A Computational Dual-Process Model of Fixation-Duration Control in Natural Scene Viewing

R. Calen Walshe¹ · Antje Nuthmann²

Accepted: 11 May 2021 / Published online: 1 September 2021
© The Author(s) 2021

Abstract

Research on eye-movement control during natural scene viewing has investigated the degree to which the duration of individual fixations can be immediately adjusted to ongoing visual-cognitive processing demands. Results from several studies using the fixation-contingent scene quality paradigm suggest that the timing of fixations adapts to stimulus changes that occur on a fixation-to-fixation basis. Analysis of fixation-duration distributions has revealed that saccade-contingent degradations and enhancements of the scene stimulus have two qualitatively distinct types of influence. The *surprise effect* begins early in a fixation and is tied to surprising visual events such as unexpected stimulus changes. The *encoding effect* is tied to difficulties in visual-cognitive processing and occurs relatively late within a fixation. Here, we formalize an existing descriptive account of these two effects (referred to as the dual-process account) by using stochastic simulations. In the computational model, surprise and encoding related influences are implemented as time-dependent changes in the rate at which saccade timing and programming are completed during critical fixations. The model was tested on data from two experiments in which the luminance of the scene image was either decreased or increased during selected critical fixations (Walshe & Nuthmann, *Vision Research*, 100, 38–46, 2014). A counterfactual method was used to remove model components and to identify their specific influence on the fixation-duration distributions. The results suggest that the computational dual-process model provides a good account for the data from the luminance-change studies. We describe how the simulations can be generalized to explain a diverse set of experimental results.

Keywords Computational modelling · Random walk · Fixation duration · Saccades · Naturalistic scenes · Double step

Introduction

The study of eye movements during naturalistic scene viewing involves identifying the mechanisms that allow the visual system to access vital visual information. In typical tasks, humans make three to four saccadic eye movements per second (Rayner, 2009). Whenever the eyes shift to a new location, the visual system requires time to process

the updated visual information acquired during the current fixation. An important topic of active research is the degree to which the duration of an individual fixation can be flexibly adjusted to ongoing visual-cognitive processing demands (Nuthmann, 2017, for review).

The goal of this article is to introduce a computational dual-process model of fixation-duration control during scene viewing. We start by presenting theoretical positions and empirical findings regarding the control of fixation durations in scenes. Next, we will illuminate what is known from the classic double-step paradigm about observers' ability to modify their saccade programs based on new incoming visual information. Since our computational model incorporates principles from existing models of eye-movement control in high-level tasks, we proceed by introducing selected models in more detail. We then present recent, seemingly conflicting results on the issue of unidirectional (slow down) as opposed to bidirectional (slow down and speed up) adjustment of fixation durations in scene viewing, together with an existing descriptive

✉ R. Calen Walshe
calen.walshe@utexas.edu

✉ Antje Nuthmann
nuthmann@psychologie.uni-kiel.de

¹ Center for Perceptual Systems, University of Texas at Austin, Austin, Texas 78712, USA

² Kiel University, Institute of Psychology, Olshausenstr. 62, 24118 Kiel, Germany

dual-process model to account for these findings. In the remainder of the article, we will present a simulation-based model that merges the dual-process account with recent models of eye-movement control.

Directly Controlled Fixation Durations

Eye-movement control models can be broadly contrasted as direct-control versus indirect-control models (Trukenbrod & Engbert, 2014, for review). Accounts of eye-movement control that adopt strong direct-control assumptions suggest that stimulus processing occurs rapidly enough to influence the timing of the next saccadic decision. By contrast, indirect-control models assume that the properties of the fixated stimulus do not have an impact on the duration of the current fixation. 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 et al., 2006) or when the text is presented in visually degraded format (Glaholt et al., 2014). In the context of scene viewing, evidence in support of direct control has been provided by the fixation-contingent scene quality paradigm, in which the quality of the scene is manipulated during selected critical fixations (Glaholt et al., 2013; Henderson et al., 2013; Henderson et al., 2014; Walshe & Nuthmann, 2014).

For example, Henderson et al. (2013) reduced the luminance of the scene on every n^{th} saccade, using different margins (e.g., a reduction to 60% of the original luminance, see Fig. 1a). During the next saccade, the scene returned to

its normal luminance. The duration of the critical fixation between these two saccades was immediately affected by the reduction in scene luminance, with increasing durations for decreasing luminance (see also Walshe & Nuthmann, 2014). Similarly, fixation durations were lengthened when removing high or low spatial frequencies from the scene images in a gaze-contingent manner (Glaholt et al., 2013; Henderson et al., 2014). In summary, compelling evidence exists to support the claim that some proportion of fixations are directly controlled by the stimulus content.

However, research has also demonstrated that not all fixations are directly controlled. For example, when interpreting the results of their study on direction-coded search, Hooze and Erkelens (1998) argued that performance failures could be explained by participants' inability 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.

Several studies used the stimulus onset delay paradigm (Morrison, 1984) to investigate the control of fixation durations during scene perception (Henderson & Pierce, 2008; Henderson & Smith, 2009; Luke et al., 2013; Shioiri, 1993). In the scene onset delay (SOD) paradigm, the onset of the scene stimulus is delayed by presenting a visual mask at the beginning of selected critical fixations (Fig. 1b). After the delay, the duration of which is varied, the scene (re)appears. The underlying logic is that stimulus processing can only begin after the scene stimulus has become available. SOD studies have consistently identified one population of fixations that increased in duration as the delay increased, suggesting that these durations were

Fig. 1 Gaze-contingent paradigms to test the direct-control hypothesis regarding fixation durations in scene viewing. a. Fixation-contingent scene quality paradigm. At time A, during the fixation preceding a selected critical fixation, the base scene is presented. In this illustration, the untouched original scene image serves as the base scene. After the beginning of the next saccade is detected (B), the scene image is changed (vertical broken line). Here, the original scene is replaced with a scene image in which the luminance is strongly reduced, from 100% to 60% (Henderson et al., 2013). After the start of the next saccade is detected (D), the base scene is restored via another display change (vertical broken line). Thus, the luminance-reduced scene image is

presented for the entire duration of the critical fixation. Note that the scene images are changed during saccadic eye movements when the processing of new visual information is suppressed (Ross et al., 2001). b. Scene onset delay paradigm. At time A, the normal scene image is presented. During the next saccade, which is represented by the oblique line, the scene is replaced by a noise mask. Following a delay (D), the duration of which is experimentally manipulated, the scene reappears. Thus, when the eyes begin the critical fixation (C), the scene has been removed from view. The saccade terminating this fixation can happen during the presentation of the noise image or after it is removed

directly controlled by the current scene image. However, not all fixations were affected by the presence of the SOD and its length. This pattern of results is consistent with a mixed-control account, with one population being directly controlled and the other being indirectly controlled (Henderson & Pierce, 2008). For a more comprehensive review of empirical findings on the control of fixation durations in scenes, the reader is referred to Nuthmann (2017).

A challenge that direct-control theories face arises from considering several timing constraints imposed on stimulus processing within a single fixation. For instance, it is estimated that approximately 90 ms are required to transmit visual signals from the retina to the brain and then visually encode the image for subsequent processing (Reichle & Reingold, 2013). Additionally, programming a saccade to the next location is estimated to require 175–250 ms (Salthouse & Ellis, 1980). This includes a lag of approximately 30 ms in transmitting the motor command from the brain to the muscles of the eye (Becker, 1989, 1991). Therefore, the combination of these constraints potentially leaves little time available for the foveated stimulus to exert a real-time influence on the current fixation duration. To develop plausible theories of eye-movement control that include direct-control assumptions, it is essential to give proper consideration to these timing constraints.

Reprogramming of Saccades: Double-Step Paradigm

In the fixation-contingent scene quality paradigm and in the SOD paradigm, the scene image is unexpectedly changed while observers are actively engaged in extracting information from it. The observation that fixation durations are adjusted in these situations implies that observers were able to flexibly reprogram saccades based on new incoming visual information. The extent to which a saccade program is modifiable once it has been partially or completely prepared can be investigated using the double-step paradigm (Becker & Jürgens, 1979; Ludwig et al., 2007; Westheimer, 1954; Wheelless et al., 1966). Figure 2a illustrates the procedure of a typical double-step trial, using the parameters of the static condition in our own double-step study (Experiment 1 in Walshe & Nuthmann, 2015). The trial began with the presentation of a central fixation cross. After 2,000 to 3,000 ms, a small target box appeared at 7° eccentricity either to the left or right for a short amount of time. In half of the trials, the target then stepped to a second location at 14° eccentricity. As in many previous double-step studies, the target locations were confined to the horizontal meridian (e.g., Becker & Jürgens, 1979; Westheimer, 1954; Wheelless et al., 1966). The subject's task was to follow the target with their eyes. Given that the target stepped in quick succession from 7° to 14°, the task

required the subject to lengthen the saccade amplitude they had been preparing. To investigate the temporal constraints for such reprogramming to occur, each double-step trial was presented under one of four interstep intervals: either 50, 100, 150 or 200 ms elapsed between the first and second target steps.

