Journal of Experimental Psychology: Human Perception and Performance

The Time Course of the First Decision in Scene Viewing: Perceptual and Semantic Contributions to Initial Scene Processing

Benjamin W. Tatler and James R. Brockmole

Online First Publication, May 26, 2025. https://dx.doi.org/10.1037/xhp0001333

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

https://doi.org/10.1037/xhp0001333

The Time Course of the First Decision in Scene Viewing: Perceptual and Semantic Contributions to Initial Scene Processing

Benjamin W. Tatler[1] and James R. Brockmole[2]
[1] School of Psychology, University of Aberdeen
[2] Department of Psychology, University of Notre Dame

Effective behavior requires that we respond to the environment rapidly and appropriately. Our visual system can extract a range of information from scenes within tens of milliseconds. The timing of the first eye movement, therefore, may reveal key aspects of the mechanisms underlying rapid extraction and utilization of information during this essential, initial processing, yet we lack theoretical understanding of these mechanisms. Across several data sets, this study shows differences in the timing of the first saccade. Modeling individual data sets, we tested whether the decision mechanism for the first saccade is different from that for subsequent saccades. We found that when viewing started from the scene center, the time course of the first saccadic decision was influenced by low-level features in peripheral vision. However, when viewing started noncentrally, and was aligned more predictably with informative content, low- and high-level information in both foveal and peripheral vision was accumulated and evaluated in the decision process. These findings suggest a common mechanism for initial and later processing of the scene, but with priorities for weighting the input that depend on expectations about foveal content at scene onset. This study extends existing theoretical frameworks for decision making in scene viewing to encompass initial scene processing and expectation-based flexibility in information weighting.

Public Significance Statement
Eye movements are guided by information about a scene that is accumulated over time. Thus, less information is available to guide early eye movements than later ones. How does the visual system overcome this challenge? We argue that early and late eye movements are guided by a single decision-making process that allows for the flexible and strategic prioritization of various types of information depending on their availability at different points in time and in different viewing contexts.

Keywords: scene viewing, saccade, decision, salience, semantics

We live in complex environments where we must adapt and guide our behavior efficiently. This requires that we extract relevant information from our surroundings rapidly and use it to respond appropriately, exploring the scene effectively and adapting our behavior to the current demands of our setting. The initial moments of rapid extraction of information from the scene are essential for scaffolding further scene understanding and exploration. The first eye movement (first saccade) in the scene stems from this initial processing and, thus, may reveal key aspects of its mechanisms and its subsequent influence. However, despite the crucial input it provides, the first saccade is often discarded from analyses of eye-movement behavior when viewing scenes and the development of theoretical frameworks and computational models of where and when to move the eyes (Tatler et al., 2011). Current theoretical frameworks and computational models have instead focused on what happens once scene exploration has begun, describing the roles of low-level visual features (Krasovskaya & MacInnes, 2019), high-level, semantic features (Henderson & Hayes, 2017), top-down task goals (Castelhano et al., 2009; Land & Tatler, 2009), randomness (Clarke et al., 2016), and the time course over which decisions to move the eyes are made (Tatler et al., 2017). There is good reason to question, however, whether the principles of gaze control revealed by such models can be appropriately applied to initial scene processing. Hence, in this report, we take a novel approach to modeling gaze

Nurit Gronau served as action editor.
Benjamin W. Tatler https://orcid.org/0000-0002-5316-766X
 The authors declare that they have no conflicts of interest.

Benjamin W. Tatler served as lead for conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, and writing–original draft. James R. Brockmole served in a supporting role for conceptualization, methodology, and resources. Benjamin W. Tatler and James R. Brockmole contributed equally to writing–review and editing.

Correspondence concerning this article should be addressed to Benjamin W. Tatler, School of Psychology, University of Aberdeen, King's College, Aberdeen AB24 3FX, Scotland, United Kingdom. Email: b.w.tatler@abdn.ac.uk

by systematically evaluating the information used to support observers' very first decision in scene viewing, namely, the endpoint of their first saccade, to reveal the mechanisms that underlie the initial stage of scene exploration.

The information processing demands very early in scene viewing are likely to differ from those present later in viewing for at least three reasons. First, the sudden appearance of the scene introduces a large change in visual information across the retina. Changes in visual information during scene viewing often prolong fixation durations, whether these changes are local, such as the onset of a local stimulus in peripheral vision (Pannasch et al., 2011; Pannasch & Velichkovsky, 2009), or global, such as a global change to contrast or luminance in a scene (Walshe & Nuthmann, 2021), or the reappearance of a scene after it disappears briefly (Henderson & Pierce, 2008; Henderson & Smith, 2009; Pannasch et al., 2011). These onset-related prolongations of fixation duration have been interpreted as reflecting a distraction effect (Pannasch et al., 2011; Pannasch & Velichkovsky, 2009), a surprisal response (Walshe & Nuthmann, 2021), or the need to prolong processing to integrate the new information (Pannasch et al., 2011; Walshe & Nuthmann, 2021).

Second, in most scene viewing research, initial processing occurs in the absence of any prior knowledge of what the scene image will be before its appearance. The importance of this is highlighted by the impact of a brief scene preview on subsequent scene processing and viewing behavior. A brief preview can reduce the extent to which the first eye movement is guided by low-level image features (Anderson et al., 2016) and can impact subsequent gaze behavior: When the scene was fully visible during search after a brief preview, the first two fixations were guided more efficiently (Hillstrom et al., 2012), and when peripheral information was unavailable during search, this benefit led to more efficient guidance throughout search (Võ & Henderson, 2010). These findings suggest a clear difference in how peripheral visual information is processed to guide scene exploration when a scene is first encountered without prior information compared to when prior information has been provided in the form of a brief scene preview.

Third, the content of foveal vision when the scene first appears has not been actively selected by the visual system. The importance of active selection for inspection becomes evident in paradigms where information changed during a saccade toward it. For example, Henderson (1992) asked participants to make a saccade toward either an object or a plus sign in peripheral vision. If the saccade target was a plus sign, the display was updated during the saccade to replace it with an object. Subsequent object naming times were significantly faster for the object in foveal vision if it had been present in peripheral vision before the saccade, suggesting that information gathered from the object in peripheral vision aided its identification when it reached foveal vision. Additionally, Castelhano and Pereira (2018) found that changing the identity of an object during a saccade toward it resulted in longer fixation durations on the object if it was visually dissimilar to the object present before the saccade. These findings suggest that selecting information to saccade to impacts the time course of processing that information when it is subsequently in foveal vision. Therefore, the absence of active selection at scene onset is likely to result in initial processing of foveal information in the moments after scene appearance being very different from processing foveal information later in viewing.

Taken together, these findings suggest that the information processing demands in the moments after scene onset not only differ from those later in viewing but have measurable impacts on decisions about where people look, both initially and later in viewing. Consistent with this, it has been suggested that initial scene viewing is underpinned by ambient processing of the scene, with a switch to more focal processing once the overall properties of the scene are understood sufficiently (Unema et al., 2005; Velichkovsky et al., 2005), although this proposal was developed from studies in which the period before the first saccade was excluded, so excluding the processing occurring during the initial moments of scene processing. Similarly, Schütt et al. (2019) suggested that the first saccade is guided strongly by bottom-up, and mostly low-level, image features, whereas later saccades are more strongly guided by high-level information. However, while these prior studies are consistent with there being different processing early than later in viewing, they do not directly test for differences in underlying mechanisms.

Key insights about the underlying mechanisms and information processing priorities for saccadic decisions come from studying the time course of the decision process (Carpenter, 1981; Noorani & Carpenter, 2016; Ratcliff, 2001). Different underlying decision mechanisms can be identified by comparing distributions of decision times (such as saccade latencies or fixation durations) to see whether these decision times are likely to have been generated by a common underlying mechanism (Carpenter, 1994, 1999; Noorani & Carpenter, 2016). Furthermore, modeling the factors that influence saccadic decision times can be used to infer whether and how different types of visual information in foveal and peripheral vision influence the decision to move the eyes (Tatler et al., 2017). Thus, studying the time taken to initiate the first saccade in viewing and comparing this to the time taken to initiate subsequent saccades can provide fundamental knowledge for developing a theoretical understanding of the first decision to move our eyes after a scene appears, and of how the processing occurring in the first moments of viewing a scene may contribute to this decision.

For later saccades in scene viewing, Tatler et al. (2017) showed that saccadic decisions can be well described as arising from an underlying mechanism that involves the linear accumulation of evidence from both foveal and peripheral vision to support a Bayesian-like decision process that continuously evaluates the relative benefit of prolonging a current fixation for further foveal processing or relocating the fovea to process another location in the scene. Within this "stay-or-go" decision mechanism, a saccade is made when the evidence that favors moving to a new location ("go") outweighs the evidence that favors maintaining fixation at the current location ("stay"). In their resulting gaze-control model, linear approach to threshold explaining space and time (LATEST), Tatler and colleagues showed that evidence favoring the "stay" component of the decision (and hence leading to longer fixations) includes higher edge information, visual salience, and semantic interest at the fixated location. Reciprocally, evidence favoring the "go" component of the decision (and hence leading to shorter fixations) includes higher edge density and semantic information at the location of the saccade target. Hence, an antagonism exists between central and peripheral information that informs gaze control decisions.

Given the different processing demands that are present when the scene first appears and the prior suggestions that processing might differ early in viewing compared to later, it is unclear whether the manner in which evidence is accumulated and/or the "stay-or-go" decision mechanism described by LATEST (Tatler et al., 2017) are appropriate for processing to plan the first saccade. Three possibilities exist for the mechanisms underlying processing during the

first moments of scene viewing that lead to the first decision to move the eyes.

First, the theoretical framework of a single mechanism of linear accumulation and evaluation of evidence may be unsuitable for describing how information is gathered and used in the build-up to the first saccade. If so, this will be evident in different distributions of decision times for the first compared to later saccades (Carpenter, 1981; Noorani & Carpenter, 2016). One possibility is that the initial processing of the scene is characterized by greater involvement of inhibitory processes, inhibiting the initiation of the first saccade while initial scene information is extracted. Consistent with this possibility, the latency of the first saccade tends to be longer than the durations of the fixations that follow (Castelhano et al., 2009; Einhäuser et al., 2020; Nuthmann & Clark, 2023; Over et al., 2007). Inhibiting saccades not only prolongs latency, but also changes the shape of the distribution of saccade latencies (Noorani, 2014; Noorani & Carpenter, 2016). Saccade cancelation is a feature of models of scene viewing such as Controlled Random-walk with Inhibition for Saccade Planning (CRISP; Nuthmann et al., 2010) and Inhibitory Control with Adaptive Timer (ICAT; Trukenbrod & Engbert, 2014), although these have not been applied to the first saccade in viewing. The dual process account of saccade timing proposed by Walshe and Nuthmann (2021) may also be appropriate for the first saccade. In their model, fixation duration arises not from a single mechanism accumulating and evaluating visual information, but from the combined effect of two mechanisms: one for processing visually surprising events and one for more detailed visual encoding. Scene onset in the absence of a preview represents a situation of high visual surprise because of the visual transient and because the scene content is unknown and unexpected. Thus, initial processing may be dominated by the mechanism for processing visual surprise rather than the mechanism for detailed visual encoding that dominates later saccades. This dual process model also predicts changes in the distribution of timings under conditions of high visual surprise, and so higher visual surprise very shortly after scene onset would be evident in differences in the shape of distributions between the first and later saccades.

Second, initial processing may involve the linear accumulation of evidence but not the stay-or-go mechanism that underlies later saccadic decisions. If so, this will be evident in differences in the sources of information that appear to contribute to the timing of the first saccadic decision. Given the emphasis of ambient processing at the start of viewing (Unema et al., 2005; Velichkovsky et al., 2005) and the need to extract scene layout and content for subsequent foveal exploration (Fei-Fei et al., 2007; Oliva & Torralba, 2001; Sanocki & Sulman, 2009), it may be that initial scene processing is underpinned by a mechanism that only involves the evaluation of peripheral information, with no evaluation of foveal information. In support of this, Nuthmann (2014) found that masking peripheral vision influenced the latency of the first saccade, whereas masking foveal vision did not. This peripheral-only account of decision mechanisms for saccade targeting has been suggested previously as a mechanism for targeting the first saccade after scene onset (Mackay et al., 2012), but this model is restricted to the first saccade and does not compare mechanisms underlying the first and later saccades in viewing. Peripheral-only processing models have also been proposed for explaining later saccadic decisions in reading (Carpenter & McDonald, 2007; McDonald et al., 2005) and scene viewing (Roos et al., 2008).

Third, initial processing may be underpinned by the same stay-or-go decision mechanism as later saccades, but with differences in the strategic prioritization of how information is weighted in the decision process. If so, we would expect to see contributions of information in foveal and peripheral vision to the timing of the first saccade, as has been shown for later saccades in viewing (Tatler et al., 2017). However, the relative balance between foveal and peripheral vision and between different types of information (e.g., low level vs. high level) in the decision process might be different for the first saccade than for later saccades, because of the different processing demands of the first moments of scene viewing.

Evidence from previous studies is not sufficient to distinguish the above three possibilities. Consistent with all three possibilities, studies have shown that the first saccade may be guided by both low-level features, preferentially selecting regions of high visual salience (Anderson et al., 2015), and higher level understanding and expectations (Schütt et al., 2019), although the evidence for whether high-level, semantic information in peripheral vision influences saccadic targeting is somewhat mixed. There is evidence that the first saccade is no more likely to target objects that are highly informative about the scene than objects that are not (Henderson et al., 1999), but also contrary evidence that semantically informative locations are more likely to be targeted by the first saccade (Nyström & Holmqvist, 2008; Spotorno & Tatler, 2017). Whether the first saccade preferentially targets semantically informative objects or not may be because of task constraints, with preferential selection of semantically informative objects only when they are task relevant (Nuthmann et al., 2019). However, no account is taken in these studies of the information present in foveal vision at scene onset and whether this influences the first decision to move the eyes (albeit that in the case of Nuthmann et al., 2019, there was no foveal content when the display appeared). Moreover, where the saccade is directed to only reveals the outcome of the decision process, rather than the underlying mechanisms that led to the decision. Insights into the underlying mechanisms come from investigation of the time course of the decision process.

Few studies have investigated the time course of the processes that lead to this first saccade. Anderson et al. (2015) found that short latency first saccades were more likely to target regions of high visual salience than longer latency first saccades, an effect that was later demonstrated to be reduced but not removed by showing participants a preview of the scene for sufficient time for semantic information to be gathered before the scene appeared for overt exploration (Anderson et al., 2016). Mackay et al. (2012) similarly suggested that short latency saccades are the outcome of a fast, automatic saccade generation mechanism based exclusively on bottom-up visual salience and likely controlled by subcortical pathways. They suggested, however, that longer latency first saccades are generated by processes based upon bottom-up salience weighted by top-down signals and reflect cortical processing of the incoming visual stimulus. This account by Mackay et al. (2012) is consistent with the possibility of different mechanisms operating during the early moments of scene processing to generate the first saccade in viewing, but is not sufficient to allow this conclusion, at least in part because any possible contribution of information in foveal vision during the build-up to the first saccade was not considered.

