

---

# DUAL-PROCESS ACCOUNT FOR FIXATION TIMING IN NATURAL SCENE PERCEPTION.

---

A PREPRINT

**R. Calen Walshe**

The University of Texas at Austin

calen.walshe@utexas.edu

**Antje Nuthmann**

The University of Kiel

nuthmann@psychologie.uni-kiel.de

September 17, 2020

## ABSTRACT

During the perception of naturalistic scenes, the eyes frequently move from one location of the visual field to another in the service of tasks such as scene encoding, memorization and scene understanding. It is critical to have visually guided behaviour well adapted to the demands of the current task. By using computational modelling, we study the question of fixation timing during natural scene perception to understand the mechanisms underlying the control of visually guided actions. A stochastic process approach is used to model (a) empirically well characterized mechanisms of the human visual system, (b) hypothetical mechanisms of fixation control, and (c) to predict empirical fixation duration distributions generated from recent experiments. Moreover, the model makes a novel assumption, which is that saccade programming timers are adaptively modified by separate processes associated with (a) fast-acting image-onset surprisal and (b) variation in image encoding difficulty. With this formalization of the dual-process account we are able to recover fixation duration distributions from the fixation-contingent scene quality paradigm, in which the quality of the scene (e.g., its luminance) is manipulated during the duration of selected critical fixations. We suggest that model simulations can provide qualitative and quantitative explanations for seemingly contradictory empirical results.<sup>1</sup>

## 1 Introduction

The study of human perception of natural scenes is a large field with a variety of concurrent goals. Since the human visual system is designed in a way that provides high resolution at the center of gaze with lower resolution in the periphery this necessitates that humans make frequent eye movements around a scene to construct an accurate and high resolution visual percept as well as understand the meaning of objects and elements in the scene. Each time the eye moves to a new location the visual system requires time to process the new information available after landing on

---

<sup>1</sup>Project code and data will be uploaded to [https://github.com/calenwalshe/scene\\_fixation\\_toolbox](https://github.com/calenwalshe/scene_fixation_toolbox)

a new location. The solution that the visual system has arrived at is to allow fixation durations - the amount of time spent looking at a location before moving on to a new location - to vary depending on the specific features of the image in the scene. An open question that we address in this work is: What are the strategies that the human visual system uses to determine the variation in fixation durations during natural scene perception?

## 1.1 Directly controlled fixation durations

A topic of debate is the degree to which eye-movements are directly controlled by the stimulus content upon which gaze is currently directed. With regards to the timing of eye-movement decisions, research aims to determine the degree to which processing of content currently being assessed within a fixation may intervene in influencing the timing of the next saccadic decision (Nuthmann, 2017). Accounts of eye-movement control that adopt strong direct-control assumptions suggest that stimulus processing occurs rapidly enough to have such an influence and that the timing of eye-movement decisions is governed by this processing. Empirical evidence has provided a great deal of support for the direct-control hypothesis. Research on reading behaviour has shown that fixation durations increase when a fixation lands on a word with low predictability or low frequency (Kliegl, Nuthmann & Engbert, 2006) or when the text is presented in visually degraded format (Glaholt, Rayner & Reingold, 2014). In scene perception, direct-control effects on fixation durations have been observed under various conditions such as when a fixation encounters a luminance degraded scene or when spatial frequency content has been removed from the stimuli (Henderson, Nuthmann & Luke, 2013; Walshe & Nuthmann, 2014; Glaholt, Rayner & Reingold, 2013; Henderson, Olejarczyk, Luke & Schmidt, 2014). Such effects have also been observed for fixations that land on a task relevant vs. irrelevant stimuli (Glaholt & Reingold, 2012) as well as fixations in which a task irrelevant distractor stimulus is presented (Pannasch, Schulz & Velichkovsky, 2011). Overall, compelling evidence exists to support the claim that some proportion of fixations are directly controlled by stimulus content.

Despite such evidence for direct control, research has demonstrated that not all fixations show such an effect. Fixation durations that do not appear to be influenced by currently fixated stimulus properties are often referred to as being indirectly controlled. Indirect control postulates that the duration of an individual fixation is determined by factors that are external to the local stimulus content encountered within a fixation. Instead, fixations may be executed at random intervals with a mean rate set by global factors external to the fixated stimulus (Reingold, Reichle, Glaholt & Sheridan, 2012). In a study of direction-coded search, Hooe & Erkelens (1998) argued that performance failures could be explained by an inability for participants to match fixation durations to the processing demands of the fixated stimulus. Such a result violates the assumptions of a pure direct control eye-guidance mechanism. Henderson & Pierce (2008) used a technique known as the stimulus onset delay paradigm (SOD) (Shioiri, 1993) to investigate the control of fixation durations during naturalistic scene perception. In this paradigm participants actively explored the scene and on a selected saccade the entire scene was removed from view. After a variable delay within the subsequent fixation, the scene was replaced to view. Consistent with a direct control account, they found that one population of fixation durations increased to match the length of the delay. However, they found that another population of fixation durations was unaffected by the length of the SOD. On this basis, the authors argued for a mixed-control account with one population being directly controlled and the other being indirectly controlled.

## 1.2 Challenges for direct-control theories

A serious challenge that direct-control theories must face is how such effects are at all possible given the limited processing time that is available for stimulus processing to influence eye-movements (Reichle & Reingold, 2013). In a typical naturalistic viewing task, fixation durations are approximately 300 ms with a wide variability around this mean (Rayner, 2009). However, direct-control effects have substantially less time to exert their influence. This becomes apparent when the timing of specific neural events related to stimulus processing are taken into account (Reichle & Reingold, 2013). It is known that approximately 60 ms is required to transmit visual signals represented on the retina to regions of the brain responsible for processing these signals (Reichle & Reingold, 2013). Furthermore, an additional 90 ms is required to encode these signals and process them into a format in which they may be interpreted. The latency of the movement signal to pass from the brain to eye is thought to be approximately 20 ms (Becker, 1991). Given these timing constraints, any direct control effect of stimulus content on the timing of the resulting saccade must logically occur between the time needed for the stimulus to reach brain areas responsible for saccadic decisions and onset of transmission of the oculomotor movement signals. Any theoretical approach to understanding the underlying mechanisms must give consideration to such timing constraints.

## 1.3 Eye-movement control models in high-level tasks

Eye-movement control theories have been most actively developed within the study of reading behaviours (Engbert, Nuthmann, Richter & Kliegl, 2005; Reichle, Pollatsek, Fisher & Rayner, 1998; Legge, Klitz & Tjan, 1997; McDonald, Carpenter & Shillcock, 2005). Reading presents fertile ground for testing such theories as movement trajectories follow highly stereotyped patterns and the stimulus varies in predictable and easily measurable ways. However, recent models of the temporal properties of eye-movement decisions have been successfully extended beyond the reading domain (Nuthmann, Smith, Engbert & Henderson, 2010; Trukenbrod & Engbert, 2014; Reichle, Pollatsek & Rayner, 2012; Tatler, Brockmole & Carpenter, 2017).

These models have had success in part because they bridged the gap between low and high-level factors that influence eye movement timing. Low-level factors such as saccade programming timelines are largely known via clever experiments using the double-step paradigm. Evidence from double-step studies (Becker & Jürgens, 1979; Westheimer, 1954) shows (a) that task relevant decision making does occur within a single fixation and (b) that the timing and stages of decision making within an eye-movement can be accurately measured. In a double-step task, two stimuli are placed into conflict with one another by placing one stimulus at a temporal and positional offset from the other. A typical goal in a double-step task is to maintain a fixation until a target is presented on the screen and then make a saccade (fixate) the target as quickly as possible. A competing goal is introduced by requiring that, if a secondary stimulus appears after the first stimulus, a fixation is made to the new stimulus instead. Typically, the updated goal stimulus is located at a new position and the subject has to reformulate their decision. Critically, the double-step task is calibrated to ensure that the new stimulus is presented before an eye-movement to the first stimulus has been generated. In other words, two sources of information have been presented to the observer at a temporal offset before any response is made. By varying the interval between the first and second targets it is possible to infer how well the subject can switch between the goals as a function of time. A classic finding from this literature is that saccade motor planning can be updated after being initiated up until approximately 80 ms prior to the actual physical onset of the

saccade. More recently, Walshe & Nuthmann (2015) studied saccade plan modification in a naturalistic scene viewing task and found that the classic finding generalizes to the more ecologically valid context. They also showed evidence that saccade modification deadlines may be longer in scenes viewing than commonly reported in the literature.

### 1.3.1 CRISP model

The CRISP model was developed as a model of fixation durations in scene viewing (Nuthmann et al., 2010), but has also been applied to modelling fixation durations in reading tasks (Nuthmann & Henderson, 2012). Saccade programs are initiated at a rate determined by an autonomous timing process (Engbert, Longtin & Kliegl, 2002; Engbert et al., 2005). In the CRISP model, the timer was implemented as a stochastic random walk towards a threshold. Once the threshold value is reached, a saccade program is initiated that may then, eventually, terminate the fixation. The CRISP model may be referred to as an example of a mixed-control mechanism (c.f. direct control) due to the manner in which within-fixation stimulus influences impact the timing signal. The direct control component of the CRISP model arises from two sources. The first is due to inhibitory influences that may interfere with the rate at which the timing signal approaches threshold. For example, if a complex or difficult to process stimulus is fixated, the timer rate may be inhibited. The second direct control influence in the CRISP model arises when stimulus processing elicits a cancellation of a saccade that is within the initial stages of preparation. On the other hand, indirect control occurs in the CRISP model when saccade programming is initiated prior to the onset of the current fixation and when no processing related cancellation occurs. In these cases, since a saccade program has been initiated prior to the onset of the current fixation, modification of the timer rate will not have any impact on the current fixation duration. Nuthmann et al. (2010) showed that advanced preparation saccades tend to be those that produce short fixation durations. Therefore, given this fact the CRISP model implicitly predicts that short fixations tend to avoid the influence of direct control.