The main experimental finding of the double-step paradigm is the amplitude transition function (ATF) (Becker & Jürgens, 1979; Kalesnykas & Hallett, 1987). To obtain an ATF, the amplitude of the first saccade made by one or several observers is plotted against a time variable called Delay (D) (Fig. 2b). D measures the time from the second step of the target to the first saccade. Of course, D is influenced by the interval between the two target steps; specifically, short interstep intervals tend to be associated with long delays, and long intervals with short delays. Moreover, D is influenced by the intrinsic variability in the saccade initiation process (Findlay, 1992). Double-step studies have consistently revealed that the first saccade made by a subject will be one of three types. For small values of D , saccades are aimed at the first position of the target. For large values of D , saccades are predominantly aimed at the second position of the target. For intermediate values of D , transition responses occur in which saccades are aimed at intermediate positions. Empirical data from double-step experiments can be fit by mathematical functions like a cumulative Gaussian distribution (Ludwig et al., 2007) or a logistic function (Walshe & Nuthmann, 2015). Figure 2b depicts an example ATF that was obtained by fitting a four-parameter logistic function with a nonlinear mixed-effects regression framework.

The Delay D measures the time available for the second target step to perturb the saccade that is in preparation to the first target position (Findlay, 1992). When D is long, subjects are able to cancel or modify the initial saccade program, which allows them to ignore the appearance of the target in the first position and to move their eyes directly to the second target position (upper bound of the logistic function in Fig. 2b). When D is short, however, the perturbing second step does not affect the primary saccade. In this case, the saccade program was already well underway and could not be altered anymore. As a result, the eyes were directed to the first target position (lower bound of the logistic function in Fig. 2b). Importantly, the point at which the amplitude transition starts is interpreted as the last point in time at which it is possible to modify or cancel an ongoing saccade program. For the data depicted in Fig. 2b, this point-of-no-return was estimated to be 74 ms. Thus, during the final 74 ms prior to saccade execution, the planned movement could not be altered anymore. In computational models of eye-movement control, this final stage of saccade programming is often referred to as the

Fig. 2 Double-step paradigm. a. Procedure involved in a typical double-step trial (static condition in Walshe & Nuthmann, 2015). Participants fixate on a red cross at the centre of the screen. After 2,000 to 3,000 ms, a target appears at 7° eccentricity either to the left or right for 50, 100, 150 or 200 ms (interstep interval). In 50% of the trials, this target is replaced by a second identical target appearing at 14° eccentricity in the same direction as the first target. Participants are asked to follow the stepping target with their eyes. b. Amplitude transition function, showing how saccade amplitudes gradually transition from

the initial to the final position of the target as the Delay, measuring the time from the second step of the target to the first saccade, increases. The graph does not show the actual data points but the prediction of a four-parameter logistic function that was fit to the data from the static condition in Experiment 1 reported by Walshe and Nuthmann (2015). The horizontal solid black lines represent the physical distance of the targets from the central fixation cross (step 1: 7°, step 2: 14°). The vertical broken black line marks the point-of-no-return (PONR)

nonlabile stage, which is preceded by a labile phase during which the saccade program is still modifiable (Engbert et al., 2005; Nuthmann et al., 2010; Reichle et al., 1998; Trukenbrod & Engbert, 2014).

Classic variations of the double-step paradigm involve the presentation of simple stimuli on a uniform background. More recently, Walshe and 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. Moreover, their data suggest that saccade modification deadlines may be longer in scene viewing than commonly reported in the double-step literature.

Eye-Movement Control Models in High-Level Tasks

Eye-movement control theories have been most actively developed within the study of reading behaviours (Engbert et al., 2005; Legge et al., 1997; McDonald et al., 2005;

Reichle et al., 1998). Reading presents fertile ground for testing such theories because the eye-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 et al., 2010; Reichle et al., 2012; Tatler et al., 2017; Trukenbrod & Engbert, 2014). The model we introduce here combines principles that have been developed to explain patterns of eye movements in reading (Engbert et al., 2005; Reichle et al., 1998; Schad & Engbert, 2012; Trukenbrod & Engbert, 2014), visual search (Trukenbrod & Engbert, 2014) and scene perception (Nuthmann et al., 2010). The model's core principles 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). In the following, we therefore briefly introduce the CRISP model (Nuthmann et al., 2010) and the ICAT

model (Trukenbrod & Engbert, 2014). We also include the LATEST model (Tatler et al., 2017), which is currently the only model to explain both fixation locations and fixation durations during scene viewing.

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 et al., 2002; Engbert et al., 2005). In the CRISP model, the timer is implemented as a stochastic random walk towards a threshold. Once the threshold value is reached, a new saccade program is initiated and then completed in two stages (labile and nonlabile). CRISP incorporates direct-control mechanisms in that moment-to-moment difficulties in visual and cognitive processing can immediately inhibit (i.e., delay) saccade timing and programming, leading to longer fixation durations. When modelling empirical data collected with the SOD paradigm (Fig. 1b), the specific assumptions were that (a) current processing demands inhibit the random walk's transition rate, and (b) processing difficulties can lead to saccade cancellation. After the scene disappears from view, the mean random walk transition rate is considerably reduced until the scene reappears. Moreover, if a labile saccade program is active when the scene disappears, it is cancelled with a certain probability.

By contrast, indirect control occurs when a saccade program is initiated prior to the onset of the current fixation, with no processing related cancellation occurring. In these cases, modifications to the random walk timing signal will have limited or no impact on the current fixation duration. Finally, saccade programs that are started before the onset of the current fixation tend to produce short fixation durations (Nuthmann et al., 2010). Therefore, the CRISP model implicitly predicts that short fixations tend to avoid the influence of direct control. In summary, given the way saccade timing and programming are conceptualized in CRISP, the model may be viewed as implementing mixed control of fixation durations.

ICAT model

As in the CRISP model, saccade programs are initiated according to a random-walk timing process that may be subject to processing related inhibition. Once the timer reaches threshold, a new saccade program is initiated. In addition, ICAT permits continuous-time inhibition at the level of saccade programming. This is possible because the labile, nonlabile and execution stages of saccade programming are

all implemented as independent random walks, as in Schad and Engbert (2012). Specifically, inhibition is applied to the labile but not to the nonlabile stage (Trukenbrod & Engbert, 2014). When foveal processing difficulty increases, the transition rate of the random walk for the labile stage of saccade programming is reduced. This is different from the original formulation of the CRISP model, where labile saccade programs are not prolonged in response to increasing processing demands. Instead, in CRISP a labile program can be cancelled in the extreme case that the stimulus is taken away from view (Nuthmann et al., 2010).

ICAT makes very elaborate assumptions about the relationship between stimulus processing and the duration of saccade timing and saccade programming processes. Within ICAT, there is a complex interplay between influences on eye movements defined by a dichotomy introduced as local versus global control. On the one hand, ICAT introduces global control principles to account for general aspects of a task's difficulty and fixation history. On the other hand, local control principles account for effects that arise within a single fixation. Local control is further subdivided into local-I and local-II. Local-I refers to variability in fixation durations due to the stochastic fluctuation in the random-walk timer responsible for initiating saccade programs. By contrast, local-II represents changes in the mean rate of the random walk due to immediate stimulus changes or processing difficulties encountered during a fixation. Thus, the local-II component is responsible for modelling direct-control influences on fixation durations in ICAT.

The CRISP model (Nuthmann & Henderson, 2012; Nuthmann et al., 2010), 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 et al., 2003; Reichle et al., 2012) they all assume that saccades are programmed in multiple stages (Becker & Jürgens, 1979).

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, LATER-like 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 the LATEST model, fixation durations are assumed to reflect decision time only, without taking additional processes like saccade programming into account (Tatler et al., 2017). In the model, gaze control is viewed as a series of Stay-or-Go decisions. Each decision involves an evaluation of the benefit of remaining fixated at the current location (Stay) relative to the benefit of making a saccade to a new location (Go). 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 et al., 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).

