split by outgoing saccade amplitude. (c) Outgoing saccade amplitudes for early unit saccades, split according to whether the outgoing saccade continued in the same direction as the incoming saccade ( $\pm 22.5$  degrees) or did not continue in the same direction ( $> 22.5$  degrees from incoming saccade direction).



*Figure 18.* The interaction between outgoing saccade amplitude and semantics at the target location for corrective saccades. As the semantics at the peripheral location increases, the negative effect of eccentricity on decision rate increases.

## Appendix

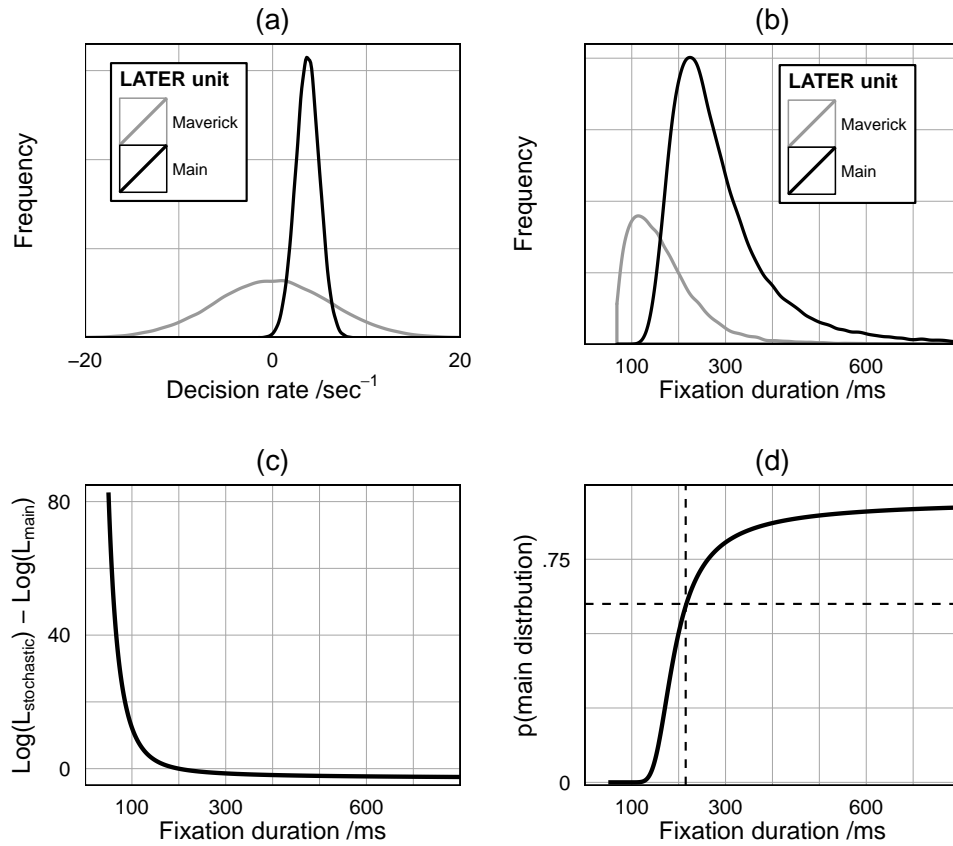


*Figure A1.* A set of 5000 simulated response latencies (rate of rise for decision signal:  $\mu = 5$ ,  $\sigma = 1$ ) plotted as a conventional frequency histogram (a), showing the obvious skewness of the distribution. (b) the same data plotted using a reciprocal scale for latency: note that for convenience the latencies still increase to the right. The distribution is now relatively symmetrical, and indeed similar to a Gaussian. (c) The same data as in (a) plotted as a cumulative histogram, using a probit scale that stretches the ends of the ordinate axis in such a way as to generate a straight line if the data is indeed Gaussian; since the latency uses a reciprocal scale, this is a reciprobbit plot. (d) Human manual responses to a visual stimulus ( $N = 825$ ) (Welford, 1959).



Figure A2. The 64 images used in this study.





*Figure A3.* (a) Simulated decision rates derived from distribution parameter estimates for a single subject in our study. These simulations are based on participant 45 in our dataset, for whom the parameter estimates were:  $\mu = 3.73$ ,  $\sigma = 1.27$ ,  $\sigma_\epsilon = 6.19$ . These simulated decision rates give rise to the distributions of durations shown in (b). The overlap between the distributions of durations arising from the main and stochastic units would result in a single observed population that is the sum of these two distributions. Based on the simulations in (a) and (b) we can calculate the log-likelihood for a fixation of a given duration (c) and use this to calculate the probability that a fixation of any given duration was generated by the main decision process (d). The dashed lines in (d) indicate the 0.6 criterion cutoff we employed to select fixations likely to have been terminated by a saccade generated by the main decision process.

Table A1  
*LATER* parameter estimates for each participant.