The Present Study

The present study was designed to distinguish between three different possible accounts of processing during the initial moments of scene viewing that contribute to the first decision to move the eyes and to

provide a first systematic investigation of whether and how information in foveal vision at scene onset contributes to this first saccadic decision.

A first step is to consider whether there is evidence that the first saccadic decision involves the same process of linear accumulation and evaluation of evidence that has been found for later saccades, or is underpinned by a different mechanism. Comparing the distributions of decision times for the first saccade to later saccades can provide key evidence to evaluate this: Different underlying mechanisms will lead to different distributions of responses (Carpenter, 1994, 1999; Noorani & Carpenter, 2016).

Previous studies have found that the time to launch the first saccade after scene onset is longer than the duration of the fixations that follow it during scene search (Nuthmann & Clark, 2023; Over et al., 2007). However, outside the context of scene search—where the task demands might favor extended peripheral processing of the scene prior to initiating overt exploration of the scene—the evidence is less direct. Differences between the times to launch the first and later saccades are evident in means reported when asked to memorize (Castelhano et al., 2009) or study the scenes carefully (Einhäuser et al., 2020), but these studies lack quantitative comparisons of the full distributions of times to launch the first and later saccades. In Study 1, therefore, we tested whether the distributions of times to launch the first saccade differed statistically from the durations of each of the fixations that followed, using data sets collected across different labs and under three different task conditions (free viewing, memorization, and search). By modeling this across and within each data set, we were able to determine whether differences between the first and subsequent saccades were found irrespective of viewing task and other experimental conditions, and thus ensure that the more detailed explorations in Studies 2 and 3 of the present report were unlikely to be parochial to the data sets collected for these subsequent studies. In Studies 2 and 3, we specifically tested whether first saccade decision times can be described using a single linear accumulation to threshold model of the underlying decision process. Following this, we explored whether and how information in foveal and peripheral vision contributed to the timing of the first saccade, and compared this to the factors that influenced the timing of subsequent saccades.

Taken together, the three studies presented here allowed us to develop a first theoretical framework for the way in which early scene processing informs the first decision to move the eyes, determining whether this first saccadic decision involves the same underlying mechanism as later saccades in viewing. In doing so, these studies are a key step to advance understanding of scene viewing and utilization of information from scenes to guide inspection behavior.

Study 1

Study 1 collected nearly half a million fixation records from 11 previously published studies of scene viewing. Our goal was to determine whether the first saccade in viewing is likely to arise from the same or a different mechanism that has been found to underlie later saccades by comparing the distributions of decision times for the first saccade to later saccades. If different mechanisms underlie the generation of first and subsequent saccades, then we would expect different distributions of saccadic latencies to describe each set of fixations (Carpenter, 1994, 1999; Noorani & Carpenter, 2016).

Method

Data Sets

We collated a data set of 463,309 fixations on complex scenes made by 361 observers across 11 previous studies. The sources from which data were drawn can be found in Table 1.

Eye movement data from five studies (370,247 fixations from 169 observers) were taken from the open source database published by Wilming et al. (2017a), drawing from a range of prior studies (Kaspar & Koenig, 2011; Onat et al., 2014; Ossandón et al., 2014). The remaining 18 studies in the Wilming et al. (2017a) database could not be used because they did not include the fixation that was ongoing when the scene appeared (15 studies), because they used faces as stimuli (two studies), or because they included an experimental manipulation of the gap between the disappearance of the pretrial marker and the onset of the scene (one study). Additionally, eye movement data from six studies by Tatler and various colleagues (93,062 fixations from 192 observers) were also included in the data set, drawn from a range of prior studies (Spotorno et al., 2014, 2015; Tatler, 2007; Tatler, Baddeley, & Gilchrist, 2005; Tatler et al., 2017; Tatler, Gilchrist, & Land, 2005). In Spotorno et al. (2014, 2015), some experimental scenes included objects placed in implausible locations in the scenes. In Spotorno et al. (2015), implausibly placed objects were associated with a longer first saccade latency, so for the present study, we excluded trials in which there were implausibly placed objects. No such effect was found in Spotorno et al. (2014), so all trials were included for analysis in the present study. For both Spotorno et al. (2014, 2015), data from filler trials were included for analysis in the present study, which were not analyzed in these prior publications. For these filler trials, all objects were placed in plausible locations. In the sections that follow, the data sets are named to reflect the instructions given to participants and do not follow the same data set naming conventions as used in Wilming et al. (2017a). For the correspondence between these data sets names and those used in Wilming et al. (2017a), please see https://osf.io/s5h8b/?view_only=58768b126db0439e946a49d94403d3a6.

While viewing times varied considerably between studies, only the first 10 fixations (and saccades) of each trial were included in the data set for the present study. The previous studies that were selected for the present data set included three different tasks (free viewing, memorization, and search) and a range of viewing times (see Table 1). For all data sets included in the present study, saccades were detected using the SR Research algorithm with the default sensitivity settings.

Data Analysis

We ran linear mixed models (LMMs) using the lme4 package (Bates et al., 2024) in the R statistical programming environment (R Core Team, 2024) to compare the (reciprocal) latency of the first saccade to the (reciprocal) durations of the next nine fixations in viewing. If we view a saccade as the end point of a decision to move the eyes, based on evaluation of visual information in the scene, then the reciprocal of the time to launch each saccade (thus latency for the first saccade and fixation duration before each subsequent saccade) conveniently describes the rate at which the decision to move the eyes is made and thus expresses the rate of rise of the underlying decision process (Tatler et al., 2017). Ordinal saccade number was entered into the model as a dummy-coded categorical variable with first

Table 1
A Summary Table of the Data Sets Analyzed in Study 1

Name	Source	Number of trials	Viewing time (s)	Number of participants[a]	Number of fixations[a]
Free viewing I	Onat et al. (2014)	255	6	23	50,522
Free viewing II	Ossandón et al. (2014)	255	6	43	90,027
Free viewing III	Kaspar and Koenig (2011)	240	6	45	96,957
Free viewing IV	Kaspar and Koenig (2011)	180	6	34	56,036
Free viewing V	Wilming et al. (2017a)	360	6	24	76,705
Free viewing VI	Tatler (2007)	120	5	22	22,379
Memory I	Tatler, Baddeley, and Gilchrist (2005)	48	Variable	13	4,414
Memory II	Tatler et al. (2017)	64	10	71	37,690
Search I	Spotorno et al. (2014)	126	Variable	24	6,876
Search II	Spotorno et al. (2015)	126	Variable	32	10,248
Search III	Tatler (2007)	120	Variable	30	11,455

[a] For the numbers of participants and fixations reported here, this refers to those selected for analysis in the present study (see Method section for data selection criteria).

saccade as the reference level, to compare the first saccade to each of the next nine saccades in turn. An overall model across data sets was run, followed up with separate models for each data set to see whether the overall pattern was also found in each data set. Because we planned these follow-up tests on each data set, we used a Bonferroni-corrected α of .025 for both the overall model and each follow-up model to reflect the fact that each data set was tested twice. In all models, participant was entered as a random effect; in the overall model, data set was also included as a random effect. The package lmerTest (Kuznetsova et al., 2020) was used to calculate p values, and ggplot2 (Wickham et al., 2024) was used for plotting data.

Transparency and Openness

In Study 1, data from Wilming et al. (2017b) were downloaded from https://datadryad.org/stash/dataset/doi:10.5061/dryad.9pf75 and data suitable for the current study were extracted to a .csv file using Matlab R2020a (see Study 1 Data Analysis section or details of inclusion criteria). Data from Tatler and colleagues used in Study 1 comprised data from a range of published studies and were extracted to a .csv file using Matlab R2020a. The data used in Study 2 were from a data set collected for Tatler et al. (2017), but focused on analyzing the first two saccades in viewing, which were excluded from the analyses in the previous publication, and have therefore not been analyzed previously. As a comparison, but not as the main focus of Study 2, the data previously analyzed in Tatler et al. (2017) were included in the present study and analyzed using a simpler model than previously to provide a comparison for the main focus of the present study. The sample size for Study 3 was based on that for Study 2 (collected previously for Tatler et al., 2017). The results of Study 2 provided a means of checking that the sample size for Study 3 was appropriate, using a power calculation based on the effects found in Study 2 (see Participants section of Study 3 for details). Data for the main experiment in Study 3 were collected between June 2017 and March 2020. Raw data files from the eye tracker were converted to ASCII files using the SR Research EDF2ASC program, outputting raw samples and events detected by the SR Research software. Data were extracted from the ASCII files using Matlab, and variables needed for analysis in the present study were collated into .csv files. Study 3 included a rating study to collect ratings of points of semantic interest in the scenes. With no formal analyses attached to this evaluation study, the sample size was arbitrarily chosen to provide a large sample of responses from which to create the interest maps and select the final set of scenes for the main experiment. None of the three studies was preregistered. This article was prepared in R Markdown (Allaire et al., 2024) using the papaja American Psychological Association template (Aust & Barth, 2023). Supplemental data for all studies and stimuli for Studies 2 and 3 are available to download via the Open Science Framework: https://osf.io/s5h8b/?view_only=58768b126db0439e946a49d94403d3a6. The R scripts containing the article and all analyses are available at the Open Science Framework.

Results

Overall, modeled across all data sets, the decision rate for the first saccade was longer than that of each of the nine saccades that followed it (Table 2), indicating that the latency of the first saccade was longer than the durations of the fixations that followed it

Table 2
Results of LMM to Predict Decision Rate (Reciprocal Fixation Duration) Across All Data Sets

Fixed effect	b	SE	t	p
Intercept	4.53	0.204	22.24	<.001
Saccade 1v2	1.04	0.013	78.21	<.001
Saccade 1v3	0.56	0.014	41.31	<.001
Saccade 1v4	0.38	0.014	27.88	<.001
Saccade 1v5	0.31	0.014	22.47	<.001
Saccade 1v6	0.30	0.014	21.51	<.001
Saccade 1v7	0.31	0.014	22.41	<.001
Saccade 1v8	0.32	0.014	22.70	<.001
Saccade 1v9	0.31	0.014	21.97	<.001
Saccade 1v10	0.30	0.014	21.04	<.001
Random effect	**N**	**Variance**		
Participants (intercept)	361	0.290		
Database/study (intercept)	11	0.444		
Residual		4.504		
Log-likelihood		−1,006,824.1		
Conditional R^2/marginal R^2		0.013/0.152		
Deviance		2,013,648.2		
AIC		2,013,674.2		
$N_{observations}$		463,309		

Note. LMM = linear mixed model; AIC = Akaike information criterion.

(Figure 1, top left). Indeed, while decision rate decreased over Saccades 2–10, it remained above that for the first saccade (Figure 2).

This overall pattern was also found in seven of the 11 data sets analyzed in the present study (see the Open Science Framework: https://osf.io/s5h8b/?view_only=58768b126db0439e946a49d94403d3a6 for modeling results for each individual data set). For three of the remaining four data sets (Free viewing I, Memory I, Search III), the first decision rate was also slower (thus the first saccade latency was longer) than that of the second saccade, but at some point in viewing, decision rates became significantly slower (thus fixation durations became longer) than that of the first saccade (Free viewing I: Saccades 5, 6, 7, 10; Memory I: Saccades 4, 7, 8, 9, 10; Search III: Saccades 5, 6, 7, 8, 9, 10). Only for one data set (free viewing VI) was the first saccade latency never significantly longer than the fixation durations preceding any of the subsequent saccades.

Discussion

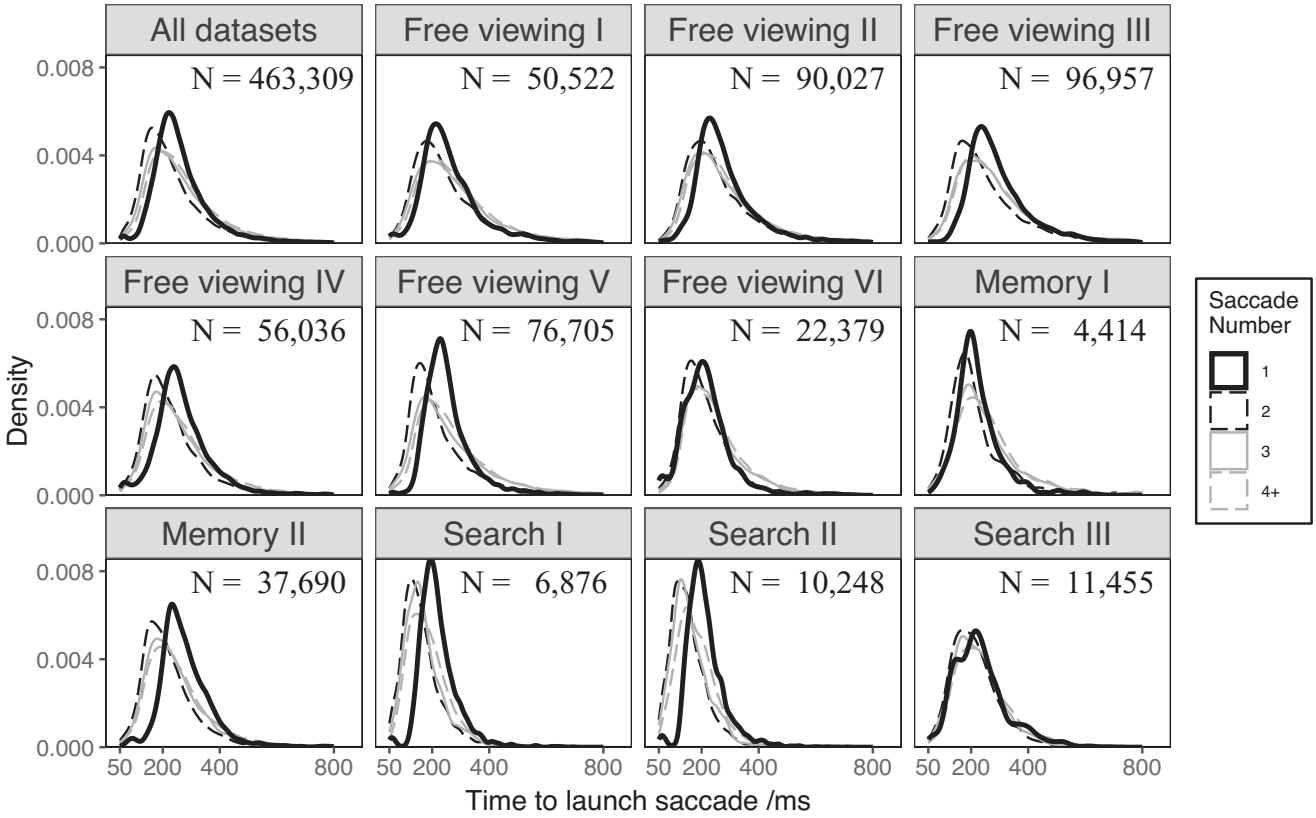
As a first step toward assessing whether the mechanisms underlying the first decision to move the eyes in scene viewing are qualitatively different from those underlying subsequent decisions to move the eyes, we compared the decision rate for this first saccade to that of the next nine saccades.