Asymmetrical Control of Fixation Durations, Dual-Process Account

As outlined above, results from studies using the fixation-contingent scene degradation paradigm indicate that fixation durations are under the immediate direct control of the currently visible scene (Henderson et al., 2013). The principled mechanism in the CRISP model to accommodate such adjustments is processing-related inhibition: When the quality of the scene is reduced at the beginning of a critical fixation, the rate at which the timer accumulates to the threshold is reduced as well. In the original formulation of the CRISP model (Nuthmann et al., 2010), modulations of the random walk timer were exclusively unidirectional (timer slowdown). However, in subsequent simulation studies it was pointed out that additional experimental work would be required to investigate the directionality or symmetry in the way fixation durations are adjusted (Nuthmann & Henderson, 2012).

Therefore, Walshe and Nuthmann (2014) conducted luminance-change experiments in which both degradations and enhancements of scene images were implemented gaze-contingently. In similar experiments by Henderson et al. (2013), the base luminance level was always 100% (see Fig. 1a). By contrast, Walshe and Nuthmann (2014)

Fig. 3 Scene enhancement manipulations in different studies using the fixation-contingent scene quality paradigm (Fig. 1a). 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. The red upward arrow means that the scene enhancement manipulation was associated with an increase in fixation duration (which was not predicted by either hypothesis), whereas a blue downward arrow signals a decrease in fixation duration. FD = fixation duration

used lower base luminance levels (Experiment 1: 80%, Experiment 2: 60%), which allowed them to either enhance or degrade the scene stimulus. In Experiment 1, the luminance was either increased from 80% to 100% (Fig. 3a) or decreased from 80% to 60%. The scene-quality changes were even stronger in Experiment 2, where the luminance was either increased from 60% to 100% (cf. Figure 1a, which shows the inverse shift from 100% to 60%) or decreased from 60% to 20%. The main question to be addressed was whether making the stimulus easier to process would lead to a decrease in fixation durations (symmetric control hypothesis) or not (asymmetric control hypothesis). The results are shown in Fig. 4. Degrading the scene stimulus by shifting luminance down resulted in an increase in fixation durations (Fig. 4b). Interestingly, enhancing the scene stimulus by shifting luminance upwards did not result in a comparable decrease in fixation durations. Instead, fixation durations were significantly increased (schematic depiction in Fig. 3a). However, the increase was much smaller in the UP condition than in the DOWN condition of a given experiment (Fig. 4b). These results, along with other results from reading and visual search tasks (reviewed in Walshe & Nuthmann, 2014), 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, an explanation that we refer to as the dual-process account has been suggested as an alternative to strictly asymmetrically controlled fixation durations in scene viewing (Glaholt et al., 2013; Walshe & Nuthmann, 2014). Glaholt et al. (2013) showed direct-control effects on fixation durations that depended on whether the scene

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. 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. In addition, 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. They argued that the removal of high spatial frequencies induces greater challenges to scene encoding processes than does the removal of low spatial frequencies. 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. 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 (Fig. 3b). The mean fixation duration was significantly longer in the colour condition than in the no-change condition. 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

Fig. 4 Empirical fixation-duration distributions and mean fixation durations replotted from Walshe and Nuthmann (2014). a. Fixation-duration distributions for the three luminance-change conditions in Experiments 1 (left) and Experiment 2 (right). For the duration of a

critical fixation, the luminance of the scene was increased or decreased by a margin of 20% in Experiment 1, and 40% in Experiment 2. b. Mean fixation durations for Experiment 1 (light grey) and Experiment 2 (dark grey)

the distribution are the result of a *surprise* effect. The surprise effect is fast acting, modifies even the shortest fixation durations, 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 increased 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 and Nuthmann (2014) to explain the distributional effects observed in their luminance-change study (shown in Fig. 4a). Like Glaholt et al. (2013), they found that fixation durations tended to increase even when the scene was made easier to process, and that 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. On every 6th saccade, the base scene was replaced by one of four clarified (i.e., less strongly filtered) stimuli, or the original unfiltered scene. In the unfiltered scene condition, fixation durations *decreased* (by about 90 ms) relative to fixation durations on the baseline stimuli (Fig. 3c). In another experiment, participants were presented with base stimuli that consisted of original unfiltered scenes. In this experiment, spatial frequencies were *removed* rather than *added*. Consistent with other studies, 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 and Nuthmann (2014) are partially at odds. Using a similar experimental design, Henderson et al. (2014) observed symmetry while Walshe and Nuthmann (2014) observed asymmetry. However, the dual-process 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 fixation durations may be observed. Unfortunately, Henderson et al. (2014) did not present distributional analyses, which are required to confirm this prediction from the dual-process account.

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 CRISP (Nuthmann et al., 2010) and 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 assumptions explicit by simulating distributions of fixation durations and comparing them against empirical observations. In what follows, we introduce model mechanisms both in an informal and in a formal manner. We then use the model to simulate the data from two luminance-change experiments reported in Walshe and Nuthmann (2014).

A Dual-Process Model of Fixation-Duration Control during Scene Viewing

Overview

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; Engbert et al., 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). Thus, the durations required to initiate and program a saccade are partially determined by the inherent unpredictability of the random walk. A brief episode of simulations with the model is shown in Fig. 5.

In the model, we explored assumptions made by the dual-process account by implementing two distinct influences by which the timing of fixations may be modified on a moment-to-moment basis (i.e., surprise and encoding modulation). The dual-process assumptions enter the model at the level of both saccade timing and saccade programming (labile and nonlabile). In particular, both surprise and encoding related signals can inhibit the rate at which saccade programs are initiated and completed (Fig. 5). We also leave open the possibility for encoding related facilitation (Henderson et al., 2014). Below, we first summarize important model principles that contribute to generating human-like fixation durations. Next, we present a formal description of the mathematical approach used to simulate the model.

Model Components

Rhythmic saccade timer

Like other models of eye-movement control in high-level tasks (Engbert et al., 2005; Nuthmann et al., 2010;

Fig. 5 Simulation of parallel random walks for saccade timing and different phases of saccade programming. The figure shows part of a simulated sequence of fixation durations for a trial in which the luminance of the scene was reduced from 60% to 20% during a selected critical fixation (Experiment 2). Labels on the top of the figure and the regions shaded with light grey identify successive fixations, including one critical fixation. Labels on the right identify the five random walks. The shaded regions above the x-axis containing Ts and Ls specify the portion of the sequence that the timer (T) and labile saccade

programming (L) are active. The red dashed line shows how completion of the timer interrupts an ongoing labile saccade program. The random walk intervals coloured orange and blue are associated with surprise and encoding modulation, respectively. In the example, there was no labile saccade program active at the beginning of the critical fixation, which is why the surprise inhibition was only applied to the random walk timer. Fixation durations are the time intervals between successive saccades, without including the duration of saccade execution (white bars)

Schad & Engbert, 2012; Trukenbrod & Engbert, 2014), the present model includes a process that simulates a random timer. The timer is continuously active and is independent of the saccade programming components responsible for generating a saccadic eye movement. Instead, the random timer is responsible for initiating the saccade programs that ultimately result in eye movements. The continuous activity of the random timer means that whenever it reaches threshold it is reset to the initial state and a new timer process is initiated (Fig. 5).

Random timing is partially motivated by the fact that eye movements occur with a regular frequency and appear to have no observable conditions in which they drastically cease to occur (Lange et al., 2018). More recently, it has been proposed that the generation of eye movements is closely linked to how attention networks function. Hogendoorn (2016) has shown evidence that the initiation of saccades tends to occur at specific points within the period of oscillating attention waves that have been studied during visual perception tasks (Fiebelkorn & Kastner, 2019).

Saccade programming

Multiple stages Saccade programming is completed in multiple distinct stages, whereby each stage is implemented as an independent random walk towards threshold. The multi-stage saccade programming assumption is derived from empirical results obtained with the double-step paradigm, which have been shown to generalize from simple arrangements (Becker & Jürgens, 1979; Ludwig et al., 2007) to a scene-viewing context (Walshe & Nuthmann, 2013, 2015). Following the nomenclature used in existing models of eye-movement control, the two main stages of saccade programming are referred to as labile and nonlabile stages (see above). 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 (see red dashed line in Fig. 5). In this case, a new labile program is activated which replaces the

ongoing labile program. Once the labile stage finishes, the random walk for the nonlabile stage begins. An additional random walk process is used to model the latency required to instruct the oculomotor muscles to physically adjust the position of the eyes. We refer to this random walk as the Motor Command random walk (or “motor” for brevity). Once the motor stage finishes, a final random walk is initiated that simulates the duration of saccade execution; that is, the time required for making a physical movement of the eyes from one location to another. The total duration between the end of the previous saccade execution stage and the beginning of the subsequent saccade execution represents the measure of fixation duration. Figure 5 shows a visualization of the five different random walks and simulated fixation durations in the model.