Participant	$\mu$	$\sigma$	$\sigma_\epsilon$
1	3.68	1.37	6.14
2	3.89	1.42	7.68
3	3.15	1.11	4.47
4	3.13	1.08	5.17
5	4.04	1.36	7.43
6	4.02	1.34	5.94
7	4.05	1.44	7.16
8	3.90	1.36	5.46
9	3.75	1.35	6.21
10	4.24	1.61	7.07
11	3.92	1.45	5.97
12	3.38	1.37	4.45
13	4.22	1.49	7.07
14	3.75	1.27	6.45
15	4.12	1.50	7.18
16	3.93	1.56	5.58
17	3.87	1.27	5.78
18	4.26	1.55	6.48
19	3.75	1.29	5.15
20	3.77	1.14	3.93
21	3.90	1.53	5.77
22	3.51	1.54	5.05
23	3.95	1.21	5.82
24	3.03	1.10	4.34
25	3.20	1.27	5.06
26	3.84	1.28	5.41
27	3.02	1.25	4.55
28	3.85	1.39	6.19
29	4.53	1.82	7.00
30	3.33	1.42	5.36
31	3.85	1.26	5.41
32	4.01	1.31	5.86
33	4.18	1.69	6.42
34	4.99	1.98	8.70
35	4.22	1.41	5.73
36	3.21	1.09	5.37
37	3.40	1.18	5.15
38	4.69	1.43	6.81
39	4.30	1.44	5.55
40	4.04	1.46	7.62
41	4.21	1.68	6.34
42	3.86	1.29	5.62
43	3.68	1.57	6.22
44	4.32	1.47	6.22
45	3.72	1.27	6.19
46	4.07	1.39	7.60
47	3.95	1.56	5.44
48	3.89	1.28	5.23
49	4.23	1.58	6.57
50	3.69	1.30	5.13
51	3.72	1.38	4.82
52	3.49	1.50	6.81
53	4.14	1.59	6.02
54	2.86	1.17	5.79
55	4.63	1.82	1.70
56	3.25	1.13	5.03
57	3.72	1.31	5.23
58	4.08	1.42	5.49
59	3.62	1.40	5.75
60	4.20	1.39	7.85
61	4.09	1.46	5.63
62	3.85	1.35	5.76
63	3.13	1.20	5.96
64	4.08	1.44	5.52
65	4.19	1.69	6.53
66	4.04	1.60	5.68
67	3.88	1.47	7.00
68	3.54	1.51	5.30
69	4.35	1.38	4.97
70	4.30	1.45	4.55

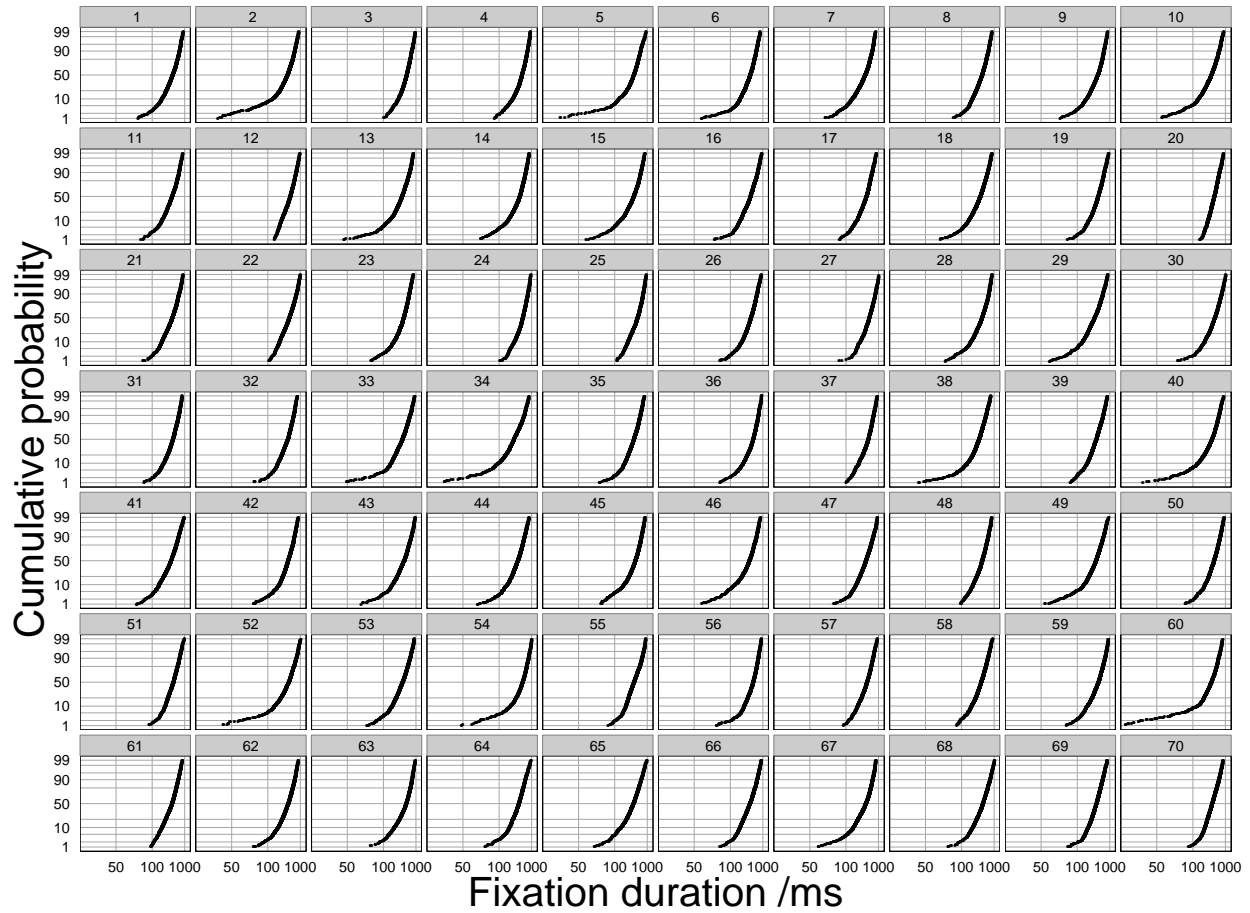


Figure A4. Reciprob plots of the distribution of observed fixation durations for each of our 70 participants.