We considered 11 data sets of eye movements collected while people viewed complex scenes under three different task conditions and a range of viewing times. As expected, decision rate decreased over saccades two to nine, as was found in Tatler et al. (2017), which corresponds with previous reports of increasing fixation duration over viewing time (Buswell, 1935). Indeed, the pattern over Saccades 2–10 corresponds well with that found in prior studies, which have considered changes in fixation duration over ordinal fixation number (Nuthmann & Clark, 2023; Over et al., 2007) or viewing time (e.g., Mills et al., 2011; Nuthmann, 2017; Unema et al., 2005).

However, the decision rate for the first saccade did not fit this pattern. Rather, when all data sets were analyzed together, the decision rate for the first saccade was lower than all of the next nine saccades that followed, a pattern that was found in most of the data sets when modeled individually. This means that it took longer on average to launch the first saccade after the scene appeared than it did to launch any of the following saccades during viewing. This finding is consistent with prior work that has reported longer first saccade latencies than subsequent fixation durations during scene search (Nuthmann & Clark, 2023) or when asked to study the scenes carefully (Einhäuser et al., 2020), but extends these previous reports by providing a quantitative test of the differences in distributions of saccade timing and showing that this finding generalizes across tasks

Figure 1

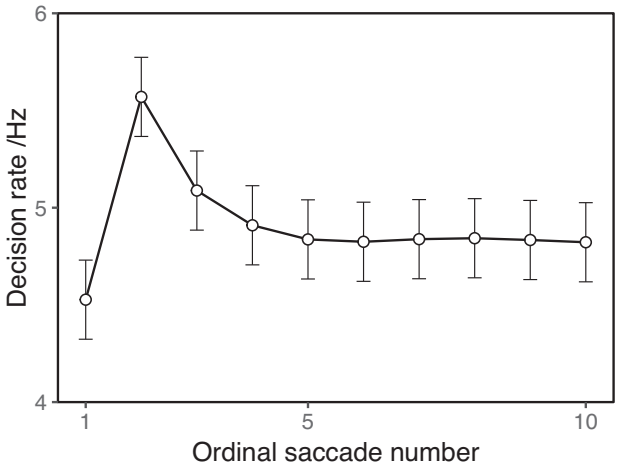
The Time to Launch Each of the Early Saccades in Scene Viewing Across Data Sets



Note. Saccade 1 (solid black line) corresponds to the first saccade made after the scene appeared, with the latency of this saccade calculated as the time from scene onset to the execution of the first saccade. For subsequent saccades, the time to launch refers to the duration of the fixation that immediately preceded the saccade. Saccades 2 (dashed black line) and 3 (solid gray line) clearly differ from Saccade 1. For ease of interpretation, Saccades 4–10 are collapsed and plotted as a single distribution (dashed gray line). Data are drawn from a range of laboratories, tasks, and experimental setups.

Figure 2

Decision Rates for the First 10 Saccades in Scene Viewing, Collapsed Across All Analyzed Data Sets



Note. Saccade 1 corresponds to the first saccade made after the scene appeared. The latency of this saccade is the time from scene onset to saccade execution. Data are plotted as partial effects from the LMM run across all data sets. Error bars show one standard error around the mean. LMM = linear mixed model.

and other experimental conditions. The result is consistent with what would be expected if inhibitory processes operate to delay the first saccade. The result is also consistent with the dual process account of saccade timing (Walshe & Nuthmann, 2021), with the visual surprise that accompanies the sudden onset of a scene acting to delay the initiation of the first saccade. Thus the results of Study 1 are consistent with the possibility that the processes underlying the first decision to move the eyes during scene viewing may differ from those underlying subsequent saccadic decisions.

The findings from Study 1 provide a necessary first step in investigating whether different mechanisms underlie how information is processed in the moments after a scene appears, and in demonstrating that differences between the first and later saccade latencies are common across tasks. However, these observed differences are not sufficient to distinguish the three possible accounts of these differences proposed earlier. For this, a more detailed analysis of a single data set is required. In Study 2, we conducted a detailed analysis of saccades within one data set collected as participants viewed complex scenes for an expected (but not administered) memory test. To test whether the type of decision mechanism proposed for later saccades (Tatler et al., 2017) can describe the first saccade, we first tested whether observed latencies for the first saccade can be modeled as arising from an underlying process involving linear accumulation and evaluation of visual information. Then, to consider whether and how information in foveal and peripheral vision contributes to the first saccadic decision, we ran an LMM to predict decision rate based on edges, visual salience, and semantic interest in foveal and peripheral vision.

Study 2

To understand the factors that influenced the time course of the first decision to move the eyes, data were extracted and analyzed from the data set collected for Tatler et al. (2017). This data set was selected because the scenes were previously coded for semantic interest (as a measure of informativeness) and thus semantic information could be included in models of information processing for saccadic decision time. To consider aspects of the time course of any change in decision priorities, data were also analyzed for the second saccade in viewing. In Tatler et al. (2017), the criteria employed for data selection meant that the first two saccades in viewing were excluded from all analyses and modeling. Thus, in the present study we modeled previously unanalyzed data for the first and second saccades in scene viewing. For comparison between the first two saccades and later saccades, we also ran models of these later saccades that differed slightly in structure from the models reported in Tatler et al. (2017) to align them with those reported here for the first and second saccades (see Data Analysis section).

The first step in Study 2 was to consider whether first saccade latencies can be modeled as linear accumulation processes. To do this, we assessed whether the latencies could be modeled as arising from a Linear Approach to Threshold with Ergodic Rate (LATER) decision process (Carpenter, 1981). We selected LATER to model linear accumulation because it has been used successfully to describe saccade latencies in a range of different experimental tasks and corresponds well to what is known about neural processes underlying saccadic decisions (see Noorani & Carpenter, 2016 for a review). Moreover, this model underlies the computational framework used for LATEST to explain saccadic decisions in scene viewing (Tatler et al., 2017).

Following this, we modeled the factors that influence saccadic decision time for the first, second, and later fixations in viewing. We tested whether visual information in central and peripheral vision influences fixation duration in a manner consistent with competition between information in central vision and potential targets of the next saccade as proposed in the model LATEST (Tatler et al., 2017) and further demonstrated by Einhäuser et al. (2020).

Method

For convenience, we report key information from the method used in Tatler et al. (2017) to collect this data set.

Participants

Seventy participants ($M_{age} = 23.3$) took part and were awarded course credit or financially reimbursed for their time. All had normal or corrected-to-normal vision. The study was approved by the local Psychology Ethics Committee.

Materials, Apparatus, and Procedure

Sixty-four images of everyday indoor and outdoor scenes were used as stimuli. These included scenes previously used in Brockmole and Henderson (2006) and Tatler (2007). Images were 800 × 600 pixels and were displayed on a 19-in. monitor positioned 63.5 cm from the viewer with a refresh rate of 100 Hz. At this distance, images subtended approximately 24.8° × 18.6° of visual angle.

Gaze data were recorded monocularly from the dominant eye at 1,000 Hz, using an SR Research EyeLink 1000 eye tracker. A nine-point calibration was followed by a nine-point validation procedure at the start of testing.

Each trial began with a central fixation marker that participants were required to fixate and which served as a single-point calibration

check. The experimenter triggered the onset of the scene once the participant was fixating the central marker. Each scene was displayed for 10 s after which it was replaced by a blank mid-gray screen. Throughout the experiment, participants believed that there would be a memory test at the end, but this was not administered, serving instead as a common task for observers.

Data Analysis

Saccades were detected using the SR Research algorithm with the default sensitivity settings. Blinks and saccades coinciding with blinks (and any corresponding fixations prior to blinks) were removed from the data set. To be included in modeling, data had to pass a strict set of quality checks. Saccades were excluded if the calibration error was worse than 0.5° across all nine calibration points or was worse than 1° at any one of the nine calibration points; if the pretrial single calibration check was worse than 1°; or if the samples within the period before the execution of the first saccade, or during the fixation that followed were noisy, identified as root-mean-squared error > 0.03 or radius of the bivariate contour ellipse area $> 0.15°$. Very small ($<0.5°$) or large ($>32°$) amplitude saccades and very short (<50 ms) or long (>800 ms) first saccade latencies were also excluded. These exclusion criteria were as used by Tatler et al. (2017), and further details of these can be found in this prior publication.

For the first saccade, the exclusion criteria resulted in the removal of 440 saccades (10.1% of the data set), leaving 3,937 first saccades for analysis. For the second saccade, the exclusion criteria resulted in the removal of 607 saccades (13.9% of the data set), leaving 3,763 second saccades for analysis.

In scene viewing, fixation durations can often be described as arising from a mixture of two distinct distributions. The first includes fixations that are correlated with visual processing and therefore arise from deliberative processes related to gaze control decisions. The second is a smaller subpopulation of fixations that are relatively short in duration and unrelated to information gathering and evaluation (Henderson & Pierce, 2008; Nuthmann et al., 2010; Tatler et al., 2017). Hence, for both the first and second saccades, we modeled the first latency using the software application Saccadic Programming and Instrumentation Computer (SPIC; Carpenter, 1994) to test whether latencies were likely to be drawn from one or two underlying distributions. This modeling procedure was used to estimate the parameters of the underlying distribution(s): the mean and standard deviation of the rate of the main distribution of latencies and the standard deviation of the rate for any identified subpopulation of unusually short saccade latencies (which we refer to as the "maverick" decision unit). These three parameters correspond to the three parameters in the LATER model of linear accumulation for saccadic decisions (Carpenter, 1981) and so provide a fit of this model to the data.

To test whether the observed latencies were consistent with having been generated by a process involving linear accumulation and evaluation of evidence, the fit between the observed data and the predictions of the fitted LATER model was calculated for each participant, based on the parameter estimates generated in SPIC (Carpenter, 1994). To estimate the fit between the predictions from the LATER model, which in some cases is the product of responses generated by two LATER decision units (the main and maverick units), a suitable approach is to use the Kolmogorov–Smirnov two-sample test (Kolmogoroff, 1941), from which the p value can provide a measure of the goodness of fit (Carpenter et al., 2009). A p value below .05 indicates a significant difference between the predictions of the model and the observed data.

The estimates from the SPIC modeling procedure were then used to calculate the probability that each saccade was generated by the main or maverick decision unit based on its execution time. These probability estimates were calculated from a simulation run using the parameter estimates (see Tatler et al., 2017 for details of this procedure). If we assume that saccades generated by the maverick unit are random and not based on evaluating visual information in the scene, then modeling the durations of these responses will not provide insights into the nature of the main decision process. Therefore, only saccades that were likely ($p > .6$) to have been generated by the main decision process were selected for analysis in the sections that follow.

After this model-fitting process, 3,146 first saccades and 2,813 second saccades were identified as likely to have been generated by the main decision unit and were used for the analyses that follow.

To model the influence of visual and other factors on decision times, an LMM was run to predict decision rate (the reciprocal of the first saccade latency) for the first saccade in viewing. Three descriptions of visual information—edges, salience, and semantic interest—were included in the model, each for the content of foveal vision and for the location in peripheral vision targeted by the saccade. These predictors cover a range of different levels of visual information in scenes, from basic low-level information in the form of edges, through intermediate-level information described by salience models—in this case, we used RARE2012 (Riche et al., 2013) for consistency with Tatler et al. (2017)—to higher level scene descriptions in the form of a measure of semantic interest. As a measure of semantic information in scenes we used semantic interest maps created for Tatler et al. (2017), which were generated in a separate experiment by asking observers to click on the five most informative locations in each scene and centering Gaussians with full width at half maximum of two degrees around each clicked location (for other applications of this method, see Krasich et al., 2020, 2021). This differs from the "meaning maps" approaches taken by Henderson, Hayes, and colleagues (Henderson et al., 2019; Henderson & Hayes, 2017; Peacock et al., 2023) for describing semantic information in scenes, which are based on degree to which the information extracted from a local scene region is informative independently of overall scene context. To account for eccentricity effects in information selection, we included both the amplitude of the outgoing saccade and the interaction between the amplitude of the outgoing saccade and each of the three visual predictors at the target location (as in Tatler et al., 2017).

Participant and scene were entered as random effects in the model. A second LMM with the same set of fixed and random effects was run to predict decision rate for the second saccade in viewing. For comparison, a third LMM was run on the data set previously analyzed in Tatler et al. (2017). This data set contained all saccades from the third in viewing (i.e., excluding Saccades 1 and 2). The LMM run here contained only those fixed effects present in the models for saccades one and two in viewing, excluding some of the fixed effects in the previously published model for this data set. Because of the pattern of results across these models, a follow-up LMM was run to test whether effects of the three visual predictors (edges, salience, and interest) differed across the first two saccades, with saccade number included as a sum-coded fixed effect, together with

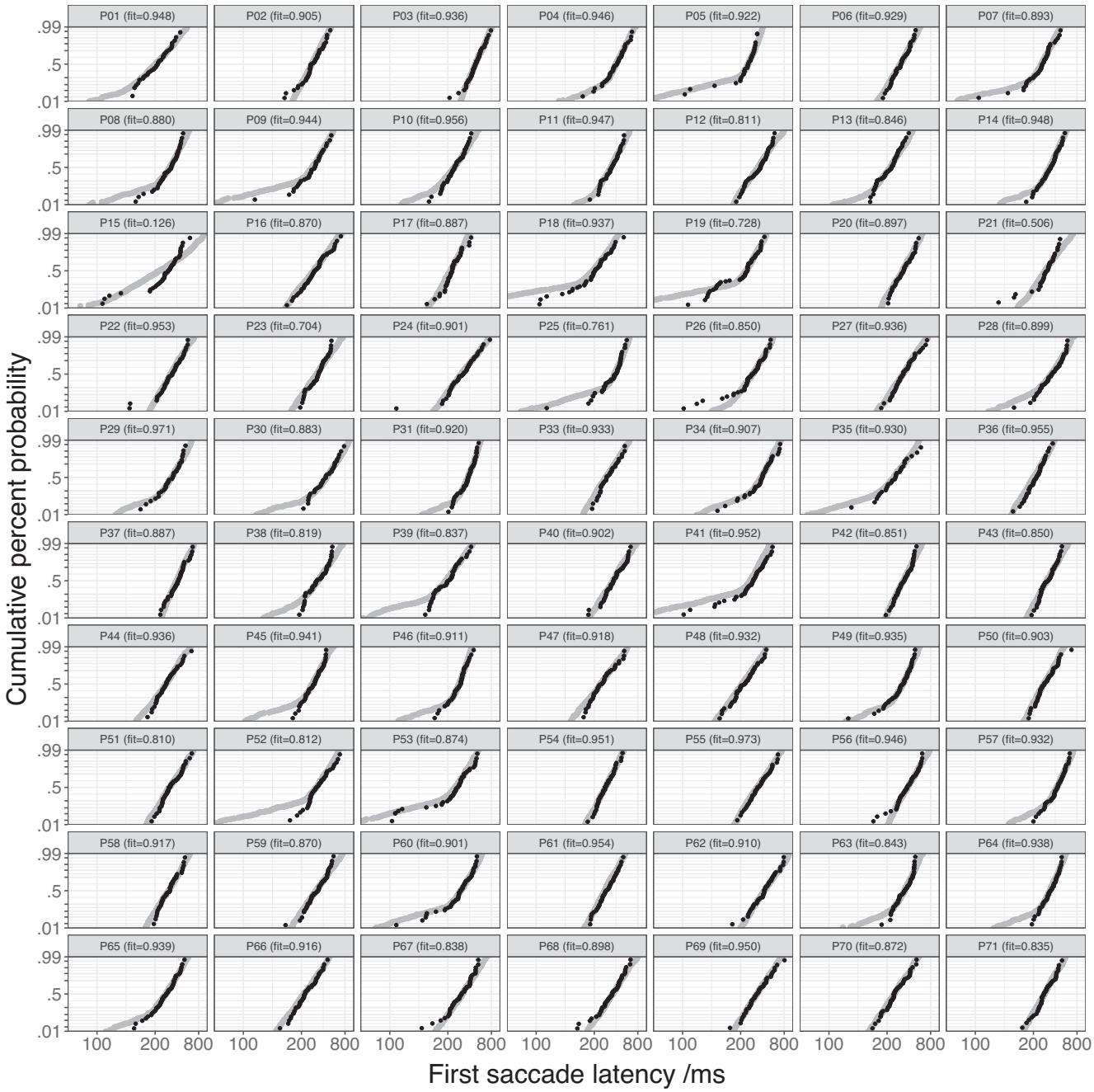
the interactions between saccade number and each of the visual predictors in foveal and peripheral vision. This post hoc model involved retesting data included in the individual models for Saccades 1 and 2, so a Bonferroni-corrected α of .025 was used. The same R packages were used as in Study 1 for analysis and data visualization.