Parallel programming of saccades Another important finding from the double-step paradigm concerns sequences of two saccades that are made in response to the two target steps (Becker & Jürgens, 1979; Caspi et al., 2004; McPeck et al., 2000). A saccade that has not yet passed the point-of-no-return often shows an intermediate landing position between the first and second target position (see above). Typically, this is followed by a corrective saccade towards the final target position. Importantly, these two saccades tend to be separated by a short intersaccadic interval (< 100 ms). This intersaccadic interval is well below the time required to program a new eye movement (175–250 ms) (Salthouse & Ellis, 1980) and decreases linearly with the amount of time that is available to prepare the corrective saccade in advance (Camalier et al., 2007; McPeck et al., 2000). These results were taken to suggest that the second saccade plan started before the first saccade plan was executed (Becker & Jürgens, 1979; McPeck et al., 2000). In the literature, such overlapping programming of saccades has been referred to as “parallel programming” (Becker & Jürgens, 1979) or “concurrent processing” (McPeck et al., 2000). 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 et al., 2016), and (b) images of naturalistic scenes (Wu et al., 2013). In the modelling approach used here, parallel programming occurs when the saccade timer initiates a labile saccade program while (a) a nonlabile program (Nuthmann et al., 2010; Saez de Urabain et al., 2017) or (b) a motor command is currently active.

Saccade timer and saccade programming rate modulations

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 occurs immediately following the onset of the fixation. Moreover, the dual-process account predicts encoding related effects of direct control, which arise when difficulties in stimulus processing are encountered. However, while surprise is fast acting, encoding modulation occurs only towards the later stages of stimulus processing. Therefore, only relatively long fixation durations will be subject to this influence.

Previous models of eye-movement timing in high-level tasks have implemented rate modulation on the timer (Nuthmann et al., 2010) or both the timer and the labile stage of saccade programming (Trukenbrod & Engbert, 2014). In these models, rate changes reflect processing demands encountered during a given fixation. In our model, surprise and encoding modulation is applied to both the timer and the labile and nonlabile stages of saccade programming. Extending the rate adjustment to nonlabile saccade programming is motivated by results from double-step studies, which suggest that the duration of the nonlabile stage is sensitive to the type of stimulus and task being performed (Ludwig et al., 2007; Walshe & Nuthmann, 2015).

Formal Model Description

Next, we develop the mathematical formulation of the model architecture. The model shares core features with other approaches to modelling fixation durations during high-level tasks. 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). Below, we first describe the Markov process for the case where saccade timing is the only process modelled as a random walk (Nuthmann et al., 2010). Later, we extend the model description for the case in which multiple Markov processes account for stochastic variability in both the initiation, preparation and execution of saccade programs.

Single random walk (timer-only)

A random walk in state m at time t is given by $S_m(t)$, with the 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 over continuous intervals of time. 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 waiting time τ is defined as the time interval to the next transition. For a discrete-state continuous-time Markov process with a constant transition probability rate from state m to state $m + 1$, the probability distribution defined over values of τ is given by the exponential distribution

$$\rho(\tau) = w_1 e^{-w_1 \tau}. \quad (1)$$

The mean waiting time for a single timer step is related to the transition probability rate of the random walk through the equation

$$w_1 = \frac{N_{timer}}{T_{timer}}, \quad (2)$$

where w_1 governs how quickly the random walk transitions between adjacent states.

Specific realizations of τ may be obtained for use in simulations by applying the function

$$\tau = -\frac{1}{w_1} \log(1 - \epsilon), \quad (3)$$

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

Multiple random walks

To fully simulate the mechanisms governing the timing of eye movements, we generalize the timer-only case described above by simulating multiple random walks that approach a bound in parallel (Gillespie, 1978). We deviate from previous implementations (Schad & Engbert, 2012; Trukenbrod & Engbert, 2014) by modelling five, rather than four, independent one-step Markov processes to account for stochastic variability in the initiation, programming and execution of saccadic eye movements.

Compared with other models, the nonlabile stage of saccade programming is split into two components in the present model: a nonlabile component and a motor command component. The nonlabile component represents a portion of saccade programming during which a pending saccade can no longer be cancelled. Therefore, once a saccade program becomes nonlabile, a saccade will imminently occur. The motor component represents the approximate amount of time required to transmit the command to move the eyes from the brain to the eye (Becker, 1989, 1991). This conceptual distinction is important for the rate modulation at the level of saccade programming. While rate modulation is allowed during the nonlabile stage, it ceases to be active as soon as the motor command is sent.

The composite description of the state of the model at time t is given by the vector $S_m(t) = (m_{timer}, m_{labile}, m_{nonlabile}, m_{motor}, m_{saccade})$. N_i is the number of steps in the i^{th} random walk where i is an index for the five random walks with $i \in (\text{timer, labile, nonlabile, motor and saccade})$ and $m_i \leq N_i$. For reasons of model parsimony, we chose to use the same number of states for all random walks, in which case N_i could simply be replaced by N . However, we leave the subscript in the model description to allow the specification of a different number of states per random walk (Trukenbrod & Engbert, 2014). The overall dynamic state of the random walk is determined by the states of each of the elementary random walks. Transitions between states in the model are identified by the notation S_{nm} which describes a single elementary state change $m_i + 1$, for some $i \in 1 \dots 5$, to the adjoining state n . In effect, a state change occurs when an elementary random walk accumulates by one step towards its threshold value N_i . In the event that a random walk reaches threshold, a subsequent random walk process ($i + 1$) is activated and the current random walk is reset to the initial state. A special case exists in the event of a saccade cancellation. To implement saccade cancellation, we assume that when the timer reaches threshold and a labile random walk is active at the same time, $m_{timer} = N_{timer}$ and $0 < m_{labile} < N_{labile}$, then the current labile random walk is reset to the initial state. Additionally, a new labile random walk is immediately activated.

Given that multiple random walks can be activated in parallel, the transition probability rate must be modified accordingly. Generalizing Eq. 2, the transition probability rate for an active 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 the i^{th} random walk. $w_i(t)$ is equal to 0 if the random walk is not active. 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 at time t is given by

$$W = \sum_i w_i(t). \quad (5)$$

The waiting time distribution when multiple random walks are activated is equal to

$$\rho(\tau) = W e^{-W \tau}. \quad (6)$$

As in the case of the single active random walk (timer-only), a specific realization of τ can be obtained from Eq. 3 by replacing w_1 with W . Once a waiting time has been sampled, a single elementary transition must be selected (Gillespie, 1978). This is done by sampling a single

elementary transition according to the relative transition probabilities given by

$$p_i = \frac{w_i}{W}. \quad (7)$$

For a more extensive and general derivation of the multiple random walk approach, the reader is referred to Trukenbrod and Engbert (2014).

Rate modulation

In the model, the transition probability rates determine how quickly each of the elementary random walks approaches threshold. Therefore, decreasing or increasing the transition probability during some time interval will slow down or speed up the random walk during that interval. We assume that the base rate of the elementary random walks can be modified in a time-dependent manner depending on the visual system's response to changes in the environment.

We model two distinct forms of random walk modulation that occur due to scene luminance changes. Surprise modulation begins immediately after the onset of a critical fixation and has a limited duration. Moreover, surprise is purely inhibitory. Surprise modulation can be expressed as follows:

$$w_i(t) = \begin{cases} S_U w_i(t) & \text{if } t_0 < t < \beta_S \text{ where} \\ & 0 < S_U \leq 1 \text{ and } i \in \{timer, lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases} \quad (8)$$

for the luminance increase and,

$$w_i(t) = \begin{cases} S_D w_i(t) & \text{if } t_0 < t < \beta_S \text{ where} \\ & 0 < S_D \leq 1 \text{ and } i \in \{timer, lab, nlab\} \\ w_i(t) & \text{otherwise} \end{cases} \quad (9)$$

for luminance decrease. In the simulations, t_0 is set to the beginning of a critical fixation. The parameter β_S specifies how long the modulations last following the onset of the critical fixation. We assume that the surprise interval β_S is the same for luminance increases (up-changes) and decreases (down-changes). S_U and S_D reflect the strength of the surprise modulation for luminance increases (S_U) and decreases (S_D). Both values are forced to less than 1, which results in a time-dependent slow-down in the rate of the random walk.