### 1.3.2 ICAT model

ICAT is a model of eye-movement control in high-level tasks (Trukenbrod & Engbert, 2014). As in the CRISP model, saccade programs are initiated according to a random timing process that may be subject to processing related interference. Once the timer reaches threshold, a labile program is initiated. However, unlike the CRISP model, ICAT implements the labile and nonlabile stages of saccade programming as independent random walk processes. In ICAT, the duration required for saccade programming is equal to the sum of the durations of each independent random walk. A novel aspect of ICAT is that it permits interference of saccade programs at the level of the labile stage of programming. When foveal processing difficulty increases, the timer rate of the labile stage of programming is decreased. This is in contrast to the CRISP model which does not permit the modification of labile durations in response to task demands.

ICAT also makes more elaborate assumptions about the relationship between stimulus processing and the duration of saccade timing and saccade programming processes. This elaboration is accomplished by introduction of local<sup>2</sup> vs global control principles. Changes to the mean rate of the saccade timer that occur at a global level are not tied to the immediate processing context (i.e. foveated content). Rather, these global effects are sensitive to properties of the

---

<sup>2</sup>Within ICAT this is technically referred to as Local-II. Local-I is a term that refers to variability in mean duration that results from the stochasticity of the random walk. Our focus is on local-II as we are interested in changes that occur to the rate of the timing process

stimulus encountered on previous fixations and reflect general aspects of the difficulty of the task. Local control is the mechanism implemented in the ICAT model to control changes in the rate at which the saccade timer and labile stage of programming adapts to currently foveated content<sup>3</sup>. Thus, ICAT’s local control mechanisms play a similar role to the direct control mechanisms implemented in the CRISP model. ICAT’s local control process is responsible for adaptive changes to timer rate that occur on the level of a single fixation.

The CRISP model (Nuthmann et al., 2010; Nuthmann & Henderson, 2012), the ICAT model (Trukenbrod & Engbert, 2014), and the SWIFT model (Engbert et al., 2005; Schad & Engbert, 2012) form a class as they all share the assumption of a random saccade timer. Moreover, like the E-Z Reader model (Reichle et al., 1998; Reichle, Rayner & Pollatsek, 2003; Reichle et al., 2012) they all assume that saccades are programmed in multiple stages (Becker & Jürgens, 1979).

### 1.3.3 Extensions of the LATER model to high-level tasks

A different class of models builds upon the LATER model, which was developed to study decision making mechanisms via measurements of eye-movements (Carpenter, 1981; Carpenter & Williams, 1995). In the LATER model, a decision signal rises linearly at a certain rate in response to the stimulus until it reaches a threshold or criterion level, at which point a saccade is initiated. The variability in saccade latencies originates from variability in the time taken to reach the threshold. The base model successfully reproduces distributions of latencies associated with “evoked” saccades; that is, the time taken to make a saccadic eye movement to look at a sudden visual target (Noorani & Carpenter, 2016). Importantly, LATER was extended to model “spontaneous” saccades in dynamic high-level tasks. Specifically, LATERian decision making has been incorporated into the SERIF model of eye-movement control in reading (McDonald et al., 2005) and in the LATEST model, which utilizes a single decision mechanism to explain both when and where observers look in scenes (Tatler et al., 2017). In the case of reading and scene viewing, the modelled latencies translate to intersaccadic intervals, which are equivalent to fixation durations in the visual-cognition literature. In LATEST, fixation durations are assumed to reflect decision time alone with no additional processes (such as saccade programming) contributing to their duration (Tatler et al., 2017). This implies that saccades are planned one by one. Due to the omission of saccade programming, neither the cancellation of saccade programs nor parallel programming of saccades is possible. To account for the subpopulation of very short fixation durations that are often observed during scene viewing, the main decision unit acts in parallel with a “maverick” unit (Roos, Calandrini & Carpenter, 2008). Only the main decision process makes full use of information from the stimulus. However, the maverick saccade generator will occasionally rise to threshold faster than the main decision unit, generating short fixation durations (Roos et al., 2008; Tatler et al., 2017).

## 1.4 Asymmetrical control of fixation durations

Research from reading and visual search tasks has led to the suggestion that there is an asymmetry in how eye-movement control mechanisms respond to increases and decreases in foveal processing difficulty. In reading, Kennison & Clifton (1995) investigated the impact of word frequency on two adjacent words embedded in single sentences. High and low word frequency adjectives were followed by high and low word frequency nouns. Parafoveal preview of the

<sup>3</sup>It should also be noted that global and local control interact. Global control mechanisms scale local inhibition depending on the anticipated processing difficulty.

noun was prevented by using the invisible boundary technique. When readers first fixated a high-frequency adjective, fixation durations on the subsequent noun showed a word frequency effect, such that longer fixation durations were observed for low-frequency than for high-frequency nouns. In contrast, no such word frequency effect was observed when readers first fixated a low-frequency adjective. Thus, increasing processing demands (high  $\rightarrow$  low) resulted in an immediate prolongation of fixation durations, whereas decreasing processing demands (low  $\rightarrow$  high) showed no immediate facilitatory effect. Such an asymmetry in the temporal control of fixation durations has also been observed in visual search. Hooge, Vlaskamp & Over (2007) used a search task in which participants were required to find a closed ring amongst distractor Cs. The distractors in their task varied in the size of the gap, such that small gap Cs were more difficult to distinguish from the target stimulus than were large gap Cs. They found that fixations on small gap Cs that were preceded by a fixation on a large gap C showed increased durations. However, a fixation on a large gap C following a fixation on a small gap did not show a corresponding decrease in fixation durations. Walshe & Nuthmann (2014) used a task in which scene quality (i.e. luminance) was gaze-contingently modified during the saccade preceding a critical fixation (McConkie & Loschky, 2002; Henderson et al., 2013). The base luminance level was 80% in Experiment 1, and 60% in Experiment 2. The quality of the scene was either increased (luminance increase from 80% to 100% in Experiment 1, Figure 1a, and from 60% to 100% in Experiment 2) or decreased (luminance decrease from 80% to 60%, or from 60% to 20%). Degrading the scene stimulus by shifting luminance down resulted in an immediate increase to fixation durations. Interestingly, enhancing the scene stimulus by shifting luminance upwards did not result in a comparable decrease to fixation durations. Instead, fixation durations were significantly increased (Figure 1a). However, the increase was much smaller in the UP condition than in the DOWN condition of a given experiment. Taken together, the results are suggestive of an asymmetric pattern of fixation durations in visual-cognitive tasks, due to the asymmetric nature of the underlying mechanisms used to control fixation timing.

Recently, several studies have been conducted that suggest an alternative to strictly asymmetrically controlled fixation durations. Glaholt et al. (2013) showed direct control effects on fixation durations that depended on whether the image had been low- or high-pass filtered. In their main experiment, during a selected critical fixation the (grey-scale) scene was changed to a high-pass or low-pass spatial frequency filtered version.<sup>4</sup> Under both conditions, fixation durations increased, with both filter conditions resulting in a general shift in the mode of the distributions that occurs for even very short fixation durations. However, it was also found that low-pass filtering produced a larger effect on fixation durations than did high-pass filtering. Using distributional analysis the authors showed that the difference between the two conditions arises primarily from an increase in long fixation durations (longer tail) but not from a difference in mode between the two distributions. In their main experiment, Glaholt et al. (2013) *decreased* the quality of the scene by *removing* spatial-frequency information. In an additional control experiment, however, they *increased* the quality of the scene by *adding* colour to grey-scale scenes (Figure 1b). The mean fixation duration was significantly longer in the colour condition than in the no-change condition, and this was related to a shift in the mode rather than the tail of the distribution.

On the basis of these results, the authors presented a dual-process account to explain the source of these two distinct distributional effects. Specifically, Glaholt et al. (2013) suggested that rapid influences on the *mode* of the distribution are the result of a *surprise* effect. The surprise effect is fast acting, modifies even the shortest fixation durations,

---

<sup>4</sup>The change occurred during the saccade prior to the fixation while processing of new visual information is suppressed (Ross, Morrone, Goldberg & Burr, 2001).

and results from a mismatch in pre- and post-saccadic stimulus content. Changes in the tail of the distribution were observed only for experimental conditions in which the change resulted in an increase in processing difficulty. On this basis, the authors argued that increases that occur specifically on the *tail* of the distribution result from *processing related* influences on fixation durations. A similar account was suggested by Walshe & Nuthmann (2014) to account for the distributional effects observed in their study with luminance stimuli. Consistent with Glaholt et al. (2013)’s dual-process account, they found that while fixation durations tended to increase even when the scene was made easier to process, this increase came from a general shift in the mode of the distribution, and not from any influence on the tail.