Results

For the first saccade, predictions generated by the fitted LATER models corresponded well with the observed latencies (Figure 3). Visually, fits vary somewhat between participants, but quantifying

Figure 3

Reciprobit Plots of First Saccade Latency in Study 2



Note. In all cases, the majority of datapoints lies along a straight line, indicating that these saccades are likely to come from a single underlying LATER-like decision process. Maverick unit contributions are evident in some but not all participants in the form of a kink to the left at the bottom of the plot, indicating more saccades executed with very short latency than would be expected given a single sensible decision process. For each participant, the observed data (black points) are plotted over the predictions from the fitted LATER model (gray points) and the fit between the predicted and observed latencies (expressed as the Kolmogorov–Smirnov *p* value) is indicated in the title of each plot. Higher values denote better fits. P = participant; LATER = linear approach to threshold with ergodic rate.

the fits by running Kolmogorov–Smirnov two-sample tests of the differences between the predicted and observed latencies and using the resulting p value as a measure of fit (Carpenter et al., 2009) showed that in all cases, there was no significant difference between the predicted and observed latencies (minimum $p = .126$) and in most cases the fit was good (Figure 3). The correspondence between predictions and observations suggests that first saccade latencies can be modeled effectively as arising from a decision process involving linear ballistic accumulation and evaluation of evidence. Of the 70 participants, 34 showed statistical evidence for a subpopulation of first saccades arising from a maverick decision unit. For the remaining participants, the first saccade latencies appeared to all be generated by a single underlying decision mechanism involving the linear accumulation and evaluation of evidence.

The results of the LMM (Table 3) for the first saccade showed that none of the predictors relating to visual information in foveal vision contributed significantly to the timing of the first saccade (solid lines in Figure 4). Conversely, the timing of the first decision to move the eyes was influenced by edge and salience information at the peripheral target location, such that the decision to move the eyes had a higher rate (thus the eyes moved sooner) when the target location contained information that was higher in edge content or salience (solid lines in Figure 4). Both of these effects were qualified by interactions with the amplitude of the first saccade, such that as first saccade amplitude increased, the effect of edges, and salience on decision time decreased. An overall effect of first saccade amplitude was also found, such that targeting locations further into peripheral vision was associated with longer decision time.

For the second saccade, the LMM (Table 3) showed a different pattern of effects (dashed lines in Figure 4). Importantly, both information in foveal vision and at the peripheral target location contributed to decision times. The effects of the significant visual predictors were as expected given the theoretical framework proposed by Tatler et al. (2017): decision rate decreased when edge content, salience, or semantic interest was higher in foveal vision (so prolonging fixation time). Furthermore, decision rate increased (so reducing fixation time) when lower level (in this case salience) or semantic interest was higher at the peripheral target location. Importantly, the pattern of effects for second saccade differed from the first not only because information in foveal vision now contributed to decision time, but also because semantic interest (both in foveal and peripheral vision) contributed to decision times for the second saccade.

Qualitative comparisons between the patterns of significant effects in the models run for the first, second, and later saccades suggest that the factors that influence the first saccade are very different from those influencing the second and subsequent saccades in viewing. In contrast, the patterns of significant effects for the second saccade were quite similar to those found for later saccades (Table 3). The different effects of visual information in foveal and peripheral vision between the first and second saccades in viewing are shown in Figure 4. A follow-up LMM was run to compare directly the differences between the first two saccades, with saccade number included as a sum-coded fixed effect that interacted with the fixed effects describing visual information in foveal and peripheral vision.

Table 3
Results of LMM to Predict Decision Rate in Study 2 for the First, Second, and Later Saccades in Viewing

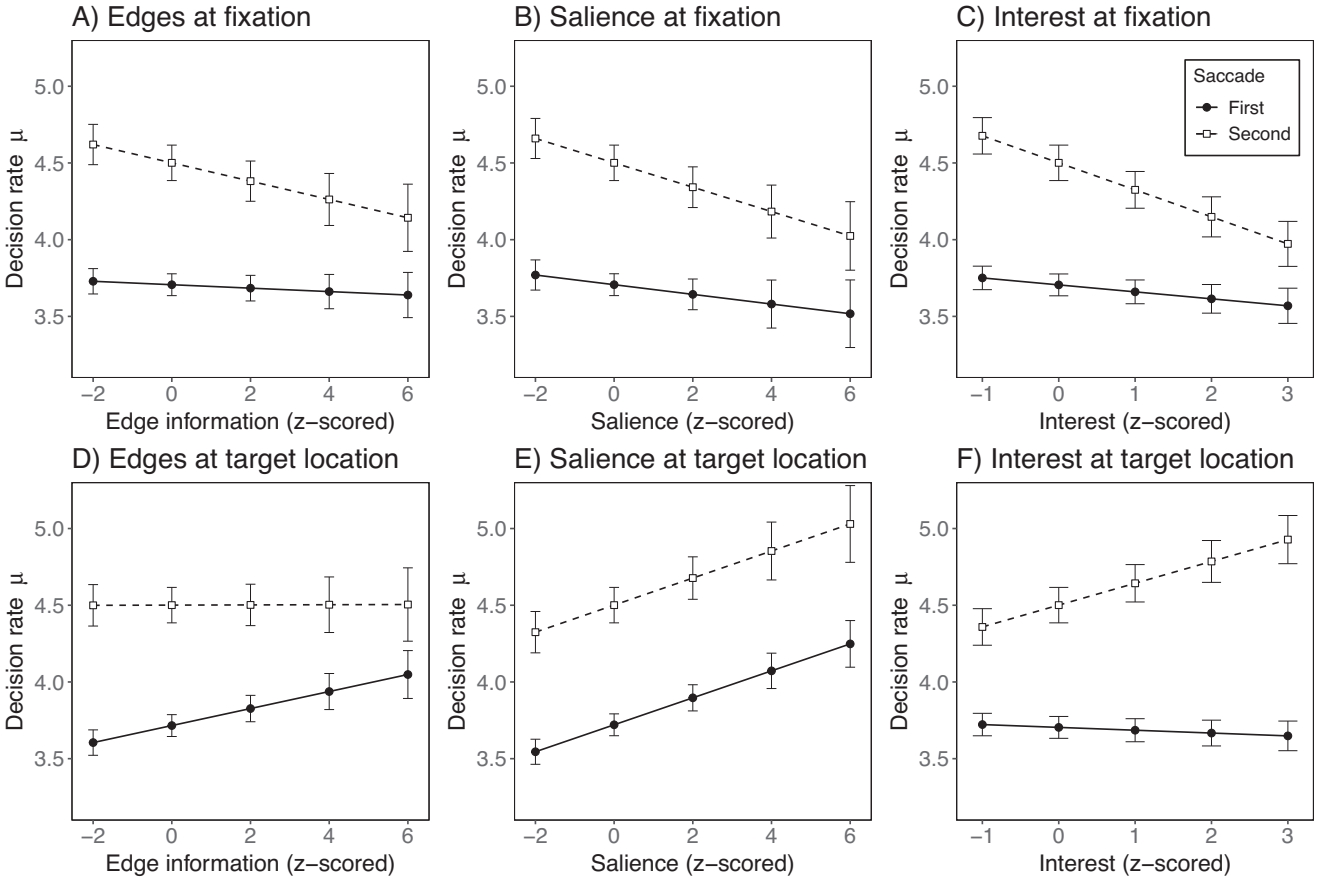
Fixed effect	First saccade				Second saccade				Later saccades			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	3.69	0.071	52.29	***	4.44	0.114	38.98	***	3.30	0.048	68.42	***
Outgoing saccade amplitude (linear)	−1.11	0.859	−1.29	ns	3.23	1.403	2.30	*	−3.56	0.856	−4.16	***
Outgoing saccade amplitude (quadratic)	−2.52	0.859	−2.93	**	−4.21	1.410	−2.99	**	−1.76	0.851	−2.07	*
Edge information at fixation	−0.01	0.021	−0.52	ns	−0.06	0.031	−1.92	ns	−0.01	0.004	−3.47	***
Edge information at target location	0.05	0.019	2.42	*	0.00	0.029	0.07	ns	0.01	0.004	2.56	*
x Outgoing saccade amplitude (linear)	−2.12	0.981	−2.16	*	0.21	1.417	0.15	ns	−1.33	0.904	−1.47	ns
x Outgoing saccade amplitude (quadratic)	−0.81	1.036	−0.78	ns	0.09	1.415	0.07	ns	−0.95	0.910	−1.04	ns
Salience at fixation	−0.03	0.035	−0.91	ns	−0.08	0.031	−2.52	*	−0.01	0.005	−3.28	***
Salience at target location	0.06	0.019	3.30	***	0.06	0.029	1.96	*	0.00	0.004	−1.16	ns
x Outgoing saccade amplitude (linear)	0.33	1.032	0.32	ns	1.05	1.669	0.63	ns	−0.82	0.908	−0.90	ns
x Outgoing saccade amplitude (quadratic)	−2.03	0.923	−2.20	*	−2.22	1.659	−1.34	ns	1.49	0.922	1.61	ns
Semantic interest at fixation	−0.05	0.030	−1.53	ns	−0.18	0.029	−5.99	***	−0.07	0.004	−16.37	***
Semantic interest at target location	−0.01	0.017	−0.30	ns	0.16	0.027	5.76	***	0.03	0.004	7.16	***
x Outgoing saccade amplitude (linear)	0.31	0.902	0.34	ns	0.11	1.389	0.08	ns	−0.87	0.896	−0.97	ns
x Outgoing saccade amplitude (quadratic)	1.08	0.847	1.28	ns	1.14	1.363	0.84	ns	1.48	0.896	1.65	ns

Random effect	*N*	Variance	*N*	Variance	*N*	Variance
Participants (intercept)	70	0.249	70	0.758	70	0.158
Scenes (intercept)	64	0.080	64	0.096	64	0.003
Residual		0.447		1.543		0.681
Log-likelihood		−3,382.0		−4,747.2		−68,110.9
Conditional R^2/marginal R^2		0.014/0.432		0.032/0.377		0.006/0.196
Deviance		6,764.0		9,494.5		136,221.7
AIC		6,800.0		9,530.5		136,257.7
$N_{observations}$		3,146		2,813		55,341

Note. The model for the later saccades is a reanalysis of the data in Tatler et al. (2017) with fewer fixed effects to build an equivalent model to those used for the first and second saccades. LMM = linear mixed model; ns = not significant; AIC = Akaike information criterion.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 4

Partial Effects of the Six Visual Predictors on the Decision Rates of the First and Second Saccades in Viewing



To avoid over complicating the model, and because this model was constructed to test specifically whether effects of visual predictors on decision rate differed between the first two saccades, outgoing saccade amplitude was not included as a fixed effect in the model. Because this LMM involved retesting the data analyzed in the previous models for Saccades 1 and 2, we used a Bonferroni-adjusted α of .025. Crucially, this LMM (Table 4) revealed significant interactions between fixation number and the effects of (a) salience in foveal vision, (b) semantic interest in foveal vision, and (c) semantic interest at the peripheral target location, suggesting that the effects of each of these three visual predictors varied between the first and second saccades in viewing (as evident in Figure 4), confirming the different patterns found for these predictors across the separate models for the first and second saccades (Table 3).

Discussion

The analyses in Study 2 were run to follow up the findings from Study 1 which showed that the decision rate for the first saccade after scene onset was significantly slower than subsequent saccades. First, we considered whether the first decision appears to arise from a process of linear accumulation and evaluation of evidence, as has been found for later saccades (Tatler et al., 2017). Second, we considered whether the stay-or-go decision mechanism, in which information in foveal and peripheral vision is evaluated (Tatler et al., 2017), is an appropriate decision mechanism for the first saccade in viewing.

First saccade latencies were well-fitted with the linear accumulation decision model LATER (Carpenter, 1994; Roos et al., 2008), with about half of the individuals showing evidence of a subset of infrequent but very short latency saccades that were unlikely to have been generated by the same decision process as the majority of latencies for those individuals. These short latency saccades can be explained as being generated by a second linear accumulation process that competes with the main decision unit, but has little to do with processing the information in the scene, a so-called "maverick" decision unit that injects randomness into viewing behavior (Carpenter, 1994; Roos et al., 2008). Competition between a main, sensible decision unit that generates the majority of saccades, and a maverick unit that interrupts this process occasionally with very short latency saccades is consistent with the LATEST framework for explaining scene viewing (Tatler et al., 2017). However, for later saccades in viewing most individuals show the presence of the fast responses consistent with the maverick decision unit (Tatler et al., 2017), whereas in the present study, for the first saccade, this population of latencies was less common across participants. This suggests that while the overall framework seems appropriate, the outcome of competition between the two units differs.

Table 4
Results of LMM to Predict Decision Rate in Study 2 Over the First Two Saccades in Viewing

Fixed effect	*b*	*SE*	*t*	*p*
Intercept	4.10	0.084	49.05	***
Saccade number	−0.41	0.014	−28.83	***
Edge information at fixation	−0.05	0.019	−2.58	*
x Saccade number	0.00	0.018	−0.26	ns
Edge information at target location	0.01	0.019	0.41	ns
x Saccade number	0.01	0.017	0.54	ns
Salience at fixation	−0.05	0.019	−2.81	**
x Saccade number	0.06	0.018	3.15	**
Salience at target location	0.07	0.019	3.84	***
x Saccade number	0.02	0.018	1.00	ns
Semantic interest at fixation	−0.12	0.018	−6.60	***
x Saccade number	0.06	0.018	3.66	***
Semantic interest at target location	0.07	0.017	3.93	***
x Saccade number	−0.10	0.016	−5.99	***
Random effect	***N***		**Variance**	
Participants (intercept)	70		0.388	
Scenes (intercept)	64		0.080	
Residual			1.093	
Log-likelihood			−8,903.5	
Conditional R^2/marginal R^2			0.118/0.382	
Deviance			17,807.1	
AIC			17,841.1	
$N_{observations}$			5,959	

Note. Significance is reported using Bonferroni-adjusted αs. LMM = linear mixed model; ns = not significant; AIC = Akaike information criterion.
* $p < .025$. ** $p < .005$. *** $p < .0005$.