Encoding modulation begins with delay β_E after the surprise modulation has finished. Relative to the beginning of the critical fixation, the onset time for the encoding modulation is thus defined as $\beta_S + \beta_E$. Any timer, labile or nonlabile random walks that are active at this time will be subjected to encoding rate modulation for the remaining duration of the critical fixation. More precisely,

the encoding modulation stops at time t_{end} , which marks the end of the scene quality manipulation. In the experiments, the scene luminance manipulation ended during the saccade following the critical fixation. For simplicity, in the model simulations the encoding modulation was applied until the end of the saccade. Of course, no encoding modulation is applied if the critical fixation ends before the encoding modulation interval begins. Encoding modulation was applied to down-changes in scene luminance only where it was implemented as an inhibitory influence by forcing E_D to be smaller than 1. In summary, the random walk rates during the encoding interval are given by

$$w_i(t) = \begin{cases} E_D w_i(t) & \text{if } \beta_S + \beta_E < t < t_{end} \text{ where} \\ & 0 < E_D \leq 1 \text{ and } i \in \{timer, lab, nlab\} \\ w_i(t) & \text{otherwise.} \end{cases} \quad (10)$$

For both surprise and encoding modulations, the total strength of modulation that occurs within a given critical fixation is determined both by the strength of the modulation and the total time that the modulation is active during a critical fixation.

Parameter Estimation

For a given experiment, the model fit was obtained by first fitting parameters to the baseline condition (no luminance change) and then holding all parameters fixed, while allowing the rate adjustment parameters to vary to best fit the empirical data. For the reported simulations, we assumed that the rates for the encoding modulation could only be inhibited, not sped up. Thus, there was no encoding related facilitation in conditions where the quality of the scene was enhanced by increasing its luminance. Rather, encoding modulations were simply not activated in this case. We found that these rate adjustment mechanisms were sufficient to explain the pattern of results observed in the experimental data. Table 1 in Appendix A.1 provides the full set of fixed and free parameters, along with their fixed or best-fitting values. Details on the model fitting procedure are provided in Appendix A.2.

Model Simulations

The main goal of the model simulations was to test the hypothesis that the empirical distributions of fixation durations observed in luminance-change experiments (Walshe & Nuthmann, 2014), and in similar experiments conducted in other laboratories (Glaholt et al., 2013; Henderson et al.,

2014), can be explained by a conceptually simple mechanism of adaptable adjustments to the timer and saccade programming rates. Specifically, the model was tested on data from two saccade-contingent scene luminance change experiments reported by Walshe and 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 image luminance was increased or decreased by a margin of 20% in Experiment 1 and 40% in Experiment 2.

For a given experiment, the best-fit model was used to generate simulated data for 10,000 independent trials. In the simulations, each trial began by setting the initial states of the random walks to 0. Then, the model was allowed to run uninterrupted until a total of six critical fixations were made. A critical fixation occurred on every 5th fixation, and the order of the experimental conditions (luminance increase, luminance decrease and baseline) was randomised within a trial.

Figure 6 compares the fixation-duration distributions derived from the simulated data against the empirical (human) fixation-duration distributions. The distributions were constructed by sorting fixation durations into 20 equally spaced bins of 60 ms. The data are plotted at the center of each bin. Fixation durations beyond 1,200 ms were filtered from the simulations to match the upper range of

fixation durations reported in Walshe and Nuthmann (2014). Fixation durations below 50 ms were excluded from the empirical data, but were included in the model simulations. Figure 8 shows the corresponding cumulative distributions (fitted model in green and empirical data in black). The results indicate that the simulated data provide a good fit to the human data.

Below, we present additional analyses exploring the model's behaviour. First, we investigate the role saccade cancellations play in generating long fixation durations in response to changes in scene quality. Next, we highlight how fixation durations generated by the model are influenced by saccade preprogramming. Finally, we use a counterfactual method to analyse how specific model components impact the fixation-duration distributions.

The Effect of Saccade Cancellation on Fixation Durations

The distributions of fixation durations observed during visual-cognitive tasks, including scene viewing, tend to be heavy-tailed. In the model, the heavy tail arises primarily through saccade cancellations. Recall that a cancelled saccade occurs when there is a labile saccade program in progress that is reset by the completion of the timer (see Fig. 5 for an example of saccade cancellation).

Fig. 6 Comparison of model and human fixation-duration distributions. The first row shows the data for Experiment 1 where the luminance of the scene was increased or decreased by a margin of 20%. The second row shows the data for Experiment 2 in which a margin of 40% was used. Each column shows the data for one of the luminance-change conditions (baseline: no change). The red line

shows the human fixation-duration distributions (see also Fig. 4a). The stacked bars show the best-fit model predictions. Model fixation durations were grouped by the number of saccade cancellations within the fixation and then a histogram of fixation durations for each group was plotted

Figure 6 compares luminance conditions for the two different experiments. The red line shows the empirical human distribution, and the histogram shows the proportion of fixation durations observed in the best-fitting model. For the model simulations, fixation durations were grouped by the number of cancellations that occurred within the fixation. Cancellations beyond four were not included as they occurred rarely. The frequency observed in each bin was normalized by the total number of fixations to create a probability distribution.

Figure 6 shows that fixation durations beyond 500 ms are almost entirely composed of fixations that included at least one cancellation. Moreover, the heavy-tailed property of the fixation-duration distribution is primarily produced by fixations that include multiple cancellations. Interestingly, the effect of cancellations on the distributions is not the same for all luminance conditions. In the luminance decrease conditions, the cancellation distributions are shifted towards longer fixation durations relative to the baseline and luminance increase conditions. The effect is stronger in Experiment 2 than in Experiment 1. These results suggest that surprise and encoding effects interact with saccade cancellation to shape the fixation-duration distribution. We address the relationship between saccade cancellation and rate modulation in “Counterfactual Analysis of Timer and Saccade Programming Rate Modulations” below.

Figure 7a further highlights the role that cancelled saccades play in generating long fixation durations. As before, fixation durations were grouped by the number of cancellations occurring within the fixation. Moreover, we identified the duration of the first labile saccade program

occurring during the critical fixation. Accordingly, fixation durations are plotted as a function of the duration of the labile stage of saccade programming (x -axis) and the number of cancellations (colour). For the visualization, we uniformly sampled a subset of 2,000 trials from the full set of 10,000 simulated trials. This was done to avoid overplotting and clearly show the shape of the bivariate distributions. Figure 7a shows that fixation durations increase as a function of both the total number of saccade cancellations and the duration of the labile stage of saccade programming.

The Influence Of Saccade Preprogramming on Fixation Durations

Given the implementation of saccade timing and programming in the model, different situations can occur at the start of a critical fixation. Oftentimes, the program for the next saccade is only started sometime during the critical fixation (see Fig. 5 for an example). In this case, the critical fixation’s duration is under direct control of the current visual scene. However, the random timer will also generate instances in which the program for the next saccade was already initiated prior to the onset of the critical fixation. We refer to this situation as some form of preprogramming but note that it is different from the parallel programming of saccades described in “Parallel programming of saccades” above. With regard to the onset of a critical fixation, different subcases can be distinguished. If saccade planning has already advanced to the motor command or to the saccade execution stage, the duration of the critical fixation will

Fig. 7 a. Fixation durations as a function of labile saccade programming duration and saccade cancellations. The x -axis refers to the amount of time that saccade programming spends in the labile stage. There is an approximately linear relationship between the duration of labile saccade programming and the eventual fixation duration. An increased number of cancellations results in longer fixation durations.

b. Saccade preprogramming. At the onset of a critical fixation, a labile saccade program may already be activated. The proportion of labile saccade programming completed at the onset of a critical fixation was binned into three intervals (0–0.33, 0.34–0.67 and 0.68–1) and coloured blue, orange and purple. Preprogramming of the labile stage is associated with shorter fixation durations

not be affected by the change in scene quality (i.e., luminance). If, however, saccade planning has only advanced to the labile or nonlabile stages, the corresponding random walks will be subject to the rate modulation.

In these cases, varying progress has been made on preparing the next saccade before the start of the current fixation. We explored this behaviour by analysing a subset of simulated critical fixations from Experiment 1 where a labile saccade program was already underway at the onset of fixation. An additional requirement was that the labile program was not cancelled before completion. For each of these cases, we identified the proportion of total labile programming that had already been completed before the critical fixation began. Based on this proportion, each fixation was placed into one of three equally spaced bins (0–0.33, 0.34–0.67 or 0.68–1). Figure 7b depicts individual fixation durations as a function of the time spent in the labile stage of saccade programming. The different completion categories are represented by different colours. As the proportion of labile programming that overlaps with the previous fixation increases, shorter durations are observed for the current fixation. Fixation durations are shortest for cases in which the majority of the labile stage of saccade programming was already completed at the onset of the fixation. Furthermore, both Fig. 7a and b show a positive correlation between the overall duration of the labile stage of saccade programming and the eventual fixation duration (see also Nuthmann et al., 2010).

Counterfactual Analysis of Timer and Saccade Programming Rate Modulations

To measure the effect that rate adjustments have on the simulated fixation-duration distributions, we employed a

counterfactual analysis approach. The underlying logic involves comparing two models that are identical in all aspects except for one critical feature (e.g., surprise). Therefore, any difference between the distributions generated by the two models must arise from that one model component. Thus, the counterfactual analysis allows us to investigate the causal role that timer and saccade programming rate adjustments play in shaping the simulated fixation-duration distributions.