A question that arises naturally from the dual-process account is whether it is possible for an increase in scene quality to be substantial enough to overcome any surprise effect that results from the detected mismatch in scene features. This question was addressed by Henderson et al. (2014) in a saccade-contingent scene-quality change study on how spatial-frequency changes impact fixation durations. Participants viewed scenes which had been reduced in quality by low-pass filtering the scene. On every 6<sup>th</sup> saccade the reduced quality image was replaced by one of four clarified (i.e. less strongly filtered) stimuli or the original unfiltered scene. What they found in the unfiltered scene condition was that fixation durations *decreased* (90 ms) relative to fixation durations on the baseline stimuli (Figure 1c). In another experiment, the authors presented participants with base stimuli that consisted of fully unfiltered scenes. In this study, spatial frequencies were *removed* rather than *added*. They observed that fixation durations *increased* when a filtered stimulus was presented. Such a pattern of results is contrary to what would be predicted from a purely asymmetrical account. For this reason, the results of Henderson et al. (2014) and Walshe & Nuthmann (2014) are partially at odds. Using a similar experimental design, Henderson et al. (2014) observed symmetry while Walshe & Nuthmann (2014) observed asymmetry. However, the dual-processing account suggests a possible explanation. Specifically, if the surprise induced by the scene change was overcome by late encoding related facilitation, then a decrease in fixations may be observed. Unfortunately, Henderson et al. (2014) did not present distributional analyses that are required to confirm this prediction from the dual-process account.

## 1.5 A dual-process model of fixation-duration control during scene viewing

In previous work, the dual-process account was tested by fitting ex-Gaussian distributions to fixation-duration data (Glaholt et al., 2013; Walshe & Nuthmann, 2014). With the present work, we go beyond that by formulating a simulation-based model that merges the dual-process account with recent models of eye-movement control in high-level tasks. Specifically, we borrow principles from both the CRISP model (Nuthmann et al., 2010) and from ICAT (Trukenbrod & Engbert, 2014) and supplement these with principles derived implicitly from the dual-process account. Thus, the computational model is used to make these implicit dual-process assumption explicit by simulating distributions of fixation durations and comparing them against empirical observations.

In brief, the model is based on a stochastic simulation approach that uses sequences of randomly generated saccade timing signals to initiate the programming of saccadic eye movements (Engbert et al., 2002, 2005; Nuthmann et al., 2010; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014). In the model, both saccade timing and saccade programming are modelled as stochastic random walk processes (Schad & Engbert, 2012; Trukenbrod & Engbert, 2014). This means that the durations required to initiate and program a saccade are partially determined by the inherent

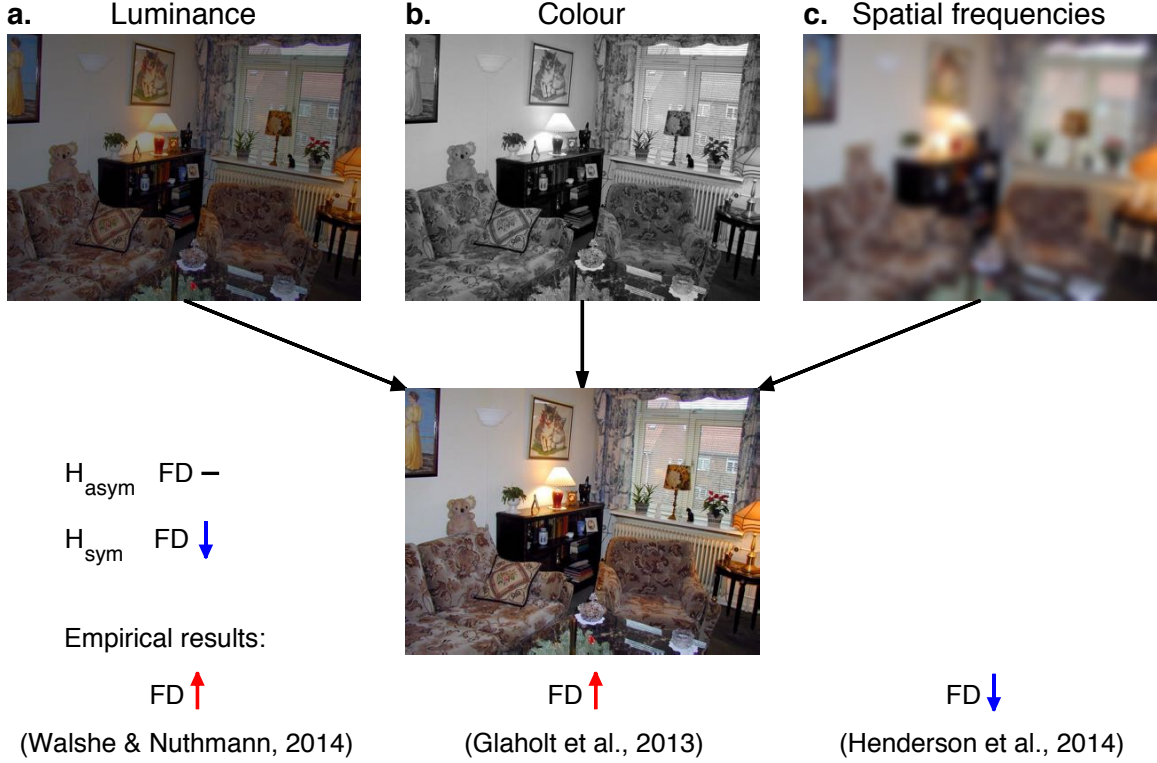


Figure 1: Scene enhancement manipulations. The base scenes were a. luminance reduced scenes (Walshe & Nuthmann, 2014, Exp. 1), b. grey-scale scenes (Glaholt et al., 2013, Exp. 2b), or c. low-pass filtered (i.e., blurred) scenes (Henderson et al., 2014, Exp. 2). On critical fixations, the quality of the scene was increased by a. increasing its luminance, b. adding colour, or c. adding high-spatial frequencies. Asymmetric control predicts no change in the duration of critical fixations ( $H_{asym}$ ), whereas symmetric control predicts a decrease in fixation duration ( $H_{sym}$ ). The pattern of empirical results is schematically depicted at the bottom of the figure. FD = fixation duration.

unpredictability of the random walk. The decision to move the eye from the current fixation is made once saccade programming is completed. The dual-process account predicts two methods by which the mechanisms responsible for gaze shifts may be modified on a moment-to-moment basis (i.e. surprise and encoding modulation). Thus, the dual-process assumptions enter the model at the level of saccade programming. That is, both surprise and encoding related signals can inhibit the rate at which saccade programs are completed. We also leave open the possibility for encoding related facilitation of saccade programming (Henderson et al., 2014). In what follows, we introduce model mechanisms both in an informal and in a formal manner. We then use the model to simulate data reported in Walshe & Nuthmann (2014).



## 2 Methods

### 2.1 Model

Here, we present a summary of important model principles that contribute to generating human-like fixation durations. Additional implementation details, best fit model parameters (Table 1) and fitting methods are provided in the appendix. Figure 2b shows a model schematic.

#### 2.1.1 Rhythmic saccade timer

Rhythmic timing is a core principle used in our modelling approach and is a shared principle with other models of eye-movement control (Engbert et al., 2005; Nuthmann et al., 2010; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014). These models all share the basic assumption that the mechanisms underlying eye-movement decisions rely on processes that generate behaviourally relevant signals at random intervals defined by a mean rate of completion. The timing signal is continuously active which allows for the spontaneous generation of saccades that are not tied to any particular stimulus or goal.

#### 2.1.2 Saccade programming

**Labile vs nonlabile stages** Saccade programming is completed in multiple distinct stages of processing. Each stage of saccade programming is implemented as an independent random walk towards threshold. This multi-stage saccade programming assumption is derived from empirical investigation into saccade programming in both simple (Becker & Jürgens, 1979; Ludwig, Mildinhal & Gilchrist, 2007) and high-level tasks (Walshe & Nuthmann, 2013, 2015). In the reading and scene viewing literature the distinction between these two stages of saccade programming has been referred to as the labile and non-labile stages of saccade preparation. We adopt the standard nomenclature for these stages. Saccade programs that are within the labile stage are subject to cancellation while those in the nonlabile stage are not. Saccade cancellation occurs when the saccade timer initiates a new labile saccade program while a previous labile saccade program is currently active. In this case a new labile program is activated that replaces the ongoing labile program (see Nuthmann et al., 2010, for explorations using simulations). An additional random walk process is used model the latency required to instruct the oculomotor muscles to physically adjust the position of the eyes. We refer to this random walk as "motor".

**Concurrent saccade preparation** It is known that saccades may be programmed in parallel, a result which can be inferred from studies using double-step targets (Becker & Jürgens, 1979; McPeck, Skavenski & Nakayama, 2000; Caspi, Beutter & Eckstein, 2004) and neurophysiological data (Shen & Paré, 2014). In double-step tasks, saccades that have not yet passed the point-of-no-return often show landing positions that are between the old and new target positions. When a saccade lands between the two target positions, a corrective saccade is typically made to the final target position. However, concurrency is revealed in the reduction in the time that is required to program this corrective saccade. This inter-saccadic interval is typically well below the duration of an ordinary saccade and decreases linearly with the amount of time that is available to process the corrective saccade with the initial saccade (McPeck et al., 2000; Camalier, Gotler, Murthy, Thompson, Logan, Palmeri & Schall, 2007). Parallel programming of saccades is not restricted to simple saccade-targeting tasks, but extends to the free-viewing of (a) configurations of real-world objects

(Wu, Chua & Yen, 2016), and (b) images of naturalistic scenes (Wu, Gilani, van Boxtel, Amihai, Chua & Yen, 2013). In the modelling approach used here, parallel programming occurs when the saccade timer initiates a labile saccade program while a nonlabile program is currently activated (Nuthmann et al., 2010).