While the overall framework for the first saccadic decision is consistent with that proposed for later saccades, the way that visual information in foveal and peripheral vision contributes to this first decision was very different. The results from Study 2 suggest that information in foveal vision did not contribute to the initial evaluation of the scene for planning the first saccade. Rather, this first decision was based upon information in peripheral vision, as shown by the effects of visual information at the location targeted by the first saccade. Moreover, only low-level information—edges and visual salience—at the target location contributed to the timing of the first saccadic decision, with no evidence that semantic information either in foveal or peripheral vision contributed to saccade timing. By the second saccade, the pattern of effects was very different, with foveal and peripheral information contributing to saccade timings and the pattern of effects closely matching those found for later saccades in the present study (and found previously in Tatler et al., 2017).[1]

The findings of Study 2, therefore, suggest that the first saccadic decision is qualitatively and quantitatively different from later decisions to move the eyes when viewing scenes. The results suggest that the first saccade is planned based on the evaluation of low-level information in peripheral vision, without contributions from foveal vision or higher level interpretation of scene contents. The possibility that the first saccade is primarily guided by salience has been debated considerably in the scene viewing literature (see Tatler et al., 2011). A frequent claim is that salience dominates saccade planning soon after scene onset, with higher level factors and top-down control emerging later in viewing (e.g., Parkhurst et al., 2002), but other authors have argued against this claim (e.g., Henderson & Hayes, 2017; Tatler, Baddeley, & Gilchrist, 2005).

Moreover, this result argues against the stay-or-go decision mechanism that was proposed in LATEST (Tatler et al., 2017), in which the relative benefits of prolonging foveal inspection or relocating the fovea to a peripheral location are evaluated. In this way, the time until execution of the first saccade in scene viewing might be better thought of as a latency of orienting in response to information appearing in peripheral vision when the scene suddenly appears.

Before concluding that the mechanisms underlying the first decision to move the eyes differ from those underlying later saccadic decisions, it is important to consider whether these apparent differences could arise as a "side effect" from the fact that observers were always looking at the center of the scene when it appeared. Indeed, starting in the center of the screen results in different viewing behavior compared to when participants start noncentrally (Rothkegel et al., 2016; Trawiński et al., 2023). Rothkegel et al. (2016) found that starting to the left or right of center resulted in eye-movement behavior that could not be well predicted by the SceneWalk model, which offers a good account of eye-movement behavior when starting centrally (Engbert et al., 2015). Trawiński et al. (2023) compared central versus noncentral initiation when viewing paintings and induced noncentral initiation by presenting a Navon figure to inspect prior to the painting presentation. They found that participants made shorter duration fixations, smaller amplitude saccades and were less likely to look at faces in the paintings when starting noncentrally than when starting centrally. If starting centrally leads to different spatial and temporal selection during viewing, and this selection arises from an evaluation of information to reach a decision about where to saccade to, it follows that starting in the screen center may result in different priorities for information evaluation than when starting noncentrally, even if the decision utilizes the same underlying mechanism as later saccades.

One reason for such strategic differences in information prioritization when starting centrally might be that viewers do not expect this location to be important or informative. Forcing the initial fixation point to be at the center of the image imposes an arbitrary starting point for scene processing that is likely to be less related to scene content than subsequent viewer-selected fixation points. Hence, it may be that the content of foveal vision when the scene appears is not as informative as that which is available at locations that are actively selected during viewing. If so, this would support any expectations by the viewer that this location is unimportant and not a priority for foveal processing at scene onset. Indeed, post hoc analyses conducted to explore this possibility (see the Open Science Framework: https://osf.io/s5h8b/?view_only=58768b126db0439e946a49d94403d3a6) showed that edge density, salience, and semantic interest at the initial (centrally constrained) point of fixation were lower and less variable than those available at the second point of fixation (i.e., at the location actively selected by the first saccade). Thus, the apparent lack of influence of foveal information on the first decision to move the eyes in Study 2 may be related to this reduced availability and variability of information in foveal vision when the scene first appears.

[1] Note that while we found an effect of edges at the peripheral target location for later saccades in Study 2 and in Tatler et al. (2017), Nuthmann (2017) found no such effect. This may reflect the different ways in which edge information was defined and extracted from the images or the different mix of fixed and random effects in the models.

Thus, to understand whether the qualitative and quantitative differences between the first and subsequent saccades in Study 2 reflect different underlying mechanisms in which only low-level information in peripheral vision is evaluated for the first saccade, or different priorities for weighting information within a common underlying mechanism (in this case the stay-or-go mechanism from LATEST), it is crucial to study the first saccadic decision when viewing is initiated noncentrally in the scene and the content of foveal vision manipulated.

Study 3

In Study 3, we allowed the starting location for scene viewing to occur noncentrally to experimentally manipulate the content and position of foveal vision at scene onset and provide a noncentral location for initial scene processing. This manipulation offered a direct test of whether the content of foveal vision contributes to the first decision to move the eyes in scene viewing when it contained high or low salience and high or low semantic interest, and viewing started noncentrally.

Method

Participants

Seventy-one participants[2] took part in the experiment for course credit or financial reimbursement. Sample size was guided by Study 2 (Tatler et al., 2017), which provided a means of estimating power in Study 3. Given the theoretical focus on the first saccadic decision in the present research, we conducted power calculations for detecting, in Study 3, the observed effects of edges and salience at the saccade target location on the first saccade model that were found in Study 2. We used the powerSim() function in the simr package (Green & MacLeod, 2016, 2023) to calculate power across 1,000 simulations for each of these effects.

The power to detect the observed effect of edges at the target location in the model was 80.5%, and the power to detect the observed effect of salience at the target location was 94.3%. This suggests that collecting a sample for Study 3 that matched that of Study 2 would be sufficiently powered. All participants had normal or corrected-to-normal vision and were naive to the purposes of the experiment. The study was approved by the local Psychology Ethics Committee (PEC/3583/2016/12).

Materials

Initial Selection of Stimuli. Stimuli were selected from the LabelMe database of scenes with object annotations using the LabelMe Matlab Toolbox (Russell et al., 2008). Initially, 200 images were selected, 100 indoor and 100 outdoor scenes that had an aspect ratio of 4:3 and a minimum width of 1,024 pixels. To have a variety of scenes in terms of structure and content, scenes were selected based on (a) whether or not they contained at least one person and (b) how many annotated objects they contained. Sparse scenes were scenes that contained between three and 10 uniquely identified annotated objects, with no more than a total of 20 annotated objects in the scene. Cluttered scenes were those that contained more than 20 uniquely named annotations and fewer than 70 annotations in total (to avoid any scenes containing too many repeats of a single object). The outdoor images were divided equally into four groups of 25 images depending on whether they contained a person or not and were of high or low clutter. For the indoor images, there were insufficient scenes containing people that met our criteria for aspect ratio and pixel dimensions. As a result, the 100 indoor images were split into sparse scenes without people (43 images), cluttered scenes without people (43 images), and scenes containing people (14 images). Scene sparseness, whether they were indoor or outdoor, and whether or not they contained people were varied to produce a diverse scene set, but were not dimensions analyzed or considered further.

Evaluation Study. A study was run to evaluate semantic interest in the 200 images selected from the LabelMe database. The results of this pilot were used to select 120 images for the main experiment reported in Study 3. Fifty independent judges (30 female) with a M_{age} of 22.1 ($SD = 3.2$) took part in this experiment voluntarily or for course credit. All had normal or corrected-to-normal vision and were naive to the purposes of the experiment. The study was approved by the local Psychology Ethics Committee (PEC/3583/2016/12). Participants were shown each image in turn and asked to select the five most semantically informative locations in each scene using a mouse pointer, following the procedure used by Tatler et al. (2017), and outlined in Study 2.

The maps of salience (created using the Matlab implementation of the adaptive whitening salience [AWS] model[3]; Garcia-Diaz et al., 2012) and semantic interest were used to identify four starting locations in each scene (Figure 5). These locations were (a) high in semantic interest and high in salience, (b) high in semantic interest and low in salience, (c) low in semantic interest and high in salience, and (d) low in semantic interest and low in salience. These locations were identified algorithmically in Matlab and as much as possible matched for eccentricity in the scene.

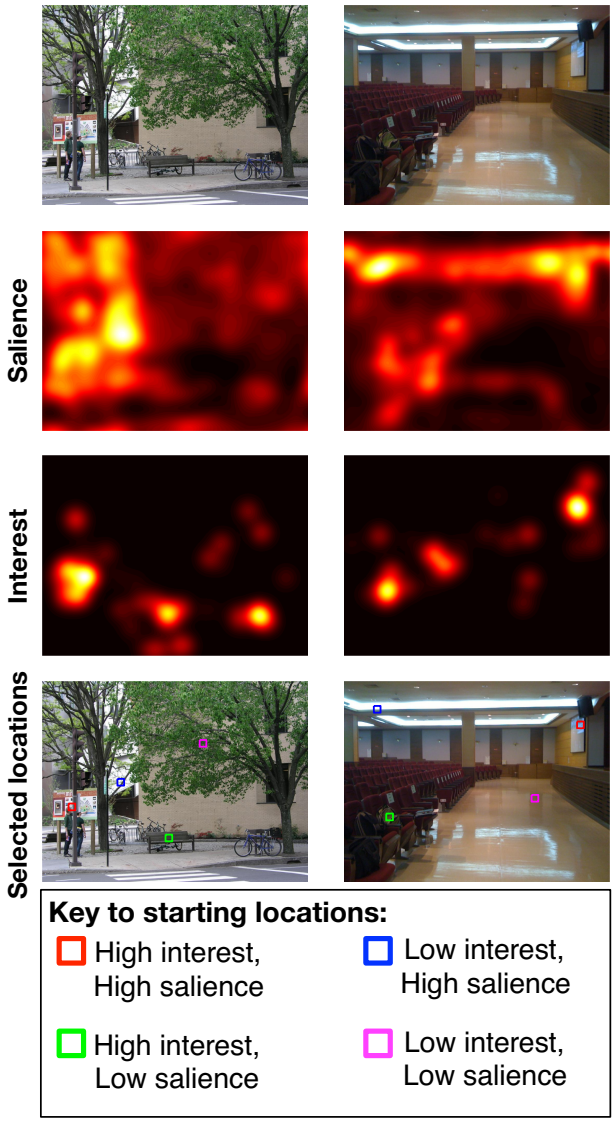
Starting locations were selected within each scene by creating binary thresholded maps of pixels with very high salience or semantics (top 10% pixel values) and low salience or semantics (bottom 55% pixel values). These four maps were combined in pairs to create maps of pixels that satisfied the four conditions for salience and semantics. For example, to identify locations that were high in salience and high in semantic interest, the two thresholded maps for high salience and high semantic interest were combined to identify pixels that overlapped between these two. From the binary thresholded condition maps, a random pixel was selected. Across the four conditions, no selected location could be within 1° of a location selected on another map. Given the often extensive number of pixels identified in the low interest, low salience map, a location was selected that had the same distance from the center of the scene as the mean of the other three starting positions in the scene. The limited overlap between some combinations of thresholded maps (e.g., high interest, low salience locations) meant that it was not possible to match distance from center across all starting positions.

Stimuli Selected for Main Experiment. A total of 120 images were selected for Study 3 that best allowed differentiation of four

[2] No demographic data were gathered from participants as these were at the time deemed unnecessary for the purposes of the present study. All participants were university students.

[3] Note that AWS was used here as the salience model to create salience maps for the evaluation study rather than RARE2012 which was used for analysis in Study 2 as it performed marginally better on the MIT/Tuebingen Saliency Benchmark (https://saliency.tuebingen.ai/results.html).

Figure 5
Two Example Images From Study 3 Showing Salience and Semantic Interest Maps and the Four Selected Starting Locations in Each Scene



Note. See the online article for the color version of this figure.

locations according to the criteria above and also provided an even number across categories of scenes. These were 80 outdoor images (20 with people and high clutter, 20 with people and low clutter, 20 with no people and high clutter, 20 with no people and low clutter) and 40 indoor images (20 with no people and high clutter, 20 with no people and low clutter).

Eye Movement Recording

Eye movements were recorded using an SR Research EyeLink 1000+ eye tracker, recording pupil and corneal reflection positions at 1,000 Hz. Eye movements were recorded from the dominant eye, which was identified using the hole-in-card test (Dolmen, 1919). A nine-point calibration procedure followed by a nine-point validation procedure was used at the start of testing. Saccades were detected using the SR Research algorithm with the default sensitivity settings. Blinks and saccades coinciding with blinks (and any corresponding fixations prior to blinks) were removed from the data set.

Procedure

Participants were seated with head resting on a chin and forehead rest positioned 72 cm from a 53.2 cm × 30 cm monitor with 1,024 × 768 pixel resolution. Images were displayed at 1,024 × 768, covering 40 × 30 cm of the screen and subtending 31.0° × 23.5° of visual angle.

At the start of the experiment, participants were asked to view each scene freely in preparation for a memory test that would be administered at the end of the experiment. The memory test was never administered.

Each trial began with a central marker that served as a calibration check, performed manually by the experimenter. Following acceptance of this calibration check, the central marker disappeared and was immediately replaced by a noncentral fixation target. This target was a 0.8° × 0.8° (26 × 26 pixels) black outlined box, which participants were required to fixate. The scene appeared once the eye tracker detected that the eye had been within this noncentral fixation target for a minimum of 200 ms. The scene remained on the screen for five seconds before it disappeared and the central calibration check for the subsequent trial appeared. The starting position from which scenes were explored was manipulated with a 2 × 2 repeated measures design in which starting locations were selected to be of high or low semantic interest for the scene and of high or low visual salience.

Participants were free to take a break at any point, and calibrations were performed whenever the pretrial calibration check indicated that calibration was slipping or after participants took a break from testing.

Data Analysis

We recorded 125,416 saccades from the 71 participants, each viewing 120 photographic scenes. Saccades were excluded if they failed to pass the same set of exclusion criteria used in Study 2. Overall, these exclusion criteria resulted in the removal of 26.5% of the data set, leaving 5,570 first saccades after scene onset, 5,943 second saccades after scene onset, and 80,709 saccades made later in viewing for subsequent analysis, from 66 participants (five participats were removed through the exclusion criteria).