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 took the best-fit model (we refer to this as the full model) for each luminance-change condition and set one of the rate modulation parameters to 1. We then simulated trials exactly as we did for the full model. The counterfactual distributions were then compared with the best-fit distributions.

The cumulative probability distributions presented in Fig. 8 allow us to identify the approximate point in time at which the fixation-duration distributions generated by different models diverge. To explore how surprise and encoding related inhibition impacts the shape of the fixation-duration distribution, we additionally computed the difference between the probability distribution obtained for a given counterfactual model and the full model; that is, $Pr_{diff} = Pr_{counterfactual} - Pr_{full}$ (Fig. 9). Note that we use these two measures as qualitative tools only. A statistical divergence point analysis (Gómez et al., 2021, for critical discussion) and a statistical evaluation of the difference scores are beyond the scope of this article.

Fig. 8 Cumulative probability for human and model fixation-duration distributions. Distributions are plotted for each condition (columns) and each experiment (rows). Black lines are human distributions. Green lines are fixation-duration distributions generated by the best-fit (or full) model. For the luminance increase and decrease conditions, additional coloured lines represent the surprised reduced model (orange) and the encoding reduced model (blue), which were compared to the full model via counterfactual comparison

Surprise reduced model

We first studied the effect that removal of the surprise component had on predicted fixation-duration distributions. Figure 8 compares the cumulative distributions for a surprise reduced model (orange) with a best-fit model with all parameters (green). The divergence point for each pair of distributions is the earliest point when the surprise modulation begins to impact fixation durations. For both experiments and both luminance-change conditions, removing surprise shifted the cumulative distribution towards short fixation durations relative to the full model. Moreover, the surprise reduced model diverges from the best-fit model as early as the second bin (centered at 90 ms) of the cumulative distribution. Thus, the effect of removing surprise impacts even relatively short fixation durations. Comparing the luminance decrease and increase conditions shows that the early surprise effect is active both in the presence and absence of encoding modulation (recall that encoding modulation is only active for luminance decreases). This demonstrates that the early onset of surprise modulation operates at least partially independently from encoding modulation.

Figure 9 additionally highlights how surprise impacts the shape of the fixation-duration distribution. In particular, the effect of removing surprise is evident in the displacement of the orange line from the value of 0. Values above and below 0 indicate that the surprise reduced model predicts a higher or lower frequency of fixation durations in that bin. For each condition, we found a positive difference for relatively short fixation durations followed by a negative difference after the early peak. This result complements what was

found in the cumulative distributions; the surprise reduced model increases the proportion of short fixation durations at the expense of decreasing the proportion of longer fixation durations.

Encoding reduced model

Next, we analysed the impact that removal of the encoding modulation has on the fixation-duration distribution. For a given experiment and luminance-change condition, Fig. 8 allows us to directly compare the cumulative distributions for the encoding reduced model (blue) and the full model (green). Recall that encoding modulation is not active during luminance increases, which is why there is no noticeable difference between the blue and green curves in the luminance increase conditions. For the luminance decrease conditions, however, removing encoding modulation decreases the proportion of very long fixation durations in the tail of the distribution. Furthermore, the impact on the tail of the distribution is more substantial in Experiment 2 than in Experiment 1. Moreover, the two distributions diverge earlier in Experiment 2 (at approximately the 7th bin centered at 390 ms) than in Experiment 1 (at approximately the 9th bin centered at 510 ms).

Figure 9 shows in more detail how encoding related inhibition influences the tail of the fixation-duration distribution in the luminance decrease conditions (right panels, blue line). For both experiments, fixation durations in the tail of the distribution shift to moderately lower fixation durations (approximately 390 ms in Experiment 2; approximately 510 ms in Experiment 1). The timing for

Fig. 9 Counterfactual analysis. Distributions of difference scores are plotted for the luminance increase (left) and luminance decrease (right) conditions in a given experiment (row). The lines in each panel are constructed by subtracting the value of the probability distribution for the full model from one of the counterfactual models. The orange and blue lines show the subtraction for the surprise reduced and encoding reduced models, respectively. A value of 0 indicates that the two probability distributions are equivalent at that fixation duration. Values above or below 0 show that the counterfactual model predicts a higher or lower probability for this fixation-duration bin

the onset of the encoding modulation is consistent with the fact that the best-fit model selected an earlier onset for encoding modulation in Experiment 2 than in Experiment 1 (see Table 1 in Appendix A.1). In summary, the results from the counterfactual analysis on encoding modulation show that this mechanism primarily impacts the tail of the distributions. It operates distinctly from surprise modulation in that short fixation durations are left unaffected.

In “The Effect of Saccade Cancellation on Fixation Durations” above, we reported that long fixation durations located in the tail of the distribution tend to be predominantly composed of fixations that include at least one cancellation (see Fig. 6). This observation implies that encoding modulation, which primarily influences the tail of the distribution, will have the most significant impact on fixations when cancellations occur.

General Discussion

Evidence for Surprise and Encoding Influences on Fixation Timing in Naturalistic Scene Viewing

It has long been known that fixation durations in visual-cognitive tasks vary with processing difficulty (Rayner, 2009, for review). In the context of scene viewing, a number of studies have recently investigated whether fixation durations are under the direct control of the quality of the current scene image. To this end, a gaze-contingent display change technique was used to manipulate the quality of the scene image during selected critical fixations (Fig. 1a). When the scene was reduced in quality via a decrease in luminance (Henderson et al., 2013; Walshe & Nuthmann, 2014) or by filtering spatial frequencies (Glaholt et al., 2013; Henderson et al., 2014), individual fixation durations were found to increase. Glaholt et al. (2013) were the first to report distributional analyses of fixation durations that uncovered a pattern of two distinct distributional effects: an influence on the central tendency and an influence on the tail. On the one hand, both low- and high-pass filtering of the scene image resulted in a shift in the central tendency of the distributions. On the other hand, the lengthening of fixation durations was stronger for low-pass than for high-pass filtered scene stimuli, which showed as an influence on the tail of the distribution. In a different study, saccade-contingent reductions in scene luminance led to both an early shift of the distribution and a late influence on the tail (Walshe & Nuthmann, 2014).

Some of these experiments also investigated the symmetry or directionality in the way fixation durations are adjusted. In case of a symmetric (or bidirectional)

adjustment to the ease of processing, enhancing the quality of the scene should lead to a shortening of fixation durations. In case of an asymmetric (or unidirectional) adjustment, no change in fixation durations should be observed. Contrary to these predictions, increased fixation durations were found in experiments where the quality of the scene was enhanced by either increasing the luminance of the scene to its normal level (Walshe & Nuthmann, 2014) or by adding colour to a grey-scale scene (Glaholt et al., 2013). Distributional analysis revealed that this increase came from a shift in the central tendency of the distribution.

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.

Modulation of Saccade Timing and Programming

A novel assumption incorporated in the simulations is the introduction of surprise and encoding modulation components that modify saccade timing and programming random walk rates at different time scales. Using the counterfactual method, we found that the surprise mechanism impacts fixation durations of all lengths and can influence even the shortest fixation durations in the distribution. In other words, the effect of surprise is consistent with a general shift in the distribution towards longer fixation durations. Unlike surprise, encoding modulation acts by increasing the duration of relatively long fixation durations located in the tail of the distribution. Thus, our counterfactual analysis demonstrates that these two hypothetical mechanisms have distinctive impacts on the shape of fixation-duration distributions.

The computational 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 random walks. Logically, this is a reasonable model assumption to make in that scene changes which make the stimulus easier to process may result in faster saccade timing and programming. 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.

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 timing and programming signals. At the same time, the model has substantial overlap with previous models that share the assumption of a random-walk saccade timer (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) the triggering of saccade programs by a rhythmic timer, and (c) multi-stage saccade programming architectures.

A specific question we want to address is whether the CRISP model (Nuthmann et al., 2010) would be able to reproduce the fixation-duration distributions observed in experiments with the scene quality change paradigm. A key assumption of the model is that moment-to-moment difficulties in visual and cognitive processing can immediately inhibit saccade initiation. When testing the model on SOD data, delaying the onset of the stimulus (cf. Figure 1b) was associated with a considerable slowdown of the random walk timer (Nuthmann & Henderson, 2012; Nuthmann et al., 2010). This alone would not prolong the duration of the critical fixation if there is a labile saccade program active at the time the scene disappears. However, in the model simulations some of these labile programs were cancelled in response to the scene disappearing from view. These two mechanisms were sufficient to generate the empirically observed data pattern.