### 2.1.3 Saccade programming rate modulation

In the model, influences of direct control occur when the mean rate of the random walk associated with saccade programming are modified. Rate modulation may occur for both the autonomous timer, labile and nonlabile accumulators. The dual-process account suggests that there are two distinct direct control influences on the timing of fixations. The first influence has been referred to as a surprise effect (Glaholt et al., 2013; Walshe & Nuthmann, 2014). This effect occurs when the eye lands on a location which contains visual features that strongly depart from what was expected prior to the onset of the eye-movement (Glaholt et al., 2013). Due to its rapid onset, surprise can influence even very short fixation durations. The surprise effect is purely inhibitory and results in a slow-down in the rate of both the labile and nonlabile programs that occurs at the minimal delay in transmission of signals from retina to brain (50 ms) (Reichle & Reingold, 2013).

## 3 Results

### 3.1 Heavy tailed fixation duration distributions

The model was tested on data from the two saccade-contingent luminance-change experiments reported by Walshe & Nuthmann (2014). Images of real-world scenes were initially presented at a baseline luminance of 80% in Experiment 1, and 60% in Experiment 2. During saccades preceding selected critical fixations, the scene stimulus was replaced with a scene which had the luminance increased or decreased by a margin of 20% in Experiment 1, and 40% in Experiment 2. Overall, the distributions produced by human observers were heavy tailed with moderate to high skewness. In the simulations, we found that the heavy-tailed property arises primarily from the fact that the aggregate fixation duration distribution is comprised of a mixture of approximately Gaussian distributions. This is shown in Figure 2a. For a given experiment and condition, the red line shows the empirical human distribution and the histogram shows the proportion of each fixation duration produced by the best fitting model. Each distribution is coloured according to the number of cancellations that occurred within that fixation. Each fixation was labeled by the total number of cancellations that occurred within the fixation. Fixations were grouped by total number of cancellations and sorted into 20 bins representing the duration of the fixation. The result was then normalized by the total number of fixations. Cancellations beyond 4 were not included as they occurred rarely. Figure 3a highlights the role that cancelled saccades play in generating long fixation durations. Most long fixation durations arise from saccade programs with multiple cancellations. The simulated distributions demonstrate another important model phenomenon; that the long tailed property of the fixation duration distribution arises as a consequence of aggregating fixations with variable numbers of underlying cancellations.

### 3.2 The influence of saccade preprogramming on fixation durations

Advanced preparation of saccades is known to play a role in the timecourse of saccadic preparation (Camalier et al., 2007; Ludwig et al., 2007; McPeck et al., 2000; Nuthmann et al., 2010; Saez de Urabain, Nuthmann, Johnson &

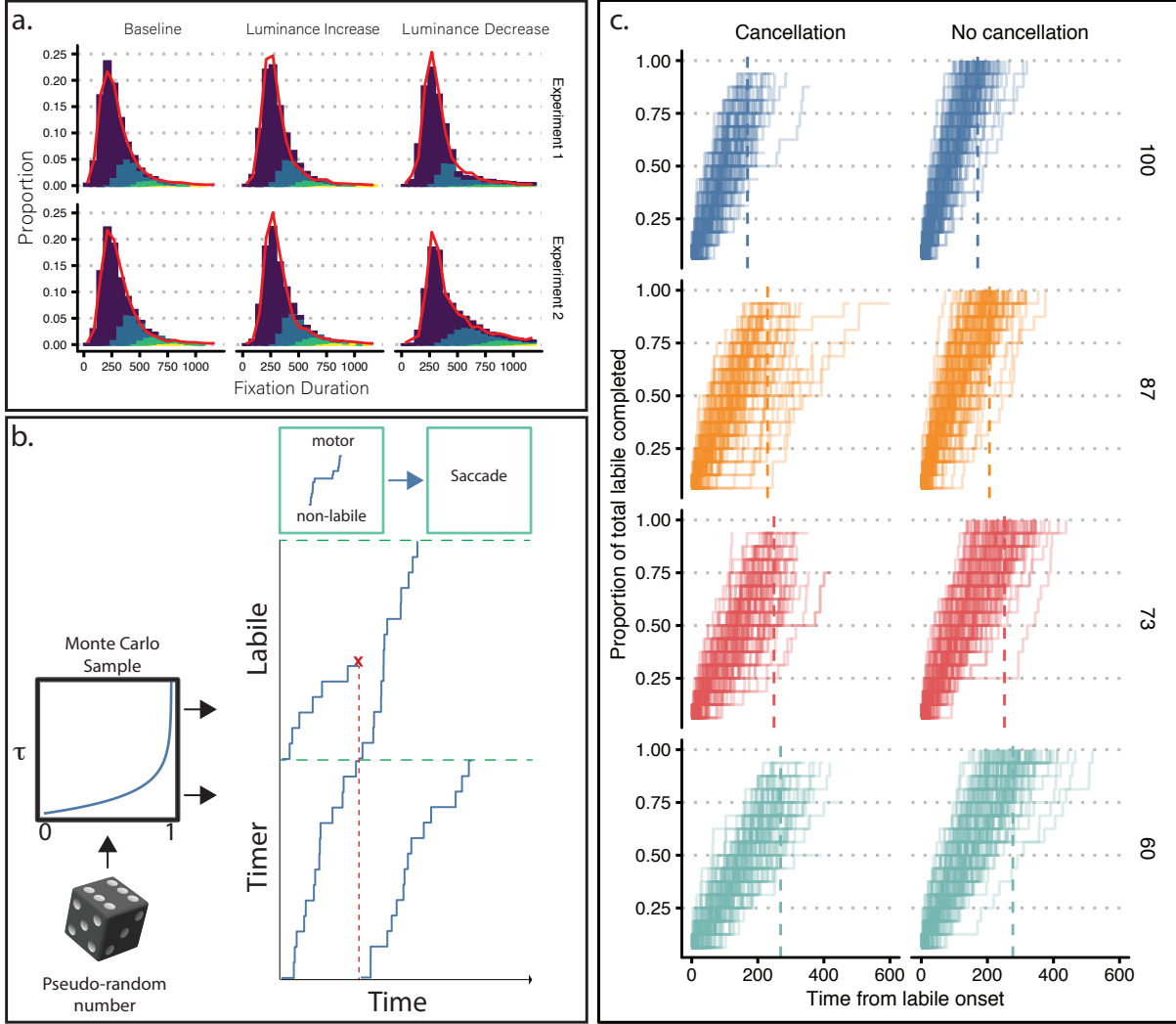


Figure 2: a. Model predictions (stacked bars) and human fixation distributions (red lines). The model distributions change in systematic ways to match human fixation distributions resulting from luminance changes. b. Model schematic. c. Example labile random walks from Experiment 1. The x-axis refers to the length of time since the onset of the labile stage of saccade programming. The columns compare sample random walks for non-cancelled and cancelled random walks. The rows show the effect that timer adjustment has on the duration of the random walk. Each row contains random walks whose rates have been slowed compared to baseline (100). The vertical dashed line shows the mean.

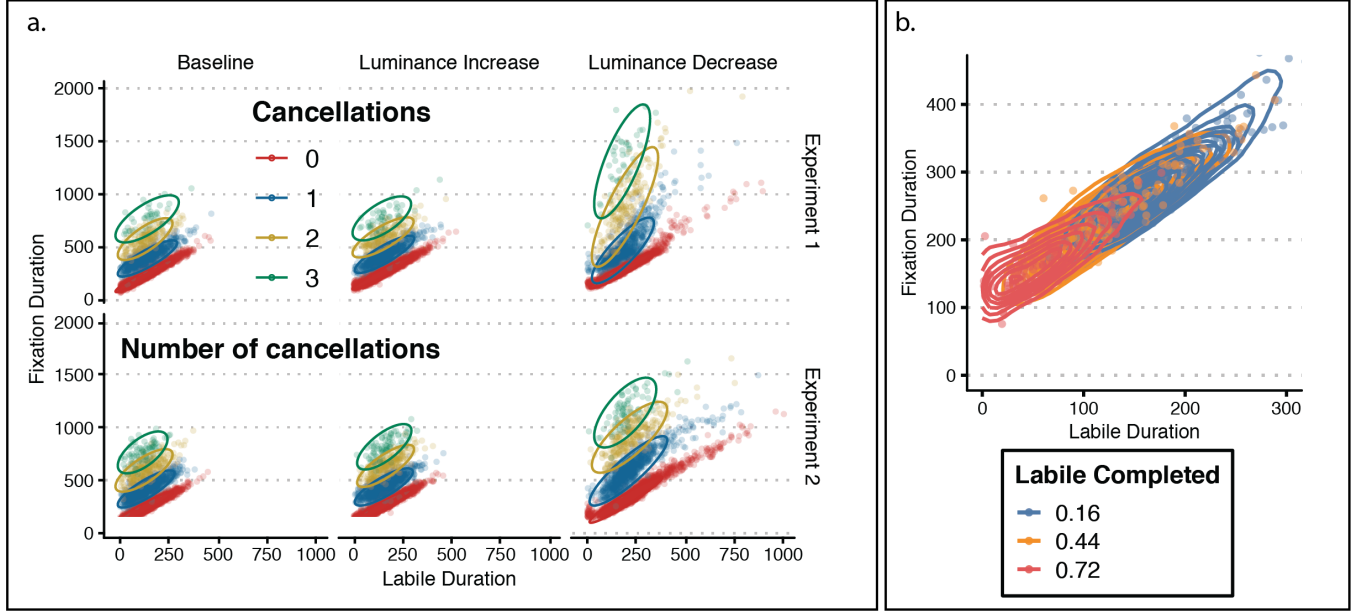


Figure 3: a. Variability in the duration of labile saccade programs. The x-axis refers to the amount of time that saccade programming spends in the labile stage. A linear relationship holds between the duration of labile programming and the eventual fixation duration. Fixation durations are grouped by number of fixations (colours). Increased number of cancellations results in longer fixation durations and the relationship between labile programming and fixation durations holds for each cancellation group. The difference in labile duration observed between luminance conditions is a direct consequence of the modification to the rate of the timer. Due to linearity, fixation durations increase proportionally to the increased labile duration. b. Saccade preprogramming. Fixation durations are labelled according to the degree of labile preprogramming. When a critical fixation begins, a labile program may already be activated. Colour indicates the level of preprogramming. Red fixation durations are those in which the labile program was near to completion when the luminance change begins. Preprogramming is associated with short fixation durations.

