



**HAL**  
open science

## Prior knowledge about events depicted in scenes decreases oculomotor exploration

Marek A Pedziwiatr, Sophie Heer, Antoine Coutrot, Peter Bex, Isabelle  
Mareschal

### ► To cite this version:

Marek A Pedziwiatr, Sophie Heer, Antoine Coutrot, Peter Bex, Isabelle Mareschal. Prior knowledge about events depicted in scenes decreases oculomotor exploration. *Cognition*, 2023, 238, pp.105544. 10.1016/j.cognition.2023.105544 . hal-04154195

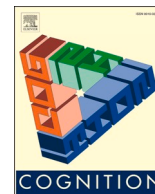
**HAL Id: hal-04154195**

**<https://hal.science/hal-04154195v1>**

Submitted on 6 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Prior knowledge about events depicted in scenes decreases oculomotor exploration

Marek A. Pedziwiatr<sup>a,\*</sup>, Sophie Heer<sup>a</sup>, Antoine Coutrot<sup>b</sup>, Peter Bex<sup>c</sup>, Isabelle Mareschal<sup>a</sup>

<sup>a</sup> School of Biological and Behavioural Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom

<sup>b</sup> Univ Lyon, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69621 Lyon, France

<sup>c</sup> Department of Psychology, Northeastern University, 107 Forsyth Street, Boston, MA 02115, United States of America

### ARTICLE INFO

#### Keywords:

Eye movements  
Visual perception  
Prior knowledge  
Scene viewing

### ABSTRACT

The visual input that the eyes receive usually contains temporally continuous information about unfolding events. Therefore, humans can accumulate knowledge about their current environment. Typical studies on scene perception, however, involve presenting multiple unrelated images and thereby render this accumulation unnecessary. Our study, instead, facilitated it and explored its effects. Specifically, we investigated how recently-accumulated prior knowledge affects gaze behavior. Participants viewed sequences of static film frames that contained several 'context frames' followed by a 'critical frame'. The context frames showed either events from which the situation depicted in the critical frame naturally followed, or events unrelated to this situation. Therefore, participants viewed identical critical frames while possessing prior knowledge that was either relevant or irrelevant to the frames' content. In the former case, participants' gaze behavior was slightly more exploratory, as revealed by seven gaze characteristics we analyzed. This result demonstrates that recently-gained prior knowledge reduces exploratory eye movements.

### 1. Introduction

Human experience of visual-information intake is continuous in time (Wurtz, Joiner, & Berman, 2011) and encompasses the ongoing transition from the past to the present (Hogendoorn, 2022). At any given moment, information about recent events is necessary for individuals to control their behavior (Tatler & Land, 2011). While some of that information might remain available in the environment (for example, in the form of results of completed actions), much of it must be remembered. Here, we recorded people's eye movements to investigate how the memory of the recent-past, or – in other words – recently-gained prior knowledge, influences humans' ongoing exploration of their surroundings.

While it is known that recent experiences influence the processing of previously unviewed static stimuli across a range of tasks (Awh, Belopolsky, & Theeuwes, 2012; Fischer & Whitney, 2014; Horstmann, 2015; Loschky, Larson, Smith, & Magliano, 2020), their effects within the context of the perception of novel, complex naturalistic scenes remain underexplored. This is because in the vast majority of studies that involve viewing scenes and/or performing tasks related to them, each scene constitutes a separate, self-contained trial, independent of the

other trials (see Tatler, Hayhoe, Land, & Ballard, 2011 and Dorr, Martinetz, Gegenfurtner, & Barth, 2010 for further critique of such studies). Consequently, such experimental settings inevitably render the accumulation of information across different, consecutively viewed scenes unnecessary. Yet, in the real world, where the visual input reaching human eyes usually has a continuous temporal structure and carries information about events that unfold over time, the accumulation of information is vital (Fairhall, Albi, & Melcher, 2014; Harrison, Bestmann, Rosa, Penny, & Green, 2011). Therefore, most laboratory-based studies involving image viewing do not tap into this process despite its importance and ubiquity in everyday life (with Hutson, Chandran, Magliano, Smith, & Loschky, 2022 being a notable exception). The goal of the present study was to address this gap and investigate how the information about recent past events, accumulated while observing their temporal development, influences human eye-movements.