Formalizing the dual-process account to explain fixation-duration distributions from scene quality change

experiments would require extensions to the CRISP model. Both the present model and the CRISP model conceptualize the timer as a random walk process. The random walk creates a trajectory over time, which can be modulated by visual-cognitive events at any point. While surprise inhibition and encoding modulation can easily be applied to random walk timing signals in CRISP, implementing similar modifications to saccade programming components is less straightforward. Processing-related saccade cancellation, used for the SOD simulations (Nuthmann et al., 2010), is not a plausible mechanism in case of stimulus enhancements, or mild degradations. Therefore, for preliminary (unpublished) simulations using data from a saccade-contingent scene luminance reduction experiment (Henderson et al., 2013, Exp. 1) a different implementation was used. If a labile program was active at the beginning of a critical fixation, its duration was prolonged by an amount reflecting the extent of the applied degradation. With this modification, CRISP was able to reproduce the empirical fixation-duration distributions reasonably well.

However, the dual-process account makes a theoretically important distinction between early surprise and late encoding related inhibition (Glaholt et al., 2013; Walshe & Nuthmann, 2014). In the CRISP model, it is not possible to modify saccade programs on a moment-to-moment basis, because the durations of the different phases are drawn from gamma distributions. Therefore, in the new model we implemented all stages of saccade programming as independent, parallel discrete random walk processes (Gillespie, 1978). Given that this kind of stochastic simulation framework has also been used in other random-timer models (Schad & Engbert, 2012; Trukenbrod & Engbert, 2014), we chose not to introduce the model presented here as CRISP 2, but as an unnamed computational model (UCM) instead. 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 LATEST model (Tatler et al., 2017) belongs to a different class of models. In LATEST, 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 the effects of scene luminance adjustments observed in our task. Decreasing the luminance of the scene would decrease information supply at both foveal and peripheral locations, resulting in a reduced rate of rise of the Stay-or-Go decision signal. This would result in the observed increase in fixation durations. The surprise effect observed

in the present study could result from either an increased Stay signal (the unexpected foveal information resulting in an increased benefit for prolonged foveal processing) or a shift in the decision criterion.

Predictive Model of Surprise Modulation

A question raised by our study is the degree to which the surprise modulation effects implemented in our simulations are related to models of predictive perception and action. For example, in studies of saccadic reaction times, it has been shown that saccade latencies are sensitive to the probability that a target is located at the endpoint of an upcoming saccade (Carpenter & Williams, 1995). It has also been shown that saccadic reaction times increase with the degree to which the location of a target violates prior expectations (Vossel et al., 2014). These studies suggest that eye-movement control is at least partially susceptible to what the visual system expects to encounter on a future fixation. More generally, it has been argued that eye-movement control is tightly linked to predicted outcomes based on scene context derived from a viewer's knowledge about the environment (Henderson, 2017). Therefore, one interesting possibility is that the large-scale stimulus changes in gaze-contingent scene quality studies induce a violation of expectations between the pre- and post-saccadic fixations. In relation to the account we have introduced, the surprise mechanism that inhibits timer and saccade programming rates could be interpreted more generally in terms of a mismatch resulting from perceived and predicted expectations.

Additionally, in predictive coding accounts of perceptual processing, it is claimed that perceptual circuits convey signals related to the difference between anticipated and observed inputs (Huang & Rao, 2011). Predictive signals have also been suggested to act as a method of smoothing errors and deemphasizing perceptual samples arising due to chance fluctuations (Burr & Cicchini, 2014). In the present context, the large-scale visual change in the image structure encountered at the onset of a critical fixation may result in the generation of such an error signal. This error signal could be used by eye-movement planning areas of the brain such as the frontal eye fields or the superior colliculus to slow down the timing and execution of saccadic eye movements when surprising visual events are encountered.

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.

Appendix

A.1 Model Parameters

Table 1 provides the parameters for the models that best fit the human data in Experiment 1 and Experiment 2. The parameter listed under the heading Timer and Programming Parameters refer to the base parameters required to simulate fixation durations in the baseline condition where no luminance change was made on a critical fixation. The heading Modulation Parameters specifies the parameters used to define the strength and duration of the rate modulation. The column titled Equation Symbol can be used to cross-reference the parameters in Table 1 with the description of the model given in the section on the formal derivation of the model.

The parameters listed as Fixed were not free to vary during the model fitting procedure. Saccade duration was fixed at 20 ms. This value is within the plausible range given by Devillez et al. (2020), although we note that it is at the low end of the range for many typical scene-viewing tasks they studied. The Motor Command parameter reflects the fact that there is a transmission lag between the time that the eye-movement system generates a command to move the eye and the time that the signal arrives and a saccade is made. In the simulations, the motor parameter was fixed at 30 ms. The exact value of the lag is not known, but 30 ms is within the plausible range given by neurophysiological and psychophysical estimates (Becker, 1989, 1991).

A.2 Model Fitting Via Maximum Likelihood Estimation

We used a maximum likelihood based method of assessing how well the parameters of the model fit the fixation durations from the empirical data sets. The metric for goodness of fit that was used was defined as the likelihood $\ell(\theta; x_1, \dots, x_n)$ of observing the empirical data $x_1 \dots x_n$ given the parameters of the candidate model θ . Other objective functions were tested with very similar results. We optimized this function using Bayesian adaptive direct search (Acerbi & Ji, 2017).

More concretely, we selected model parameters that minimized the negative of the log likelihood of the

Table 1 Parameter values for best-fit models

Parameter name	Equation symbol	Value		Range
		Experiment 1	Experiment 2	
<i>Timer and Programming Parameters</i>				
Number of Steps	N	14	16	10–30
Timer Duration (ms)	T_{timer}	226.27	230.25	100–300
Labile Duration (ms)	T_{labile}	174.24	192.24	100–250
Nonlabile Duration (ms)	$T_{nonlabile}$	71.60	83.67	50–100
Motor Command (ms)	T_{motor}	30.00	30.00	Fixed
Saccade Execution (ms)	$T_{saccade}$	20.00	20.00	Fixed
<i>Modulation Parameters</i>				
Surprise Offset (ms)	β_S	42.67	60.56	0–100
Surprise (Up)	S_U	.49	.54	.1–1
Surprise (Down)	S_D	.24	.15	.1–1
Encoding Offset (ms)	β_E	387.81	247.19	200–400
Encoding (Down)	E_D	.37	.62	.1–1

Number of Steps, Timer Duration, Nonlabile Duration, Motor Command Duration and Saccade Duration refer to the mean duration of the random walk. Surprise Offset specifies how long the surprise modulation is activated after the onset of a luminance change. Encoding Offset refers to when the encoding modulation begins relative to the value of Surprise Offset. The parameters labelled Surprise (Up) and Surprise (Down) refer to the strength of modulation during the surprise interval and Encoding (Down) refers to the strength of modulation during the encoding interval. Both the duration and the strength of the modulation jointly determine the effect of modulation during a critical fixation

observed fixation-duration distributions conditional on the parameters of the model. Distributions for both the human measurements and models were taken by binning fixation durations into 20 equally spaced bins of 60 ms. The frequency of observations in each bin was then normalized by the total number of fixation durations to ensure that a probability distribution ($\sum_i p_i = 1$) was obtained, where p_i equals the probability of observing a fixation duration in bin i .

The method of maximum likelihood is simplified when the observations are independently and identically distributed, as they are in the present case (DeGroot & Schervish, 2012). Then, the likelihood function takes the following form:

$$\mathcal{L}(\theta; x_1, \dots, x_n) = \prod_{i=1}^k P(x_i|\theta). \quad (11)$$

Since the logarithm is a monotonic function of the likelihood, maximizing the likelihood function is equivalent to maximizing the log-likelihood function:

$$\operatorname{argmax}_{\theta}(\log(\mathcal{L}(\theta; x_1, \dots, x_n))) = \operatorname{argmax}_{\theta}(\sum_{i=1}^k \log(P(x_i|\theta))). \quad (12)$$

Computing $P(x_i|\theta)$ is obtained by

$$\begin{aligned} \log(\mathcal{L}(\theta; x_1, \dots, x_n)) &= \sum_{i=1}^k \log(q_i^{n_i}) \\ &= \sum_{i=1}^k n_i \log(q_i), \end{aligned} \quad (13)$$

where q_i is the probability of the i^{th} fixation duration bin determined by the model with parameters θ , and n_i is the frequency of observing fixation durations in the i^{th} bin of the empirical data. Typically, it is simpler for software to search for a minimum value so we optimize the negative of the log-likelihood:

$$\operatorname{argmin}_{\theta}(-\log(\mathcal{L}(\theta; x_1, \dots, x_n))). \quad (14)$$

Declarations

Acknowledgements We thank Ben Tatler for helpful discussions about the LATEST model.