Smith, 2017). We did not specifically design our simulations to incorporate preprogramming of eye-movements. However, we found that the model produces overlapping saccade plans which can naturally be interpreted as saccadic preprogramming and is consistent with psychophysical results from double-step tasks. The logic of the simulations assumes that saccade programming occurs after a random timer initiates a saccade program. The form of saccade preprogramming in the simulations arises when a labile or non-labile saccade program is already active at the beginning of a fixation. In other words, progress has already been made on preparing the next eye-movement before the start of the current fixation. We explored this behaviour by looking at the proportion of labile saccade programming that had been completed at the start of a critical fixation. Figure 3b shows critical fixations in which the preparation of the labile program had already been partially completed at the start of the critical fixation. The three colours shows labile programs that had completed 16%, 44% and 72% of the maximum duration at the onset of the fixation. A greater proportion of labile stage preprogramming results in shorter fixation mean fixation durations. Within each of these groups, labile programming duration is linearly related to an increase in fixation duration. The results support the idea that fixation timing in natural environments is partially determined by concurrency of saccade preparation (Nuthmann et al., 2010; Saez de Urabain et al., 2017; Trukenbrod & Engbert, 2014).

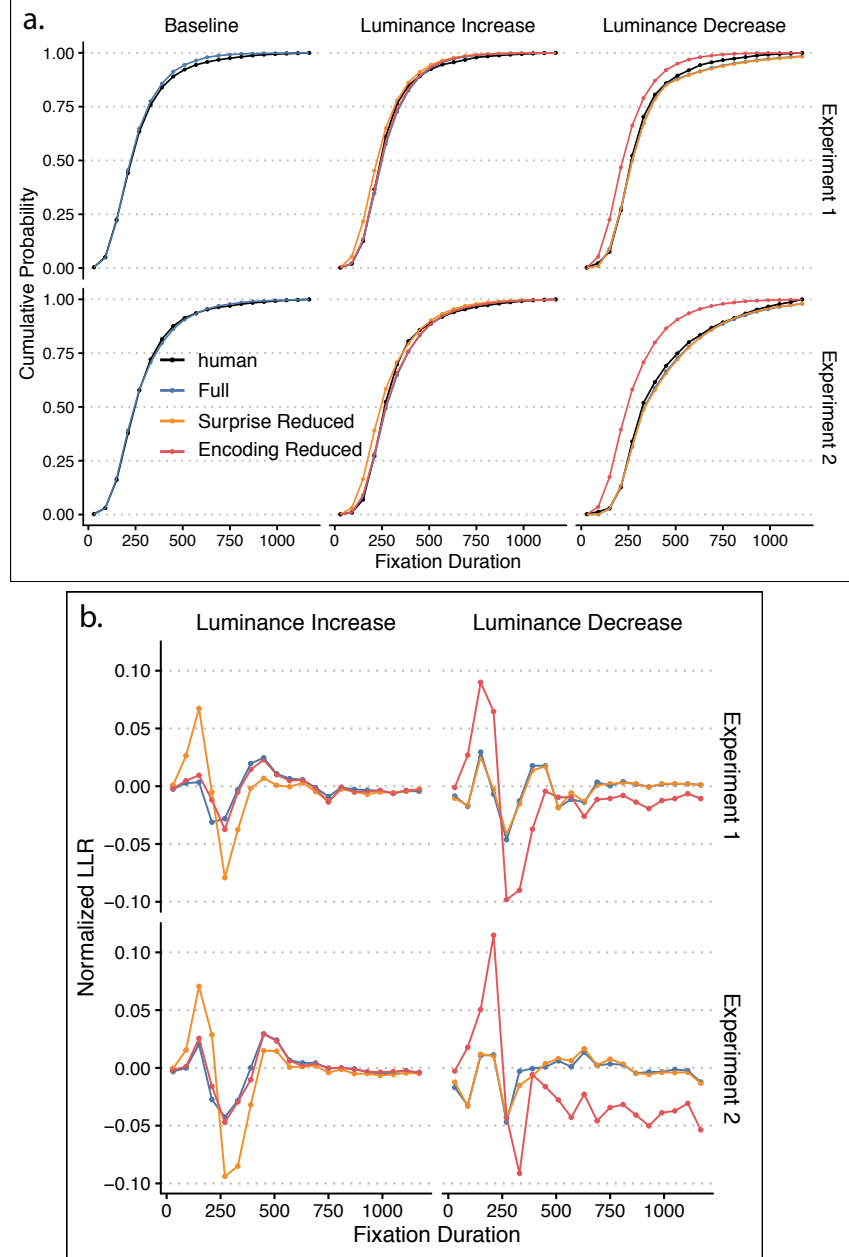


Figure 4: a. Cumulative distributions. Black is the human cumulative distribution, colours are models. The tight match between the Full (full model) and human cumulative distributions demonstrates the rate adjustments successfully characterize how fixation durations adapt to luminance changes. Removing the rate modulations has specific and predictable effects on the distribution (red and orange curves). Surprise adjustment has a short duration and its impact is therefore localized to short fixation durations. Encoding rate adjustment has a long duration and the removal has an impact on fixation durations of all lengths, relative to human densities. b. log-likelihood ratio normalized by the probability of the human fixation duration. A flat line would be a theoretically perfect fit. Blue approximates the theoretical fit. Removing surprise and encoding modulation separately shows their effective contribution to the full model. Removing surprise results in a greater proportion of short fixation durations, whereas leaving surprise and removing encoding results in a generalized impact on fixation durations of all lengths. The asymmetry implies that both mechanisms contribute to the characteristics of human fixation densities.

### 3.3 Counterfactual analysis of adaptive saccade programming rates

The main goal of the model simulations was to test the hypothesis that the empirical distributions of fixation durations observed in the luminance task (Walshe & Nuthmann, 2014), and in other tasks used by other laboratories (Glaholt et al., 2013; Henderson et al., 2014), is explained by a conceptually simple mechanisms of adaptable adjustments to the saccade programming rates. We found that the timing signals interact with a saccade cancellation process to produce either more or less skewed distributions - depending on how many cancellations occurred. This partially explains the heavy tailed property of fixation duration distributions according to our model.

The rate at which the timing mechanisms accumulate to the boundary is tied to the luminance level in the scene. For example, a large decrease in luminance resulted in a large negative influence on the saccade programming rates. The parameters for the accumulation were free to vary, but we assumed that the rates could only be inhibited, not sped up. We found that the rate adjustment mechanisms were sufficient to explain the pattern of results observed in the luminance adjustment conditions. The two-component (dual-process) rate-adjustment model operated as follows: first, a surprise inhibition to the rate was activated shortly after the onset of the luminance change. The duration of the surprise inhibition had a short duration. A second inhibition process, labeled "encoding", was activated later in the fixation; that is, after the surprise inhibition had terminated. The values for the onset of the surprise and encoding related inhibition were left as free parameters in the simulations. Figure 2a compares the empirical proportion against the model predictions. The model fits are obtained by first fitting parameters to a baseline condition and then holding all parameters fixed, while allowing the rate adjustment parameters to vary to best fit the empirical data. Figure 4a shows the cumulative distribution for the fitted and empirical distributions (fitted model in blue and empirical data in black) for all conditions in the two luminance experiments. Our result indicates that assuming these two rate adjustment mechanisms provides a good fit to the human data. A substantially worse fit is obtained by omitting these model components, or by including only one component.