In this preregistered study, we investigated whether identical images of scenes, presented under identical task conditions, can elicit different viewing behaviors depending on the content of the images immediately preceding them. All our experiments followed a similar design. In each trial, participants viewed a sequence of film frames extracted from films directed by Alfred Hitchcock. We created two experimental conditions.

\* Corresponding author.

E-mail address: [marek.pedziwi@gmail.com](mailto:marek.pedziwi@gmail.com) (M.A. Pedziwiatr).

In both conditions, the final frames in the sequences, called the ‘critical frames’, were the same. The conditions differed only with respect to the frames presented prior to them, called the ‘context frames’. In one condition, the context frames were frames preceding the critical frame in the film from which it was extracted. In the other condition, the context frames came from a different film than the critical frame. Therefore, depending on the condition, the same critical frame depicted events that were either a natural continuation of the events from the previous frames or were unrelated to them. Consequently, in the latter condition, the prior knowledge about the recent past was useless for identifying image regions that do not need to be attended because they contain information that has already been acquired. We hypothesized that this would lead to more exploratory viewing behavior than in the former condition, where prior knowledge could guide the gaze. To preempt our findings, our hypothesis was confirmed. We found that identical stimuli (i.e. critical frames) were sampled differently, depending on the information participants acquired before viewing them. Specifically, we analyzed seven metrics characterizing oculomotor behavior and found differences related to the presence of relevant prior knowledge in six of them. This result demonstrates that prior knowledge about the recent past contributes to the control of ongoing oculomotor behavior.

**Open practices statement.** This study was preregistered. The pre-registration protocol is available from the Open Science Framework repository ([https://osf.io/et7mr/?view\\_only=6f86dc8211d845c7b2c09ef6f45baf64](https://osf.io/et7mr/?view_only=6f86dc8211d845c7b2c09ef6f45baf64)), together with the data from all experiments and a script for downloading our stimuli.

## 2. Experiment 1

Forty-eight participants viewed sequences of film frames (static images taken from films). In each N-frame sequence, frames from N to N-1 served as a ‘context’ for the last frame, called a critical frame (see Fig. 1). In one condition, the context frames came from the same film as the critical frame and depicted a continuous (coherent and uninterrupted) course of events that naturally preceded events from the critical frame (Continuous condition). In the other condition, the context frames also depicted a continuous course of events but, crucially, came from a different film and therefore were unrelated to the critical frame (Discontinuous condition). Each participant viewed all critical frames but – given that the conditions were counterbalanced between two groups of participants – half of the frames were presented to them in the Continuous condition and the other half in the Discontinuous condition. None of the critical frames were ever used as context frames and vice versa. Consequently, none of the frames were presented more than once

to the same participant. To ensure that participants remained attentive throughout the experiment, we presented an attention-check question after each critical frame (participants were asked: ‘Which description best matches the last image?’) and they had to select the correct description of the image they had just seen out of four descriptions presented, for example a) ‘men in a room with a woman holding a gun’, b) ‘a group of people in an elevator’, c) ‘people walking down a hallway’, d) ‘a group of men looking through a stack of documents’. Please refer to our Supplemental Materials for all questions used in our study. We hypothesized that in the Discontinuous condition, where prior knowledge about events from context frames was irrelevant to the content of the critical frame, participants would exhibit a more exploratory viewing behavior.

### 2.1. Method

**Participants and sample size.** Forty-eight participants (40 women, mean age: 22 years) took part in the experiment, which is approximately twice the number of participants used in three similar studies (Coco, Nuthmann, & Dimigen, 2020; End & Gamer, 2019; Walter & Bex, 2021). Participants were all over the age of 18 and were predominantly staff and students of Queen Mary University of London. They all reported having normal or corrected to normal vision, provided written informed consent prior to their participation, and received remuneration (course credits or cash) afterwards. If a participant responded correctly to fewer than 80% of the attention-check questions or if a large amount of their gaze data was missing, we excluded them and tested another person instead (there were 5 such cases, see *Excluding participants* section in Supplemental Materials for details). All our experimental procedures were approved by the Queen Mary University of London research ethics board (QMERC20.307) and the Institutional Review Board of North-eastern University (#14-09-16).