We modeled the latency of the first saccade (the time from scene onset to saccade initiation) and fixation duration for the subsequent saccades, which both reflect the time to make a decision to move the eyes (Carpenter, 1981; Tatler et al., 2017) and are thus referred to collectively below as decision times. Data were modeled using the software application SPIC (Carpenter, 1994) to test (a) how well a linear accumulation mechanism can model the observed decision times and (b) whether decision times were likely to be drawn from one or two underlying distributions. This modeling procedure was used to estimate the parameters of the underlying distribution(s): the mean and standard deviation of the rate of the main distribution of decision times and the standard deviation of the rate for any identified subpopulation of unusually short decision times (i.e., "maverick" saccades).

As in Study 2, we calculated the fit between the model predictions and observed data for the first saccade in viewing to determine whether a linear accumulator model—LATER—was an appropriate mechanism for accounting for the observed saccade latencies. A Kolmogorov–Smirnov two-sample test (Kolmogoroff, 1941) for each participant was run, with the p value used as a measure of the goodness of fit for the model (Carpenter et al., 2009). The model parameter estimates were used to restrict subsequent analyses to saccades that were likely to have been generated by the main decision process (so excluding responses likely to have come from the "maverick" decision process). After this model-fitting process, 4,820 first saccades, 4,041 second saccades, and 43,575 later saccades were identified as likely to have been generated by the main decision unit and were used for the analyses that follow.

To model the influence of visual and other factors on decision times, three LMMs were run to predict decision rate (the reciprocal of the saccade latency/fixation duration) for (a) the first saccade in viewing, (b) the second saccade in viewing, and (c) later saccades in viewing. The models were constructed with the same fixed and random effects as in Study 2. While the evaluation study used the AWS model to construct salience maps, for consistency with Study 2, RARE2012 was used for extracting salience in central and peripheral vision for the analyses that follow (see Results footnote for a comparison between using RARE2012 and AWS for the models in Study 3). The same R packages were used as in Studies 1 and 2 for analysis and data visualization.

Our 2×2 design to manipulate salience and interest in foveal vision at scene onset presented the opportunity to analyze the data in two different ways: either with salience and interest in foveal vision entered into the LMM as sum-coded, two-level (high, low) categorical fixed effects to reflect the design, or as continuous fixed effects to reflect the variability in salience and semantics within these starting coalitions and to allow better comparisons to models of the second and later saccades. We ran both versions of these models and used Bayesian information criterion (BIC) to decide which fitted the data better (because of the different number of fixed effects between the two models). While the 2×2 design lends itself to the model with categorical predictors, our preference was to use the model with continuous predictors because this allows easier comparisons between models for the first, second, and later saccades within Study 3 and comparisons to the findings in Study 2. Thus, our preference was to report the model with continuous predictors for the first saccade as long as the BIC values indicated that this model was similar or better than the categorical predictors model (also see the Open Science Framework: https://osf.io/s5h8b/?view_only=58768b126db0439e946a49d94403d3a6 for distributions of salience in the different categories of salience and interest).

Results

Predicted first saccade latencies corresponded well with the observed data (Figure 6), with good statistical fits between predictions from the fitted LATER model parameters and the observed first saccade latencies and no participants for which there was a significant difference between the observed and predicted data (minimum $p = .381$). Thus, modeling saccade latencies as arising from a linear accumulation and evaluation of evidence offers a good explanation for first saccade latencies. Of the 66 participants, 22 showed statistical evidence for a subpopulation of first saccades arising from a maverick decision unit. For the remaining participants, the first saccade latencies appeared to all be generated by a single underlying decision mechanism (Figure 6).

For our analysis of the first saccade, we report the results of the LMM with salience and interest in foveal vision entered as continuous fixed effects for two reasons. First, because it provided a marginally better fit to the data (BIC = 15,602) than the model with these two fixed effects entered as two-level categorical predictors (BIC = 15,632). Second, because it allows better comparisons between models across the first, second, and later saccades in Study 3 and better comparisons to Study 2, which all used continuous predictors for salience and interest in foveal vision.

The LMM for the first saccade (Table 5)[4] showed that the timing of the first decision to move the eyes in Study 3 was influenced by visual information in foveal vision, with slower decisions to move when foveal vision contained higher edge information, higher salience or higher semantic interest (Figure 7, solid lines). Visual information at the peripheral location targeted by the first saccade also influenced the time course of the first decision to move the eyes, with the decision being slower as semantic interest at the target location increased. The effect of semantic interest at the target location was qualified by an interaction with the distance to this location in peripheral vision.

The pattern of effects found for the first saccade was very similar to that found for the second and later saccades (Table 5), with the exceptions that salience in foveal vision did not significantly predict decision time for the second saccade, and that higher salience at the target location predicted shorter decision times to move the eyes. While semantic interest at the peripheral target location significantly predicted decision rate for all saccades, the direction of effect was different for the first saccade than the second or later saccades: Higher interest slowed decision rate for the first saccade but increased decision rate for the second and later saccades (Figure 7F).

To explore the similarities and differences between the effects of the visual predictors for the first and second saccades, an LMM was run in which saccade number (first or second) was included as a sum-coded categorical fixed effect that interacted with each visual predictor. As in Study 2, this model was simplified by removing the effects associated with outgoing saccade amplitude, and a Bonferroni-corrected α of .025 was used to reflect the fact that this model retested data tested in the models above. The only significant interaction involving saccade number was with interest at the target location (Table 6), confirming the clear difference in effects between the first and second saccades evident in Figure 7. There was no evidence from this LMM that effects of any of the other predictors associated with visual information in foveal or peripheral vision differed between the first and second saccades.

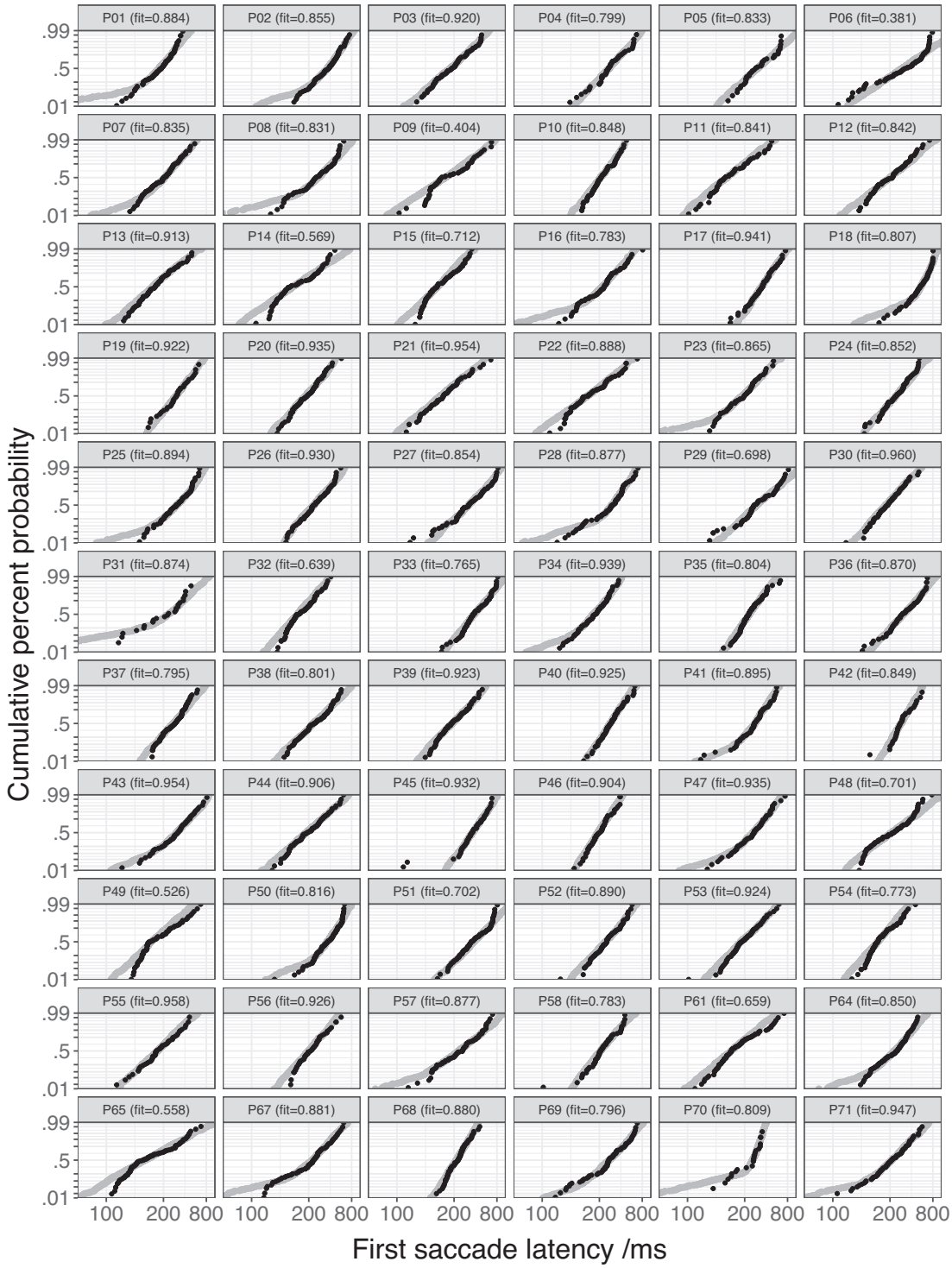
Discussion

The findings from Study 3 suggest that when scene viewing began from a noncentral position at which information content in foveal

[4] We also ran these models using AWS as the salience model instead of RARE2012 to ensure that the results were not parochial to this choice of salience model, and because the evaluation study used AWS for selecting starting locations. The pattern of effects was very similar when using AWS with the following differences: Saccade 1: the effect of salience at fixation was not significant using AWS; Saccade 2: the effects of edges at fixation and salience at the target location were not significant using AWS; Later saccades: No differences between models using RARE2012 and AWS.

Figure 6

Reciprobit Plots of First Saccade Latency in Study 3



Note. For each participant, the observed data (black points) are plotted over the predictions from the fitted LATER model (gray points) and the fit between the predicted and observed latencies (expressed as the Kolmogorov–Smirnov p value) is indicated in the title of each plot. Higher values denote better fits. P = participant; LATER = linear approach to threshold with ergodic rate.

Table 5
Results of LMM to Predict Decision Rate in Study 3 for the First, Second, and Later Saccades in Viewing

	First saccade				Second saccade				Later saccades			
Fixed effect	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Intercept	4.37	0.118	36.99	***	4.41	0.140	31.40	***	3.28	0.074	44.38	***
Outgoing saccade amplitude (linear)	17.10	1.391	12.29	***	1.50	1.477	1.02	ns	−3.83	1.043	−3.68	***
Outgoing saccade amplitude (quadratic)	−10.66	1.326	−8.04	***	−5.62	1.469	−3.82	***	−2.34	1.094	−2.14	*
Edge information at fixation	−0.05	0.023	−2.29	*	−0.06	0.026	−2.22	*	−0.01	0.006	−2.34	*
Edge information at target location	0.04	0.021	1.88	ns	0.03	0.026	1.26	ns	0.01	0.006	2.40	*
x Outgoing saccade amplitude (linear)	0.73	1.216	0.60	ns	1.04	1.484	0.70	ns	0.01	1.091	< 0.01	ns
x Outgoing saccade amplitude (quadratic)	1.36	1.138	1.20	ns	−0.69	1.507	−0.46	ns	0.17	1.165	0.15	ns
Salience at fixation	−0.07	0.026	−2.58	**	−0.03	0.033	−1.00	ns	−0.01	0.007	−1.94	ns
Salience at target location	0.04	0.024	1.77	ns	0.07	0.029	2.28	*	0.00	0.006	−0.69	ns
x Outgoing saccade amplitude (linear)	−2.25	1.366	−1.65	ns	−1.85	1.634	−1.13	ns	−2.32	1.223	−1.90	ns
x Outgoing saccade amplitude (quadratic)	−1.97	1.291	−1.53	ns	1.70	1.498	1.14	ns	0.94	1.347	0.70	ns
Semantic interest at fixation	−0.28	0.021	−12.94	***	−0.32	0.028	−11.28	***	−0.12	0.006	−20.31	***
Semantic interest at target location	−0.08	0.021	−3.90	***	0.23	0.026	8.74	***	0.04	0.006	6.93	***
x Outgoing saccade amplitude (linear)	−4.02	1.257	−3.20	***	−4.34	1.680	−2.58	**	−2.73	1.204	−2.26	*
x Outgoing saccade amplitude (quadratic)	1.93	1.224	1.58	ns	1.66	1.631	1.02	ns	−0.27	1.284	−0.21	ns
Random effect		*N*	Variance			*N*	Variance			*N*	Variance	
Participants (intercept)		66	0.073			65	0.099			66	0.004	
Scenes (intercept)		120	0.859			120	1.174			120	0.356	
Residual			1.332				1.795				0.947	
Log-likelihood			−7,724.6				−7,094.7				−60,875.0	
Conditional R^2/marginal R^2			0.096/0.468				0.044/0.441				0.011/0.283	
Deviance			15,449.3				14,189.4				121,749.9	
AIC			15,485.3				14,225.4				121,785.9	
$N_{observations}$			4,820				4,041				43,575	

Note. LMM = linear mixed model; ns = not significant; AIC = Akaike information criterion.
* $p < .05$. ** $p < .01$. *** $p < .001$.

vision was manipulated orthogonally by aligning the location from which viewing was initiated with locations of high or low salience and high or low semantic interest, there were clear effects of visual information in foveal vision. Saccades were delayed when foveal vision contained high edge information, high salience or high semantic interest. The effect of interest in foveal vision on how long it takes to decide to move the eyes was consistent with previous studies that have shown that objects that are highly informative—for example by being a semantic outlier in the scene—are looked at for longer (De Graef et al., 1990; Henderson et al., 1999). The effects of edges and interest in foveal vision on the first decision to move the eyes in Study 3 were the same as those for later saccades in Study 3 and for later saccades in viewing found by Tatler et al. (2017).

Information in peripheral vision also contributed to the timing of the first saccade: Higher semantic interest at the peripheral target location was associated with lower decision rates, meaning that it took longer to initiate saccades to peripheral locations that were high in semantic interest. Processing semantic information in peripheral vision prior to the first saccade in viewing is consistent with findings from studies of how people search natural scenes, where there is evidence that the first saccade is targeted toward locations of high semantic interest (Nyström & Holmqvist, 2008), and regions where the search target is likely to occur (Spotorno et al., 2014, 2015; Torralba et al., 2006).

For the second and later saccades, the effect of semantic interest was opposite to that for the first saccade: Higher semantic interest at the peripheral target location was associated with higher decision rates, meaning that saccades were launched sooner to semantically interesting locations. This direction of effect for semantic interest is as was found in Tatler et al. (2017).