To determine how the rate adjustments impact the distributions, we employed a counterfactual analysis approach. The counterfactual analysis allows us to inspect the causal role that the timing adjustment plays in shaping the fixation duration distribution. A counterfactual distribution is defined as the fixation distribution that arises by setting a subset of parameters to a new value and holding all others at their original value. To obtain the counterfactual distributions we take the best fit distributions for each luminance condition and set one of the rate parameters to 1. This effectively removes the influence of the rate adjustment. The counterfactual distributions are acquired for each luminance condition and each rate adjustment type (surprise and encoding) separately by applying the same procedure to each rate parameter (setting the parameter to unity). The counterfactual distributions are then compared to the best fit distributions. The difference between these distributions may be interpreted as the causal influence of the rate adjustment on the final distributions. The comparison shows the effect that a single parameter has on the observed distribution. Figure 4a and b show the results of this analysis. Panel b shows the normalized log-likelihood ratio of the counterfactual model relative to the baseline model. We used the following statistic:

$$LLR_i = \log \left( \frac{\Pr_h(\mathbf{x}_i)}{\Pr_m(\mathbf{x}_i)} \right) * \Pr_h(\mathbf{x}_i)$$

where  $i$  is the  $i_{th}$  fixation duration bin and  $h$  and  $m$  are the human and model probabilities respectively. The log-likelihood ratio is normalized by the human empirical probability as this provides a better measure of deviation between model and empirical distributions by placing greater weight on parts of the distribution that have a high empirical likelihood. A flat line would be a theoretically perfect fit. Deviations from the unity line identify a lack of fit, and the sign of the difference indicates whether the model predicts a higher or lower probability of fixation duration. Using this method we can identify the components of the theoretical model that result in a good fit to the human data.

Eliminating surprise from the model removes the impact of rapid onset timer adjustment. The removal of surprise primarily affects the early portion of the distribution, with no impact on the tail of the distribution (Figure 4b). This is evident for the luminance increase conditions, where the surprise is the only active rate modulation. The surprise reduced model (orange line) shows an excess of short fixation durations and, relative to the full model (blue line), there is little impact on the tail of the distribution. We applied the same logic to the removal of the encoding related rate modulation. There is no impact on the distribution for the luminance increase conditions, because encoding modulations are not activated when the quality of the stimulus is increased. However, encoding modulations occur in the luminance decrease conditions. In this case, a strong impact on the tail of the distribution is observed, which can be seen by comparing the red curve in Figure 4b with the blue curve. This is likely due to the fact that encoding modulations occur later in fixations. In summary, this counterfactual analysis demonstrates that these two influences, surprise and encoding, can in principle be disassociated from one another and that they have highly distinct impacts on the structure of the fixation duration distribution.

## 4 General Discussion

### 4.1 Evidence for surprise and encoding influences during naturalistic scene viewing

The model that has been introduced here combines principles that have been developed to explain patterns of eye-movements in reading (Reichle et al., 1998; Engbert et al., 2005; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014), visual search (Trukenbrod & Engbert, 2014), and scene perception (Nuthmann et al., 2010). The core principles developed here bear most resemblance to recent random-walk timer based accounts of eye-movement decision making (Nuthmann et al., 2010; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014). While many differences exist in the way these models are implemented, they form a class. The family resemblance derives from (a) their implementation of eye-movement decisions as arising from a sequential sampling process, (b) saccade programs are triggered by a rhythmic timer (c) multi-stage saccade programming architectures. Rather than presenting a competing account of eye-movement control, the modelling results presented here complement recent efforts by showing that the basic principles of this mechanism can generalize to novel task domains. Furthermore, the simulations show that additional assumptions (rate adjustment) not made by other models in this class are required to account for recently observed empirical results in the scene-viewing literature.

The present simulations focused on luminance changes to scenes (Walshe & Nuthmann, 2014). It is important to note that the distribution effects that were observed in the luminance studies have been observed in other scene adjustment paradigms as well. Glaholt et al. (2013) provided a detailed analysis of how eye-movements adapt to scene-quality changes that are made during a saccadic eye-movement. They reported distributional analyses that uncovered a pattern

of two distinctive distributional effects: an influence on the central tendency and an influence on the tail. They found that both low- and high-pass filtering resulted in a shift in the central tendency of the distributions. However, they also found that low-pass filtering resulted in a larger influence on the tail of the distributions. They argued that the removal of high-spatial frequencies induces greater challenges to scene encoding processes than does removal of low-spatial frequencies alone. Interestingly, this influence on the tail of the distribution was no longer present when the entire scene was flipped vertically or horizontally, suggesting that such a large-scale modification of the scene eliminates the benefit from high-pass relative to low-pass filtering.

Glaholt et al. (2013) suggested an explanation for these results that we refer to as the dual-process account (see also Walshe & Nuthmann, 2014). Specifically, it has been suggested that such direct influences on fixation durations in scene viewing may arise from two possible sources. The first is a surprise influence. It is hypothesized that this mechanism arises due to sudden, unexpected visual changes. An important aspect of surprise is that it does not necessarily depend on stimulus complexity or higher-order features. Therefore, it can be triggered by detection of simple visual changes across saccadic eye-movements. Thus, the surprise effect may occur very rapidly. In contrast, influence on the tail of the distribution was speculated to arise once a more detailed level of analysis has been conducted on the stimulus. This late, encoding related influence on the distribution occurs systematically in instances in which the stimulus changes are likely to result in additional processing difficulties. On this basis it was hypothesized that, complementing the early surprise effect, a late-onset encoding-related influence on fixation durations is liable to occur. The present modelling efforts represent a formalization of this hypothesis.

## 4.2 Relationship to other models

Our model extends previous models of eye-movement control in high-level tasks by assuming surprise-related and processing-related modifications to the duration of saccade programming signals. At the same time, the model has substantial overlap with previous models of eye-movement control in high-level tasks. In the CRISP model (Nuthmann et al., 2010), a timer initiates saccade programs at random intervals that is not directly tied to fixated content. Thus, the saccade timer creates a partial decoupling of the initiation of saccade programming from stimulus processing. For example, the saccade timer can initiate a saccade program immediately after fixation onset. In this case, a saccade program is initiated that may terminate the fixation but very little, if any, stimulus processing has occurred. This situation is more extreme when the saccade timer initiates a saccade program prior to the onset of a fixation. In these cases, if the saccade program is not cancelled, the saccade program that will eventually terminate the fixation was initiated on the previous fixation. However, the decoupling of the saccade programming and saccade timing is only partial. The saccade timer may be inhibited in the case that processing difficulty is high. This inhibition can result in delays in the time required to initiate saccade programs and therefore can result in extensions in fixation durations. This, along with saccade cancellation, is the primary mechanism of direct control in the CRISP model. Importantly, the CRISP model does not permit any stimulus dependent changes in the duration of saccade programming (i.e. labile and nonlabile stages), which were allowed in the present simulations.

Unlike the CRISP model, the ICAT model (Trukenbrod & Engbert, 2014) produces stimulus dependent changes to the rate at which saccade programs are initiated by the saccade timer and in the duration required for a labile program to complete. Thus, increases in fixation durations can arise from delays in the initiation of a saccade program or, more



directly, in the time required for saccade programming to complete. Furthermore, since ICAT permits the duration of labile programs to be modified on a moment-to-moment stimulus dependent basis, it accounts for adjustments in the duration of saccade programs that were initiated on a previous fixation. For this reason, ICAT can, in principle, account for increases in very short fixation durations. However, ICAT does not assume any distinction between surprise or encoding related influences on the initiation or programming of saccades. As such, it is difficult to see in what way ICAT would predict differential effects on the mode and tail of the distribution that have been observed in various scene viewing tasks (Glaholt et al., 2013; Walshe & Nuthmann, 2014).

In the LATEST model (Tatler et al., 2017), fixation times are assumed to reflect the time required to make a decision about the relative benefit of making an eye movement to a new location. Conceptually, this model could account for some effects of scene luminance adjustments observed in our task. For example, it is likely that decreasing the luminance of the scene results in increased decision time and therefore increased overall fixation times. However, because LATEST allows no impact of saccade programming it is difficult to see how it would account for the distribution effects, in particular the early shift due to surprise.

### 4.3 Inhibition of saccade programming

A novel assumption in the present modelling is the introduction of surprise and encoding modulation signals that modify the rate of saccade programming at different time scales. The surprise signal modifies the transition probability relatively early in a fixation, while the encoding signal modifies transition probabilities of longer fixations. Analysis of the simulations revealed that the surprise signal is crucial for producing the early shift in the distribution but plays very little part in extending very long fixation durations. We used a counterfactual method of analyzing the influence of a specific model component. The logic of this method involves the assumption that two models are identical in all aspects except for one critical feature, for example, surprise. Therefore, any difference in the distribution between the two models must arise from that one single model component. In Figure 4 **a** and **b** it is clear that by removing surprise, the early shift in the distribution is nearly completely removed. However, removing surprise does not eliminate the effects present in the tail of the distribution. On the other hand, removing encoding modulation had very little influence on the early parts of the distribution. However, removal of encoding modulation results in a clear reduction in the proportion of long fixation durations. Our counterfactual analysis thus demonstrates that these two hypothetical mechanisms have distinctive impacts on the shape of fixation duration distributions.

The modelling framework we have introduced allows for modelling the role of these influences in other tasks, which may shed light on contradictory findings. For example, Henderson et al. (2014) found shorter fixation durations when replacing a blurred scene with a clear unfiltered scene. This seems to suggest that there was encoding related *facilitation* at work, and that it outweighed the surprise related inhibition. In principle, this could be tested with a model in which the encoding related influence is allowed to *speed up* the timer. Logically, this is a reasonable model assumption to make in that scene changes which make the stimulus easier to process may result in quicker saccade programming timelines. Moreover, simulations with the model may be used to identify experimental conditions in which existing encoding related facilitation was outweighed by surprise related inhibition, yielding no observable shortening of mean fixation durations.