**Stimuli.** We used film frames from the “1000 frames of Hitchcock” website ([https://the.hitchcock.zone/wiki/1000\\_Frames\\_of\\_Hitchcock](https://the.hitchcock.zone/wiki/1000_Frames_of_Hitchcock)), which stores frames from all films directed by Alfred Hitchcock (1000 per film). The frames are spaced several seconds apart and were extracted using a semi-automatic procedure involving human inspection to ensure that they were not – according to the creator of the website – ‘blurred or looking odd’.

We manually selected 936 frames that constituted our stimuli set. 80 of them were the ‘critical frames’. Each was paired with two sequences of ‘context frames’ that presented different, but always consistent and logical, courses of events: Continuous and Discontinuous. Continuous context-frames were frames preceding the critical frame in a film from which it originated. Therefore, each critical frame was a natural

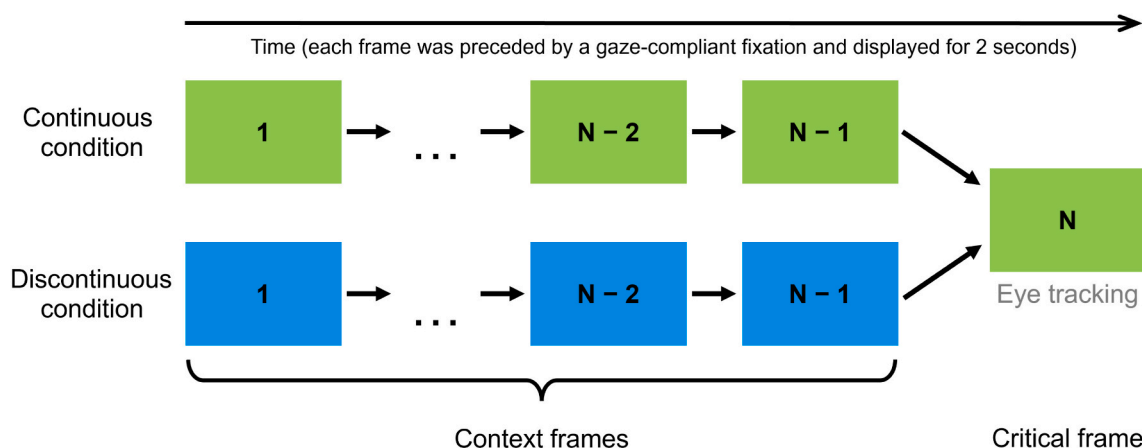


Fig. 1. Experimental design.

Note. Different colors indicate different films. The dots indicate that the number of context frames could differ between sequences (but for each different critical frame the number of context frames preceding it was identical in both conditions).

continuation of the context frames. Discontinuous context-frames came from a different film and were unrelated to the critical frame. Camera position and focus could change from frame to frame within a sequence. Context lengths varied across critical frames ( $M = 5.35$  frames,  $SD = 1.16$ ) but not within them: contexts paired with a given critical frame always had equal lengths in the two conditions. All frames were in color and had 16:9 aspect ratio. None of them contained violent or sexual content and none of the critical frames depicted a single person. We estimated that in the films from which the frames originated, the average time between two consecutive extracted frames amounted to 7.29 s ( $SD = 0.72$ ). The number of critical frames was constrained by the amount of available stimulus material. For each critical frame, we created four descriptions of its content, of which only one was correct. These served as response options in the attention-check question presented after each critical frame.

**Stimuli set validation.** We validated our stimuli set in two online experiments (see Supplemental Materials for details). Validation Experiment V1.1 confirmed that naïve participants could tell whether critical frames depicted a natural occurrence of events following from the context frames (for the Continuous contexts) or did not (for the Discontinuous contexts). Validation Experiment V1.2 confirmed that the attention-check questions were suitable (that is, sufficiently easy to answer). We iteratively modified our stimuli set (selected frames and question response options) and re-ran both experiments until, for each critical frame in both experiments, at least 8 out of 10 participants responded correctly.

**Procedure and design.** The experiment lasted approximately 45 minutes and consisted of four blocks of 20 trials, interleaved with breaks. In each trial, participants freely viewed a sequence of frames (context frames followed by a critical frame) and then responded to the attention-check question. Half of the sequences viewed by each participant belonged to the Continuous condition, half to the Discontinuous (counterbalanced between groups to which the participants were assigned using an ABAB schema). The assignment of sequences to blocks was randomized for each participant.