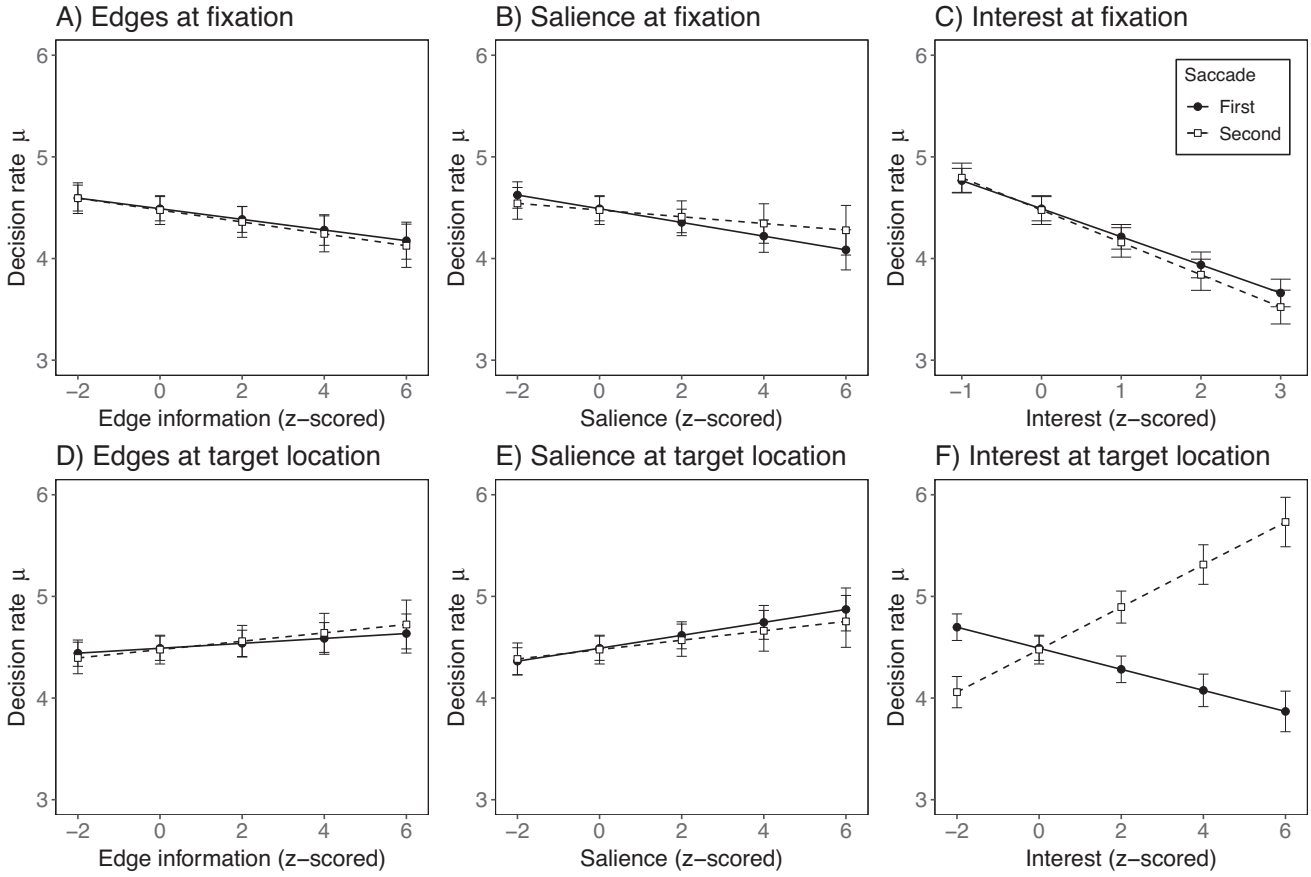
The change in direction of effect for semantic interest from the first to second (and later) saccades may reveal the manner in which semantic information about a scene is extracted and used to scaffold scene exploration. The delay in first saccade execution when selecting more semantically interesting locations may reflect longer processing time needed to extract higher level understanding from peripheral vision when the scene first appears. If this is the reason for the direction of effect observed for the first fixation in Study 3, the implication is that for later saccades in viewing, where high semantic information in peripheral vision speeds rather than slows decision times, semantic processing in peripheral vision must benefit from processing that has occurred during previous fixations and the developing semantic understanding of the scene being viewed. The finding that decision times were speeded by peripheral semantic interest even for the second saccade in viewing (in both Studies 2 and 3) shows that sufficient semantic information about the scene must have been extracted by the time the first saccade is executed that even the processing in the build up to the second saccade benefits from the initial semantic processing of the scene before the first saccade.

General Discussion

The aim of this work was to consider whether the initial processing of a scene involves different underlying mechanisms or priorities for information processing than later, once overt exploration of the scene has begun. Current models of scene viewing fail to include

Figure 7

Partial Effects of the Six Visual Predictors on the Decision Rates of the First and Second Saccades in Viewing in Study 3



the initial processing of the scene that informs the first saccade, with a need to extend theoretical frameworks to encompass this initial scene processing.

Study 1 showed that across a range of data sets, the first saccade was made after a longer time than subsequent saccades. Thus, the first saccadic decision constitutes a distinct population of responses from later decisions to move the eyes. This may suggest fundamental differences in how information is extracted and used in the first moments of scene processing. However, data from Study 2 (and Study 3) argue against this possibility because the distributions of latencies for the first saccade are well fitted as arising from a process of linear accumulation and evaluation of evidence.

A second possibility for the differences in first saccade latencies is that while the underlying decision process is still a linear accumulation and evaluation of evidence, the mechanism for evaluating the evidence differs from that underlying later saccadic decisions. For example, while later saccades are well described as arising from a "stay-or-go" mechanism of evaluating the relative value of moving to a peripheral location over prolonging processing of the current foveal target (Tatler et al., 2017), this mechanism may not be used for initial processing of the scene. The results of Study 2 are consistent with this possibility, suggesting a mechanism in which only low-level information in peripheral vision is evaluated for the first saccadic decision, with no evaluation of information in foveal vision, or of semantic information in foveal or peripheral vision.

A different mechanism for the first decision compared to later in viewing is consistent with converging evidence that initial information processing in scenes differs from later processing. Unema et al. (2005) suggested that initial scene processing involves ambient processing of the scene, followed by more focal processing of scene content later in viewing (see also Velichkovsky et al., 2005). However, these authors did not suggest that the initial ambient processing was restricted to the period before the launch of the first saccade, and indeed they excluded this period from their modeling, instead describing processing over the first couple of seconds after overt exploration of the scene began. Other authors have suggested more specifically that the processing in the buildup to the first saccade may be distinct from that which follows the first saccade. Schütt et al. (2019) suggested that the first saccade is guided strongly by bottom-up, and mostly low-level, image features, but later saccades were more strongly influenced by high-level information. Mackay et al. (2012) also suggested a specific bottom-up mechanism for the first saccade, proposing that short latency saccades are the outcome of a fast, automatic saccade generation mechanism based on bottom-up visual salience and likely controlled by subcortical pathways, but with the qualification that longer latency first saccades are generated by processes based upon bottom-up salience weighted by top-down signals. Taken together with the findings from Study 2, these accounts appear consistent with the possibility that the mechanism underlying the first saccadic decision is different from that

Table 6
Results of LMM to Predict Decision Rate in Study 3 Over the First Two Saccades in Viewing

Fixed effect	*b*	*SE*	*t*	*p*
Intercept	4.50	0.113	39.78	***
Saccade number	−0.05	0.015	−3.44	**
Edge information at fixation	−0.05	0.018	−2.83	**
x Saccade number	0.02	0.018	1.07	ns
Edge information at target location	0.03	0.018	1.45	ns
x Saccade number	−0.01	0.016	−0.38	ns
Salience at fixation	−0.08	0.021	−3.71	***
x Saccade number	−0.01	0.020	−0.46	ns
Salience at target location	0.07	0.020	3.65	***
x Saccade number	0.00	0.018	0.24	ns
Semantic interest at fixation	−0.31	0.017	−18.23	***
x Saccade number	−0.01	0.017	−0.69	ns
Semantic interest at target location	0.05	0.018	2.59	*
x Saccade number	−0.17	0.016	−10.63	***
Random effect	***N***		**Variance**	
Participants (intercept)	66		0.081	
Scenes (intercept)	120		0.782	
Residual			1.866	
Log-likelihood			−15,553.1	
Conditional R^2/marginal R^2			0.054/0.353	
Deviance			31,106.3	
AIC			31,140.3	
$N_{observations}$			8,861	

Note. Significance is reported using Bonferroni-adjusted αs. LMM = linear mixed model; ns = not significant; AIC = Akaike information criterion.
* $p < .025$. ** $p < .005$. *** $p < .0005$.

underlying subsequent saccades and is based on accumulating and evaluating bottom-up, mainly low-level information in peripheral vision.

If the first saccade is produced by a different underlying mechanism from later saccades, this may be because early information processing goals differ from those later in viewing. Indeed, early processing must serve to understand the meaning of the scene (Biederman, 1972; Potter, 1975, 1976), its global properties (Oliva & Torralba, 2001), and the layout of the scene (Sanocki & Sulman, 2009), all of which serve to scaffold the overt exploration of the scene that follows. These initial information-gathering goals are not present after these initial moments of scene processing.

While the findings from Studies 1 and 2 are consistent with there being a different mechanism underlying the first saccadic decision, the findings from Study 3 challenge this possible conclusion. In Study 3, the initial view of the scene was manipulated by forcing participants to begin from a noncentral location (in contrast to Study 2 and the vast majority of prior studies of scene viewing). Furthermore, the information in foveal vision was manipulated such that it was either high or low in salience and high or low in semantic interest. The findings of Study 3 contrasted those for Study 2 in two key ways. First, information in foveal vision contributed to the timing of the first saccade in Study 3 but not in Study 2. Second, semantic interest in both foveal and peripheral vision contributed to the timing of the first saccade in Study 3 but not in Study 2. Moreover, the factors in foveal and peripheral vision that contributed to the timing of the first saccade in Study 3 were very similar to those that contributed to the timings of later saccades in Studies 2 and 3, consistent with what would be expected if they arose from the "stay-or-go" mechanism proposed by Tatler et al. (2017).

There are at least two possible reasons for the differences between the findings of Studies 2 and 3. First, it may be that the mere act of starting noncentrally changes the mechanisms that underlie saccadic decision making. Rothkegel et al. (2016) found that viewing behavior was different—in terms of spatial selection—when starting noncentrally at an experimentally controlled location than when starting centrally. They suggested that, at least in part, the differences in viewing behavior they observed could be explained by inhibitory tagging of the noncentral starting location, reducing the tendency to return to that location. If so, the difference between noncentral and central starting positions may arise because any inhibitory tagging of the center is masked by the strong tendency to fixate (and return to) the center of the screen during viewing, independently of visual information in the scene (Tatler, 2007). Thus, starting centrally may result in the recruitment of underlying mechanisms for scene processing that are not recruited when starting noncentrally.

Alternatively, starting centrally or noncentrally may lead to strategic differences in how information is prioritized in a common mechanism, rather than the recruitment of different underlying mechanisms. Trawiński et al. (2023) found that starting noncentrally reduced the probability that faces would be looked at during scene viewing. This suggests a change in priorities between low- and high-level information that persists throughout viewing. However, while this result is consistent with the possibility of a change in priorities for different types of information in the scene when starting in the center or noncentrally, the nature of the effect is different from our findings. Trawiński et al. (2023) found a lower priority for high-level information in peripheral vision (less likely to select a face in peripheral vision for subsequent inspection by foveal vision) when starting noncentrally, whereas in Study 3 we found that semantic interest in peripheral vision played an important role in saccadic decision making from the onset of scene viewing. Furthermore, we found no differences in the factors that contributed to saccade timing later in scene viewing when starting centrally (Study 2) or noncentrally (Study 3). Rather any strategic differences were restricted to the first decision to move the eyes.

In Study 3 (noncentral start), compared to Study 2 (central start), any strategic differences in information priority for the first saccade resulted in a greater prioritization of semantic information in foveal and peripheral vision. This may have arisen because in 50% of trials the starting location coincided with a location of high semantic interest for the scene, alerting the participants to the importance of semantic interest in where they began their scene exploration from. If this is the case, the implication is that starting in the center, where there is no expectation that the starting location may be important for scene viewing or may contain semantically interesting content, leads to a strategic deprioritization of semantic interest—and indeed of information in foveal vision more generally—when planning the first saccade.

That semantic information may be deprioritized strategically during the first saccades after scene onset has been shown by Hwang et al. (2011). These authors showed that while semantic information contributed to saccade targeting from the first saccade in a general scene inspection task, in a scene search task, this information was deprioritized during the first few saccades in favor of visual guidance related to the search target. Moreover, previous work has shown that whether observers expect a source of information in the scene to be important influences whether that information is prioritized. For example, Underwood and Foulsham (2006) found that highly salient objects were more likely to be fixated when participants viewed

scenes in preparation for a memory task than when they viewed the same scenes to search for a low-salience target. This finding suggests that whether or not salience was prioritized in peripheral vision for planning saccades depended on the task requirements.

The findings from Study 3 suggested that while semantic interest did influence even the timing of the first saccade in viewing, the manner in which it contributed differed from later saccadic decisions. Specifically, higher semantic information at the peripheral target location was associated with longer decision times for the first saccade but shorter decision times for subsequent saccades. This finding is as would be expected if semantic information about the scene is accumulated over the first moments of scene viewing to build up a semantic representation and understanding of the scene when it first appears, with this initially constructed semantic representation of the scene then underpinning subsequent semantic processing in peripheral vision during later scene exploration. Prior studies have shown that semantic information about a scene can be accumulated in the first tens of milliseconds after it appears (Biederman, 1972), that semantic information can be extracted in peripheral vision (Li et al., 2002; Underwood et al., 2008), and that peripheral semantic information can guide even the first saccade in viewing, increasing the probability that the first saccade targets scene regions in which objects are expected (Spotorno et al., 2014) and that the first saccade selects one of the most semantically informative (diagnostic) objects in the scene (Spotorno & Tatler, 2017). An initial glimpse of a scene prior to viewing—in the form of a brief preview of the scene which can be as short as 75 ms—can allow enough information to be extracted that it can be used to guide eye movements during subsequent search of the scene (Hillstrom et al., 2012; Võ & Henderson, 2010), suggesting that later search is scaffolded on the semantic information extracted during the first moments of scene processing. The possibility that semantic information builds up over time prior to launching the first saccade is supported by previous work that has shown that salience is less prominently involved in target selection for the first saccade in viewing when the saccade is launched after a longer latency than when it is launched after a shorter latency (Anderson et al., 2015; Mackay et al., 2012).

In Study 2, the first decision to move the eyes was not influenced by semantic interest in foveal or peripheral vision. However, the effects of semantic interest on the second saccade suggest that this information was extracted and represented during the buildup to the first saccade and available to scaffold even the second saccadic decision. The possibility that semantic information may be extracted and retained even though it is not used to inform the first saccadic decision is consistent with previous reports that scene semantics are extracted in an obligatory fashion when viewing scenes. For example, Greene and Fei-Fei (2014) found that scene semantics were encoded even when the scene was presented as a task-irrelevant background in a word classification task. Similarly, Cornelissen and Võ (2017) found that objects which were semantically inconsistent with the scene background were fixated for longer even when the semantics of the object and scene were task irrelevant, suggesting not only the obligatory encoding of scene semantics but also that scene semantics extracted early in viewing influence later timing of decisions to move the eyes.