#### 4.4 Predictive model of surprise modulation

Surprise related changes in the timing of saccadic eye-movements may be understood within the context of predictive models of perception and action. Within this perspective, the surprise related changes in the planning of an impending eye-movement results from the mismatch between the statistics of the visual environment that are expected to be present prior to an eye-movement, and the statistics (i.e. lower luminance environment) that are discovered at the onset of the next fixation. To the extent to which these expectations are violated, the system responds by reevaluating the motor control parameters that are used to govern the timing of saccades. In studies of saccadic reaction times it has been shown that saccadic latency increases with decreasing target probability (Carpenter & Williams, 1995). This fact has been used to suggest that saccade related decisions are impacted by the probabilistic context in which a target is embedded. It has also been shown that saccadic reaction times increase with the degree to which the location of a target violates prior expectations (Vossel, Mathys, Daunizeau, Bauer, Driver, Friston & Stephan, 2014). The large-scale stimulus change that occurs in gaze-contingent scene quality studies may induce such a violation of expectations. In predictive coding accounts of perceptual processing, it is claimed that perceptual circuits convey signals relative to the difference between anticipated and observed inputs (Huang & Rao, 2011). One functional significance of such a coding system is that redundancies in perceptual inputs may be reduced or removed, allowing for a more compact and efficient neural code. Predictive signals may also be used as a method of smoothing errors and deemphasizing perceptual samples arising due to chance fluctuations (Burr & Cicchini, 2014). However, an adaptive system must also retain the capacity to adjust its behaviour when the sensory evidence suggests a large-scale violation of expectations. Such an error signal is used by the motor system to adjust saccade amplitudes in the context of saccadic (Wallman & Fuchs, 1998) and motor adaptation tasks (Berniker & Kording, 2008), where the motor gain on future trials depends on the observed error in the current response. Predictive signalling has been shown to exist as early as the retina (Hosoya, Baccus & Meister, 2005) and at later stages of the visual pathway including the lateral geniculate nucleus (Dong & Atick, 1995) and the visual cortex (Rao & Ballard, 1997). In the present context, it is possible that the large scale visual change in the structure of the image induces a change in the spatial structure, which leads to an error signal being generated. This error signal may rapidly reach the frontal eye fields or superior colliculus, resulting in a modification to saccade planning, which is here modelled as a slow-down in saccade related decision signals. This hypothesis remains to be further explored by experiments. However, we note that it is consistent with numerous accounts of how sensory decisions are made, and with the underlying neurophysiology.

##### 4.4.1 Conclusion

This work has been motivated by the hypothesis that the eye-movement control system adapts to stimulus changes that occur on a fixation-to-fixation basis. The novel hypothesis that we introduce is that there are two qualitatively distinct types of influence, a rapid onset surprise and a delayed onset encoding modulation. It is shown that this hypothesis can explain fixation duration distributions measured from scene luminance adjustments, which may generalize to other recent experiments on the control of fixation durations. Future work is needed to address the generalizability of these simulations to novel task domains and stimulus variants.

## 5 Appendix

### 5.1 Formal derivation of the model

Here, we develop the mathematical formulation of the model architecture. The model shares many core features with other random walk approaches to modelling fixation durations during high-level tasks (Nuthmann et al., 2010; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014). A unifying feature of these models is that a random walk process is implemented to account for stochastic variability at the level of saccade initiation intervals. Following previous models (Nuthmann et al., 2010; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014), we implement the saccade timer as a discrete-state, continuous-time Markov process (Gillespie, 1978). First, we derive the random walk process for the case where the saccade initiation timer is the only active mechanism. Later, we extend the model derivation for the case in which multiple saccade programming mechanisms are activated in parallel.

### 5.2 Timer-only random walk

A random walk in state  $m$ , at time  $t$ , is given by  $S_m(t)$  with initial state given by  $S_0(0)$ . An elementary transition occurs when the random walk changes state from  $S_m$  to an adjacent state  $S_n$ . The random walk continues until it reaches a threshold value such that  $n = N$ . Once the random walk reaches threshold it is reset to the initial state  $S_0$ . Elementary transitions of the random walk occur at continuously valued intervals. Therefore, a random walk that is in state  $m$  at time  $t$  will transition to the next state  $n$  at time  $t + \tau$ :  $S_m(t) \rightarrow S_n(t + \tau)$ . The duration that the random walk pauses in a given state is referred to as the waiting time. The mean waiting time  $\tau$  of a timer step is related to the transition probability rate of the random walk through the equation

$$w_1 = \frac{N_{timer}}{T_{timer}}, \quad (1)$$

where  $w_1$  governs how quickly the random walk transitions between adjacent states. The rate parameter is equal to the proportion of the total number of states  $N_{timer}$  of the timer random walk to the expected duration required for the timer to reach threshold  $T_{timer}$ .

For a discrete-state continuous-time Markov process with constant transition probability rate with respect to  $\tau$ , the probability distribution defined over values of  $\tau$  is given by the exponential distribution

$$\rho(\tau) = W_n e^{-W_n \tau}. \quad (2)$$

In the simple case, in which only the random walk of the timer is active, the rate parameter of the exponential distribution  $W_m$  is given by the transition probability rate of the random walk such that  $W_m = w_1 = w_{timer}$ . Specific realizations of  $\tau$  may be obtained by applying the function

$$\tau = \frac{1}{W_m} \log(1 - \epsilon), \quad (3)$$

where  $\epsilon$  is a pseudorandomly generated number over the interval  $0 \leq r \leq 1$  (Gillespie, 1978).

The formulation provides a description of the evolution of the random walk process under the conditions that the random timer is the only active stochastic process (see Nuthmann et al., 2010).

### 5.3 Multiple random walks

To fully simulate the mechanisms governing the timing of eye-movements, we rely on an approach to simulate random walks that approach a bound in parallel. This method has previously been applied to a model of visual search and reading (Trukenbrod & Engbert, 2014) and is a generalization of the Monte Carlo approach used to simulate the discrete-state continuous-time Markov process introduced previously for the timer-only model (Gillespie, 1978). We model five independent one-step processes to account for stochastic variability in the preparation and execution of saccadic eye-movements. We also describe how properties of the environment are modelled by parameters of the random walk.

The composite description of the state of the model at time  $t$  is given by the vector  $S_m(t) = (m_1, m_2, m_3, m_4, m_5)$ <sup>5</sup> where  $\{m_i \in \mathbb{N} | m_i \leq N_i\}$  and  $N_i$  is the number of steps in the  $i^{th}$  random walk. Therefore, the overall dynamic state of the random walk is determined by the state of each of the elementary random walks. Transitions between states in the model are given by  $S_{nm}$  and is equivalent to a single elementary state change  $m_i + 1$ , for some  $i \in 1 \dots 5$ . In effect, a change in state occurs when an elementary random walk accumulates by one step towards its threshold value  $N_i$ . An elementary random walk is reset to an initial state in the case that it reaches threshold  $m_i = N_i$ .

Given that random walks may be activated in parallel, the transition probability rate must be modified accordingly. Previously, in the timer-only case, only a single transition was possible i.e.  $S_{nm} = m_1 + 1$  and the total transition probability rate was determined by the rise-rate of this single random walk. Generalizing equation 1, the transition probability rate for an elementary random walk at time  $t$  is given by

$$w_i(t) = \frac{N_i}{T_i}, \quad (4)$$

where  $w_i$  is associated with random walk  $m_i$ . When multiple random walks are active in parallel, we compute the total transition probability rate as the sum of the individual transition probability rates associated with active elementary time-dependent random walks. Therefore, the total transition probability rate  $W_n(t)$  is given by

$$W_n = \sum_{i \in A} w_i(t), \quad (5)$$

where  $A$  gives the index of active random walks.

In the model, changes in the rate at which the labile and nonlabile elementary random walks accumulate to threshold is subject to time-dependent changes. These rate modulations reflect the impact that visual changes sampled during a fixation have on the programming of an upcoming saccadic eye-movement. Two distinct forms of timer adjustment are modelled. What is referred to as surprise occurs shortly after the onset of a saccade. The timecourse of surprise is also modelled as a change that has a limited duration. Surprise is purely inhibitory as can be seen from the following equation,

$$w_i(t) = \begin{cases} \alpha_i w_i(t) & \text{if } t \leq \beta_S \text{ and } 0 < \alpha_i \leq 1 \text{ and } i \in \{timer, lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases}, \quad (6)$$

---

<sup>5</sup> $m_1$  gives the state of the random timer,  $m_2$  the labile state,  $m_3$ , the nonlabile state,  $m_4$  the efferent transmission time,  $m_5$  saccade execution time.

565 for the luminance increase and,

$$w_i(t) = \begin{cases} \epsilon_i w_i(t) & \text{if } t \leq \beta_S \text{ and } 0 < \epsilon_i \leq 1 \text{ and } i \in \{timer, lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases}, \quad (7)$$

566 for luminance decrease. The parameters  $\beta$  and  $\gamma$  provide the temporal boundary during which modulation influences  
 567 the random walk. The onset of timer-related encoding changes occur after the onset of surprise modulation and is  
 568 given by

$$w_i(t) = \begin{cases} \zeta_i w_i(t) & \text{if } t \geq \gamma_E \text{ and } 0 < \zeta_i \leq 1 \text{ and } i \in \{timer, lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases}. \quad (8)$$

#### 569 5.4 Model fitting

570 The objective function that was used to assess candidate model fitness was defined as the likelihood  $\ell(\theta; x_1, \dots, x_n)$   
 571 of observing the empirical data  $x_1 \dots x_n$  given the parameters of the candidate model  $\theta$ . We optimized this function  
 572 using Bayesian adaptive direct search (Acerbi & Ji, 2017).