Funding Open Access funding enabled and organised by Projekt DEAL.

Data and code availability Project code and data are available at https://github.com/calenwalshe/scene_fixation_toolbox.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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). Retrieved from <http://papers.nips.cc/paper/6780-practical-bayesian-optimization-for-model-fitting-with-bayesian-adaptive-direct-search.pdf>.
- Becker, W. (1989). Metrics. In R. H. Wurtz, & M. Goldberg (Eds.) *The neurobiology of saccadic eye movements* (pp. 13–67). Elsevier.
- Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.) *Eye Movements* (pp. 95–137). CRC Press.
- Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19(9), 967–983. [https://doi.org/10.1016/0042-6989\(79\)90222-0](https://doi.org/10.1016/0042-6989(79)90222-0).
- Burr, D., & Cicchini, G. M. (2014). Vision: Efficient adaptive coding. *Current Biology*, 24(22), R1096–R1098. <https://doi.org/10.1016/j.cub.2014.10.002>.
- 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. <https://doi.org/10.1016/j.visres.2007.04.021>.
- 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). 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. <https://doi.org/10.1038/377059a0>.
- 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. <https://doi.org/10.1073/pnas.305329101>.
- DeGroot, M. H., & Schervish, M. J. (2012). *Probability and statistics* (4th ed.). Pearson.
- Devillez, H., Guyader, N., Curran, T., & O'Reilly, R. C. (2020). The bimodality of saccade duration during the exploration of visual scenes. *Visual Cognition*, 28(9), 484–512. <https://doi.org/10.1080/13506285.2020.1830325>.
- 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. [https://doi.org/10.1016/S0042-6989\(01\)00301-7](https://doi.org/10.1016/S0042-6989(01)00301-7).
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777–813. <https://doi.org/10.1037/0033-295X.112.4.777>.
- Fiebelkorn, I. C., & Kastner, S. (2019). A rhythmic theory of attention. *Trends in Cognitive Sciences*, 23(2), 87–101. <https://doi.org/10.1016/j.tics.2018.11.009>.
- Findlay, J. M. (1992). Programming of stimulus-elicited saccadic eye movements. In K. Rayner (Ed.) *Eye movements and visual cognition* (pp. 8–30). Springer. https://doi.org/10.1007/978-1-4612-2852-3_2.
- 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. [https://doi.org/10.1016/0021-9991\(78\)90060-8](https://doi.org/10.1016/0021-9991(78)90060-8).
- 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. <https://doi.org/10.3758/s13414-013-0522-1>.
- 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. <https://doi.org/10.1080/13506285.2014.891542>.
- Gómez, P., Breithaupt, J., Perea, M., & Rouder, J. N. (2021). Are divergence point analyses suitable for response time data?. *Behavior Research Methods*, 53(1), 49–58. <https://doi.org/10.3758/s13428-020-01424-1>.
- Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, 21(1), 15–23. <https://doi.org/10.1016/j.tics.2016.11.003>.
- 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. <https://doi.org/10.1037/a0031224>.
- 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. <https://doi.org/10.1080/13506285.2014.897662>.
- 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. <https://doi.org/10.3758/PBR.15.3.566>.
- Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition*, 17(6–7), 1055–1082. <https://doi.org/10.1080/13506280802685552>.
- Hogendoorn, H. (2016). Voluntary saccadic eye movements ride the attentional rhythm. *Journal of Cognitive Neuroscience*, 28(10), 1625–1635. <https://doi.org/10.1162/jocn.a.00986>.
- Hooge, I. T. C., & Erkelens, C. J. (1998). Adjustment of fixation duration in visual search. *Vision Research*, 38(9), 1295–1302. [https://doi.org/10.1016/S0042-6989\(97\)00287-3](https://doi.org/10.1016/S0042-6989(97)00287-3).
- Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5), 580–593. <https://doi.org/10.1002/wcs.142>.
- Kalesnykas, R. P., & Hallett, P. E. (1987). On plotting amplitude-transition functions for voluntary eye saccades. *Vision Research*, 27(4), 675–679. [https://doi.org/10.1016/0042-6989\(87\)90051-4](https://doi.org/10.1016/0042-6989(87)90051-4).
- 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), 12–35. <https://doi.org/10.1037/0096-3445.135.1.12>.
- Lange, E. B., Pieczykolan, A., Trukenbrod, H. A., & Huestegge, L. (2018). The rhythm of cognition - Effects of an auditory beat on oculomotor control in reading and sequential scan-

- ning. *Journal of Eye Movement Research*, 11(2), Article 9. <https://doi.org/10.16910/jemr.11.2.9>.
- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review*, 104(3), 524–553. <https://doi.org/10.1037/0033-295X.104.3.524>.
- Ludwig, C. J. H., Mildinhall, J. W., & Gilchrist, I. D. (2007). A population coding account for systematic variation in saccadic dead time. *Journal of Neurophysiology*, 97(1), 795–805. <https://doi.org/10.1152/jn.00652.2006>.
- Luke, S. G., Nuthmann, A., & Henderson, J. M. (2013). Eye movement control in scene viewing and reading: Evidence from the stimulus onset delay paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 10–15. <https://doi.org/10.1037/a0030392>.
- 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. <https://doi.org/10.1037/0033-295X.112.4.814>.
- McPeck, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research*, 40(18), 2499–2516. [https://doi.org/10.1016/S0042-6989\(00\)00102-4](https://doi.org/10.1016/S0042-6989(00)00102-4).
- Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 667–682. <https://doi.org/10.1037/0096-1523.10.5.667>.
- Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience and Biobehavioral Reviews*, 64, 229–251. <https://doi.org/10.1016/j.neubiorev.2016.02.018>.
- 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. <https://doi.org/10.3758/s13423-016-1124-4>.
- Nuthmann, A., & Henderson, J. M. (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. <https://doi.org/10.1080/13506285.2012.670142>.
- 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. <https://doi.org/10.1037/a0018924>.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–506. <https://doi.org/10.1080/174702109.02816461>.
- 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. <https://doi.org/10.1037/0033-295X.105.1.125>.
- 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. <https://doi.org/10.1037/a0026473>.
- 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), 445–526. <https://doi.org/10.1017/S0140525X03000104>.
- Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience*, 7(7), 361. <https://doi.org/10.3389/fnhum.2013.00361>.
- Roos, J. C. P., Calandrin, 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. <https://doi.org/10.1007/s00221-008-1304-1>.
- 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. [https://doi.org/10.1016/S0166-2236\(00\)01685-4](https://doi.org/10.1016/S0166-2236(00)01685-4).
- 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. <https://doi.org/10.1016/j.visres.2016.10.015>.
- Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *American Journal of Psychology*, 93(2), 207–234. <https://doi.org/10.2307/1422228>.
- 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. <https://doi.org/10.1080/13506285.2012.670143>.
- Shioiri, S. (1993). Postsaccadic processing of the retinal image during picture scanning. *Perception & Psychophysics*, 53(3), 305–314. <https://doi.org/10.3758/BF03205185>.
- 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. <https://doi.org/10.1037/rev0000054>.
- Trukenbrod, H. A., & Engbert, R. (2014). ICAT: A computational model for the adaptive control of fixation durations. *Psychonomic Bulletin & Review*, 21(4), 907–934. <https://doi.org/10.3758/s13423-013-0575-0>.
- 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. <https://doi.org/10.1093/cercor/bhs418>.
- 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 M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.) *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1569–1574). Berlin: Cognitive Science Society. Retrieved from <https://mindmodeling.org/cogsci2013/papers/0293/>.
- Walshe, R. C., & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research*, 100, 38–46. <https://doi.org/10.1016/j.visres.2014.03.012>.
- Walshe, R. C., & Nuthmann, A. (2015). Mechanisms of saccadic decision making while encoding naturalistic scenes. *Journal of Vision*, 15(5), Article 21. <https://doi.org/10.1167/15.5.21>.
- Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *Archives of Ophthalmology*, 52(6), 932–941. <https://doi.org/10.1001/archophth.1954.00920050938013>.
- Wheless, L. L., Boynton, R. M., & Cohen, G. H. (1966). Eye-movement responses to step and pulse-step stimuli. *Journal of the Optical Society of America*, 56(7), 956–960. <https://doi.org/10.1364/JOSA.56.000956>.
- Wu, E. X. W., Chua, F.-K., & Yen, S.-C. (2016). Saccade plan overlap and cancellation during free viewing. *Vision Research*, 127, 122–131. <https://doi.org/10.1016/j.visres.2016.07.009>.
- 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), Article 17. <https://doi.org/10.1167/13.12.17>.