If, as our data suggest, semantic information is extracted but not used for the first saccade in Study 2, but provides scene understanding that can later allow faster selection of semantically interesting content in peripheral vision, it seems that there is no need to suggest that the mechanisms underlying the first decision to move the eyes are different from those underlying subsequent saccadic decisions. Rather, the differences between the first and later saccades arise from strategic differences in information prioritization—in the form of weighting information sources differently—within a common decision mechanism. The differences in the information influencing the first saccade found between Studies 2 and 3, therefore, are likely to arise from differences in expectations about what information should be prioritized for the first decision to move the eyes. When there is low expectation of important content at fixation, as is likely when viewing begins in the center of the scene on every trial (Study 2), information in foveal vision is deprioritized making no significant contribution to the time course of the first decision to move the eyes. This results in a decision process that appears to be based only on peripheral visual information. When viewing begins noncentrally, and important scene content frequently coincides with this starting location, information in foveal vision is prioritized and so the first decision to move the eyes appears to involve the evaluation of information in both foveal and peripheral vision. What differs in Study 3 between the first and later saccades is not the relative prioritization of different sources of information but the availability of semantic information, with the first saccade being made while a semantic representation of the scene is still under construction, whereas later saccades benefit from the semantic representation that is constructed during these early moments of scene viewing.

Strategic differences in the prioritization of information but a common mechanism for the first and later decisions in scene viewing provide an explanation for previous differences between published studies about whether and to what extent semantic information might guide the first saccade after scene onset. When the task emphasizes semantic information, such as searching for an object in a scene (Kanan et al., 2009; Spotorno et al., 2014; Torralba et al., 2006) semantics are prioritized and influence the first decision to move the eyes. However, when the task does not emphasize semantics, such as during free viewing, semantics are not prioritized and the first decision to move the eyes is dominated by other, lower level sources of information in peripheral vision (Study 2 and Anderson et al., 2015; Mackay et al., 2012).

Taken together the findings of three studies suggest that while the first saccadic decision in scene viewing appears to be qualitatively and quantitatively different from later decisions to move the eyes, involving a different time course and different use of information in foveal and peripheral vision, these differences are likely to arise from strategic differences in the priorities within a decision mechanism that is common across saccades. Moreover, these strategic differences can result not only in different priorities between tasks, but also in dynamic changes to priorities within the course of a single episode of viewing. Thus we can extend the theoretical framework proposed in LATEST (Tatler et al., 2017) by demonstrating that it is an appropriate framework for understanding even the first saccadic decision in viewing and by demonstrating the dynamic flexibility with which information sources are prioritized in the mechanisms that underlie each decision to move the eyes when viewing a scene.

References

Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2024).

Rmarkdown: Dynamic documents for R. https://github.com/rstudio/rmarkdown

Anderson, N. C., Donk, M., & Meeter, M. (2016). The influence of a scene preview on eye movement behavior in natural scenes. *Psychonomic Bulletin & Review*, *23*(6), 1794–1801. https://doi.org/10.3758/s13423-016-1035-4

Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, *15*(5), Article 9. https://doi.org/10.1167/15.5.9

Aust, F., & Barth, M. (2023). *Papaja: Prepare American Psychological Association journal articles with R markdown*. https://github.com/crsh/papaja

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2024). *lme4: Linear mixed-effects models using Eigen and S4*. https://github.com/lme4/lme4/

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*(4043), 77–80. https://doi.org/10.1126/science.177.4043.77

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. https://doi.org/10.1080/13506280500165188

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art*. University Chicago Press.

Carpenter, R. H. S. (1981). Oculomotor procrastination. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements* (pp. 237–246). Lawrence Erlbaum: Eye Movements: Cognition and Visual Perception.

Carpenter, R. H. S. (1994). SPIC: A PC-based system for rapid measurement of saccadic responses. *Journal of Physiology (Proceedings)*, *480*(2), Article 4P.

Carpenter, R. H. S. (1999). A neural mechanism that randomises behaviour. *Journal of Consciousness Studies*, *6*(1), 13–22.

Carpenter, R. H. S., & McDonald, S. A. (2007). LATER predicts saccade latency distributions in reading. *Experimental Brain Research*, *177*(2), 176–183. https://doi.org/10.1007/s00221-006-0666-5

Carpenter, R. H. S., Reddi, B., & Anderson, A. (2009). A simple two-stage model predicts response time distributions. *The Journal of Physiology*, *587*(16), 4051–4062. https://doi.org/10.1113/jphysiol.2009.173955

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3), Article 6. https://doi.org/10.1167/9.3.6

Castelhano, M. S., & Pereira, E. J. (2018). The influence of scene context on parafoveal processing of objects. *Quarterly Journal of Experimental Psychology*, *71*(1), 229–240. https://doi.org/10.1080/17470218.2017.1310263

Clarke, A. D. F., Green, P., Chantler, M. J., & Hunt, A. R. (2016). Human search for a target on a textured background is consistent with a stochastic model. *Journal of Vision*, *16*(7), Article 4. https://doi.org/10.1167/16.7.4

Cornelissen, T. H., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, *79*(1), 154–168. https://doi.org/10.3758/s13414-016-1203-7

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*(4), 317–329. https://doi.org/10.1007/BF00868064

Dolmen, P. (1919). Tests for determining the sighting eye. *American Journal of Ophthalmology*, *2*(12), Article 867. https://doi.org/10.1016/S0002-9394(19)90258-3

Einhäuser, W., Atzert, C., & Nuthmann, A. (2020). Fixation durations in natural scene viewing are guided by peripheral scene content. *Journal of Vision*, *20*(4), Article 15. https://doi.org/10.1167/jov.20.4.15

Engbert, R., Trukenbrod, H. A., Barthelmé, S., & Wichmann, F. A. (2015). Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision*, *15*(1), Article 14. https://doi.org/10.1167/15.1.14

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), Article 10. https://doi.org/10.1167/7.1.10

Garcia-Diaz, A., Leborán, V., Fdez-Vidal, X. R., & Pardo, X. M. (2012). On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. *Journal of Vision*, *12*(6), Article 17. https://doi.org/10.1167/12.6.17

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Green, P., & MacLeod, C. J. (2023). *SIMR: Power analysis for generalised linear mixed models by simulation*. https://github.com/pitakakariki/simr

Greene, M. R., & Fei-Fei, L. (2014). Visual categorization is automatic and obligatory: Evidence from Stroop-like paradigm. *Journal of Vision*, *14*(1), Article 14. https://doi.org/10.1167/14.1.14

Henderson, J. M. (1992). Identifying objects across saccades: Effects of extrafoveal preview and flanker object context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 521–530. https://doi.org/10.1037/0278-7393.18.3.521

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, *1*(10), 743–747. https://doi.org/10.1038/s41562-017-0208-0

Henderson, J. M., Hayes, T. R., Peacock, C. E., & Rehrig, G. (2019). Meaning and attentional guidance in scenes: A review of the meaning map approach. *Vision*, *3*(2), Article 19. https://doi.org/10.3390/vision3020019

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

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. https://doi.org/10.1037/0096-1523.25.1.210

Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, *19*(2), 204–210. https://doi.org/10.3758/s13423-011-0205-7

Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, *51*(10), 1192–1205. https://doi.org/10.1016/j.visres.2011.03.010

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*(6–7), 979–1003. https://doi.org/10.1080/13506280902771138

Kaspar, K., & Koenig, P. (2011). Viewing behavior and the impact of low-level image properties across repeated presentations of complex scenes. *Journal of Vision*, *11*(13), Article 26. https://doi.org/10.1167/11.13.26

Kolmogoroff, A. (1941). Confidence limits for an unknown distribution function. *The Annals of Mathematical Statistics*, *12*(4), 461–463. https://doi.org/10.1214/aoms/1177731684

Krasich, K., Huffman, G., Faber, M., & Brockmole, J. R. (2020). Where the eyes wander: The relationship between mind wandering and fixation allocation to visually salient and semantically informative static scene content. *Journal of Vision*, *20*(9), Article 10. https://doi.org/10.1167/jov.20.9.10

Krasich, K., Kim, J., Huffman, G., Klaffehn, A. L., & Brockmole, J. R. (2021). Does task-irrelevant music affect gaze allocation during real-world scene viewing? *Psychonomic Bulletin & Review*, *28*(6), 1944–1960. https://doi.org/10.3758/s13423-021-01947-4

Krasovskaya, S., & MacInnes, W. J. (2019). Salience models: A computational cognitive neuroscience review. *Vision*, *3*(4), Article 56. https://doi.org/10.3390/vision3040056

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2020). *lmerTest: Tests in linear mixed effects models*. https://github.com/runehaubo/lmerTestR

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

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

Mackay, M., Cerf, M., & Koch, C. (2012). Evidence for two distinct mechanisms directing gaze in natural scenes. *Journal of Vision*, *12*(4), Article 9. https://doi.org/10.1167/12.4.9

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

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, *11*(8), Article 17. https://doi.org/10.1167/11.8.17

Noorani, I. (2014). LATER models of neural decision behavior in choice tasks. *Frontiers in Integrative Neuroscience*, *8*, Article 67. https://doi.org/10.3389/fnint.2014.00067

Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, *64*, 229–251. https://doi.org/10.1016/j.neubiorev.2016.02.018

Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 342–360. https://doi.org/10.1037/a0033854

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., & Clark, C. N. (2023). Pseudoneglect during object search in naturalistic scenes. *Experimental Brain Research*, *241*(9), 2345–2360. https://doi.org/10.1007/s00221-023-06679-6

Nuthmann, A., De Groot, F., Huettig, F., & Olivers, C. N. (2019). Extrafoveal attentional capture by object semantics. *PLoS ONE*, *14*(5), Article e0217051. https://doi.org/10.1371/journal.pone.0217051

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

Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing-both initially and overall. *Journal of Eye Movement Research*, *2*(2), 1–11. https://doi.org/10.16910/jemr.2.2.2

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175. https://doi.org/10.1023/A:1011139631724

Onat, S., Açık, A., Schumann, F., & König, P. (2014). The contributions of image content and behavioral relevancy to overt attention. *PLoS ONE*, *9*(4), Article e93254. https://doi.org/10.1371/journal.pone.0093254

Ossandón, J. P., Onat, S., & König, P. (2014). Spatial biases in viewing behavior. *Journal of Vision*, *14*(2), Article 20. https://doi.org/10.1167/14.2.20

Over, E., Hooge, I., Vlaskamp, B., & Erkelens, C. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research*, *47*(17), 2272–2280. https://doi.org/10.1016/j.visres.2007.05.002

Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, & Psychophysics*, *73*(4), 1120–1132. https://doi.org/10.3758/s13414-011-0090-1

Pannasch, S., & Velichkovsky, B. M. (2009). Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. *Visual Cognition*, *17*(6–7), 1109–1131. https://doi.org/10.1080/13506280902764422

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123. https://doi.org/10.1016/S0042-6989(01)00250-4

Peacock, C. E., Singh, P., Hayes, T. R., Rehrig, G., & Henderson, J. M. (2023). Searching for meaning: Local scene semantics guide attention during natural visual search in scenes. *Quarterly Journal of Experimental Psychology*, *76*(3), 632–648. https://doi.org/10.1177/17470218221101334

Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965–966. https://doi.org/10.1126/science.1145183

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509–522. https://doi.org/10.1037/0278-7393.2.5.509

Ratcliff, R. (2001). Putting noise into neurophysiological models of simple decision making. *Nature Neuroscience*, *4*(4), 336–337. https://doi.org/10.1038/85956

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.3.3) [Computer science]. R Foundation for Statistical Computing. https://www.R-project.org/

Riche, N., Mancas, M., Duvinage, M., Mibulumukini, M., Gosselin, B., & Dutoit, T. (2013). Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, *28*(6), 642–658. https://doi.org/10.1016/j.image.2013.03.009

Roos, J. C. P., Calandrini, D. M., & Carpenter, R. H. S. (2008). A single mechanism for the timing of spontaneous and evoked saccades. *Experimental Brain Research*, *187*(2), 283–293. https://doi.org/10.1007/s00221-008-1304-1

Rothkegel, L. O., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., & Engbert, R. (2016). Influence of initial fixation position in scene viewing. *Vision Research*, *129*, 33–49. https://doi.org/10.1016/j.visres.2016.09.012

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1–3), 157–173. https://doi.org/10.1007/s11263-007-0090-8

Sanocki, T., & Sulman, N. (2009). Priming of simple and complex scene layout: Rapid function from the intermediate level. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 735–749. https://doi.org/10.1037/a0013032

Schütt, H. H., Rothkegel, L. O., Trukenbrod, H. A., Engbert, R., & Wichmann, F. A. (2019). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, *19*(3), Article 1. https://doi.org/10.1167/19.3.1

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, *14*(2), Article 7. https://doi.org/10.1167/14.2.7

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes. *Journal of Vision*, *15*(2), Article 12. https://doi.org/10.1167/15.2.12

Spotorno, S., & Tatler, B. W. (2017). The elephant in the room: Inconsistency in scene viewing and representation. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(10), 1717–1743. https://doi.org/10.1037/xhp0000456

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

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, *45*(5), 643–659. https://doi.org/10.1016/j.visres.2004.09.017

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

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *The Quarterly Journal of Experimental Psychology Section A*, *58*(5), 931–960. https://doi.org/10.1080/02724980443000430

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

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. https://doi.org/10.1037/0033-295X.113.4.766

Trawiński, T., Mestry, N., & Donnelly, N. (2023). The effect of prior viewing position and spatial scale on the viewing of paintings. *Vision*, *7*(3), Article 55. https://doi.org/10.3390/vision7030055

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

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, *59*(11), 1931–1949. https://doi.org/10.1080/17470210500416342

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition*, *17*(1), 159–170. https://doi.org/10.1016/j.concog.2006.11.008

Unema, P. J., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494. https://doi.org/10.1080/13506280444000409

Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). *Two visual systems and their eye movements: Evidence from static and dynamic scene perception* [Conference session]. Proceedings of the XXVII Conference of the Cognitive Science Society, Mahwah, NJ (pp. 2283–2288). https://escholarship.org/uc/item/9nv0c661

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, *10*(3), Article 14. https://doi.org/10.1167/10.3.14

Walshe, R. C., & Nuthmann, A. (2021). A computational dual-process model of fixation-duration control in natural scene viewing. *Computational Brain & Behavior*, *4*(4), 463–484. https://doi.org/10.1007/s42113-021-00111-4

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & van den Brand, T. (2024). *ggplot2: Create elegant data visualisations using the grammar of graphics*. https://ggplot2.tidyverse.org

Wilming, N., Onat, S., Ossandón, J. P., Açık, A., Kietzmann, T. C., Kaspar, K., Gameiro, R. R., Vormberg, A., & König, P. (2017a). An extensive dataset of eye movements during viewing of complex images. *Scientific Data*, *4*(1), Article 160126. https://doi.org/10.1038/sdata.2016.126

Wilming, N., Onat, S., Ossandón, J. P., Açık, A., Kietzmann, T. C., Kaspar, K., Gameiro, R. R., Vormberg, A., & König, P. (2017b). *Data from: An extensive dataset of eye movements during viewing of complex images*. Dryad. https://doi.org/10.5061/dryad.9pf75

Received August 19, 2024
Revision received February 7, 2025
Accepted March 16, 2025 ■