573 In the case that observations are independently and identically distributed, the likelihood function takes the following  
 574 form

$$\mathcal{L}(\theta; x_1, \dots, x_n) = \prod_{i=1}^k P(x_i|\theta). \quad (9)$$

575 Since the logarithm is a monotonic function of the likelihood, maximizing the likelihood function is equivalent to  
 576 maximization of the log-likelihood function

$$(\log(\mathcal{L}(\theta; x_1, \dots, x_n))) = \left( \sum_{i=1}^k \log(P(x_i|\theta)) \right). \quad (10)$$

577 Computing  $P$  for a specific observed fixation duration is equivalent to computing

$$\begin{aligned} \log(\mathcal{L}(\theta; x_1, \dots, x_n)) &= \sum_{i=1}^k \log(q_i^{n_i}) \\ &= \sum_{i=1}^k n_i \log(q_i) \end{aligned} \quad (11)$$

578 where  $q_i$  is the probability of the  $i^{th}$  fixation determined by the model  $\theta$ , and  $n_i$  is the number of times the  $i^{th}$  fixation  
 579 duration was observed. In equation 12,  $n_i$  can be substituted for  $p_i$  which gives the proportion of the  $i^{th}$  fixation  
 580 duration observed in the empirical data. This substitution is possible because  $p_i > p_j$  whenever  $n_i > n_j$  for all  $i \neq j$ .  
 581 Typically, we optimize a functional minimum value so it is more convenient to instead search for a minimum of the  
 582 negative of the log-likelihood.

$$(-\log(\mathcal{L}(\theta; x_1, \dots, x_n))) \quad (12)$$

	Experiment 1	Experiment 2
Timer	226.27	230.25
Labile	174.24	192.24
Non Labile	71.60	83.67
Motor	30.00	30.00
Saccade	20.00	20.00
Number of States	14.00	16.00
Surprise Offset	42.67	60.56
Surprise (Down)	0.24	0.15
Encoding Offset	387.81	247.19
Encoding (Down)	0.37	0.62
Surprise (UP)	0.49	0.54
Encoding (UP)	1.00	1.00

Table 1: Parameter values estimated for the two experiments that best fit the human experimental data. Timer, Labile, Non Labile, Motor and Saccade refer to the mean duration of the random walk process associated with that stage of saccade preparation. Surprise Offset specifies the onset of the surprise modulation relative to the onset of fixation and Encoding Offset refers to the onset of the encoding modulation relative to the end of surprise. The rest of the parameters describe the scalar value applied to the rate of the random walk when surprise and encoding modulations are activated. Encoding (UP) was fixed at 1 because we did not allow encoding facilitation in this model. The upper and lower bounds on the parameters during optimization were very large and outside the range of plausible models.

## References

- Acerbi, L. & Ji, W. (2017). Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. In *Advances in Neural Information Processing Systems*, (pp. 1836–1846).
- Becker, W. (1991). Saccades. In R. Carpenter (Ed.), *Eye-Movements: Vol. 8 in Vision and visual Dysfunction* (pp. 95–137). London: Macmillan.
- Becker, W. & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19(9), 967–983.
- Berniker, M. & Kording, K. (2008). Estimating the sources of motor errors for adaptation and generalization. *Nature Neuroscience*, 11(12), 1454–1461.
- Burr, D. & Cicchini, G. M. (2014). Vision: Efficient Adaptive Coding. *Current Biology*, 24(22), R1096–R1098.
- Camalier, C. R., Gotler, A., Murthy, A., Thompson, K. G., Logan, G. D., Palmeri, T. J., & Schall, J. D. (2007). Dynamics of saccade target selection: Race model analysis of double step and search step saccade production in human and macaque. *Vision Research*, 47(16), 2187–2211.
- Carpenter, R. H. S. (1981). Oculomotor procrastination. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 237–246). Hillsdale, NJ: Erlbaum.
- Carpenter, R. H. S. & Williams, M. L. L. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, 377(6544), 59–62.

- Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 101(35), 13086–13090.
- Dong, D. W. & Atick, J. J. (1995). Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, 6(2), 159–178.
- Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42(5), 621–636.
- Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777–813.
- Gillespie, D. T. (1978). Monte carlo simulation of random walks with residence time dependent transition probability rates. *Journal of Computational Physics*, 28(3), 395–407.
- Glaholt, M. G., Rayner, K., & Reingold, E. M. (2013). Spatial frequency filtering and the direct control of fixation durations during scene viewing. *Attention, Perception, & Psychophysics*, 75(8), 1761–1773.
- Glaholt, M. G., Rayner, K., & Reingold, E. M. (2014). A rapid effect of stimulus quality on the durations of individual fixations during reading. *Visual Cognition*, 22(3-4), 377–389.
- Glaholt, M. G. & Reingold, E. M. (2012). Direct control of fixation times in scene viewing: Evidence from analysis of the distribution of first fixation duration. *Visual Cognition*, 20(6), 605–626.
- Henderson, J. M., Nuthmann, A., & Luke, S. G. (2013). Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 318–322.
- Henderson, J. M., Olejarczyk, J., Luke, S. G., & Schmidt, J. (2014). Eye movement control during scene viewing: Immediate degradation and enhancement effects of spatial frequency filtering. *Visual Cognition*, 22(3-4), 486–502.
- Henderson, J. M. & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, 15(3), 566–573.
- Hooge, I. T. C. & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research*, 38(9), 1295–1302.
- Hooge, I. T. C., Vlaskamp, B. N. S., & Over, E. A. B. (2007). Saccadic search: On the duration of a fixation. In R. P. G. Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 581–595). Oxford: Elsevier.
- Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, 436(7047), 71–77.
- Huang, Y. & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5), 580–593.
- Kennison, S. M. & Clifton, Jr, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: General*, 121(1), 68–81.
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1), 12.

- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological review*, 104(3), 524.
- Ludwig, C. J. H., Mildinhal, J. W., & Gilchrist, I. D. (2007). A population coding account for systematic variation in saccadic dead time. *Journal of Neurophysiology*, 97(1), 795–805.
- McConkie, G. W. & Loschky, L. C. (2002). Perception onset time during fixations in free viewing. *Behaviour Research Methods Instruments & Computers*, 34(4), 481–490.
- 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, 814–840.
- McPeck, R., Skavenski, A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision research*, 40(18), 2499–2516.
- Noorani, I. & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience and Biobehavioral Reviews*, 64, 229–251.
- Nuthmann, A. (2017). Fixation durations in scene viewing: Modeling the effects of local image features, oculomotor parameters, and task. *Psychonomic Bulletin & Review*, 24(2), 370–392.
- Nuthmann, A. & Henderson, J. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition*, 20(4-5), 457–494.
- 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.
- Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, & Psychophysics*, 73(4), 1120–1132.
- Rao, R. P. & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, 9(4), 721–763.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–506.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157.
- Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye-mind link. *Psychological Review*, 119(1), 155–185.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26(4), 445+.
- Reichle, E. D. & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience*, 7(361), 1–6.
- Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology*, 65(2), 177–206.
- 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.
- Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, 24(2), 113–121.



- Saez de Urabain, I. R., Nuthmann, A., Johnson, M. H., & Smith, T. J. (2017). Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research*, 134, 43–59.
- Schad, D. J. & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition*, 20(4-5), 391–421.
- Shen, K. & Paré, M. (2014). Predictive saccade target selection in superior colliculus during visual search. *The Journal of Neuroscience*, 34(16), 5640–5648.
- Shioiri, S. (1993). Postsaccadic processing of the retinal image during picture scanning. *Perception & Psychophysics*, 53(3), 305–314.
- Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review*, 124(3), 267–300.
- Trukenbrod, H. A. & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, 21(4), 904–34.
- Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., & Stephan, K. E. (2014). Spatial attention, precision, and Bayesian inference: A study of saccadic response speed. *Cerebral Cortex*, 24(6), 1436–1450.
- Wallman, J. & Fuchs, A. F. (1998). Saccadic gain modification: Visual error drives motor adaptation. *Journal of Neurophysiology*, 80(5), 2405–2416.
- Walshe, R. C. & Nuthmann, A. (2013). Programming of saccades to double-step targets in scene viewing: A test of assumptions present in the CRISP model. In Knauff, M., Pauen, M., Sebanz, N., & Wachsmuth, I. (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, (pp. 1569–1574), Berlin, Germany. Cognitive Science Society.
- Walshe, R. C. & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research*, 100, 38–46.
- Walshe, R. C. & Nuthmann, A. (2015). Mechanisms of saccadic decision making while encoding naturalistic scenes. *Journal of Vision*, 15(5)(21), 1–19.
- Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *Archives of Ophthalmology*, 52(6), 932–941.
- Wu, E. X. W., Chua, F.-K., & Yen, S.-C. (2016). Saccade plan overlap and cancellation during free viewing. *Vision Research*, 127, 122–131.
- Wu, E. X. W., Gilani, S. O., van Boxtel, J. J. A., Amihai, I., Chua, F. K., & Yen, S.-C. (2013). Parallel programming of saccades during natural scene viewing: evidence from eye movement positions. *Journal of Vision*, 13(12)(17), 1–14.