Before starting the experiment, participants received verbal and onscreen instructions that stated (a direct quote): ‘*You will be presented with a sequence of images. Please look at each of them carefully.*’ and were informed about the presence of the attention-check questions. Then, they completed a sample trial. Before each block, the eye tracker was calibrated. All frames were presented for two seconds and participants’ eye-movements were recorded during that time. Prior to each frame, a gaze-compliant fixation dot appeared centrally on the screen. Participants had to fixate within a 1.5 degrees’ radius from it for 750 ms for a frame to be displayed. All critical frames were followed by the attention-check question ‘*Which description best matches the last image?*’ and four possible answers, out of which one had to be selected by pressing a corresponding button (‘d’, ‘f’, ‘g’, or ‘h’, assigned ‘a’, ‘b’, ‘c’ or ‘d’ with a sticker corresponding to the response choice) on a keyboard.

**Apparatus.** The experiment took place in a dimly lit room. The frames were presented on a computer monitor against a grey background and had a height of 16 degrees of visual angle and a width of 28.4, which corresponded to 620 and 1102 pixels, respectively. Participants sat 55 cm from the monitor (iiyama Vision Master Pro 510; type: CRT; resolution: 1600 × 1200; refresh rate: 60 Hz; diagonal: 19 in.; brightness and contrast were adjusted to ensure comfortable viewing for prolonged times), with their heads stabilized in a chin rest. Their gaze was recorded using an infrared, video-based eye tracker Tobii 4C (90 Hz sampling rate). The procedure was programmed in MATLAB, using Psychophysics Toolbox Version 3 (Kleiner et al., 2007) and Tobii Pro Software Development Kit for Matlab version 1.7.1.4.

**Oculomotor metrics.** We examined the following seven metrics of oculomotor behavior: the number of fixations, average fixation duration (measured in milliseconds), average inter-fixation distance (Euclidean distance between subsequent fixation points measured in degrees of visual angle, akin to saccade amplitude), inter-observer consistency (the

similarity between the heatmap created from fixations of a given participant and the heatmap created from fixations of all remaining participants, determined by using a correlation between these heatmaps; Lyu et al., 2020), the probability of blinking (in our data, zero indicated no blinks, one – at least one blink), first-saccade latency (measured in milliseconds), and heatmap entropy (a measure of dispersion of all fixations registered on an image with values ranging from zero to one; Gameiro, Kaspar, König, Nordholt, & König, 2017). Apart from the entropy, that requires aggregating data from all participants viewing the same frame to calculate a single value, all metrics had unique values per participant per frame. The details of data pre-processing (including fixation-extraction procedure) and metrics calculation are provided in Supplemental Materials.

**Data analysis.** We used functions from the R (R Core Team, 2020) package lme4 (Bates, Mächler, Bolker, & Walker, 2015) to analyze how our metrics differed between the Continuous and Discontinuous conditions. Specifically, we fitted (generalized) linear mixed-effects models (GLMMs and LMMs) to the data. All our models included an intercept and a fixed effect of experimental condition. We used LMMs (fitted using function *lmer*) for all metrics but the probability of blinking, for which we used a GLMM (logistic regression model fitted using function *glmer*). Models’ random effects structures included random intercepts both per frame (for all metrics) and per participant (for all metrics but the entropy, for which the data were aggregated over participants).

To establish which metrics differed between conditions in a statistically significant fashion, we followed guidelines provided by Brown (2021). Specifically, we tested for which metrics our models provided better fits to the data than models that did not include the fixed effects of the experimental condition but were otherwise identical. To this end, we used likelihood-ratio tests implemented in R function *anova*. When reporting the results, we provide beta coefficients from our models and their standard errors, as well as the values of a test statistic  $\chi^2$  (always for one degree of freedom) and *p*-values from the likelihood-ratio tests. In our data, the Continuous condition was always coded as 0 and Discontinuous as 1. Therefore, in each LMM, beta coefficient for the intercept is an estimate of metrics value in the Continuous condition, while beta coefficient for the condition provides the estimate of difference between the two conditions.

## 2.2. Results

We found statistically significant effects of condition for five out of our seven oculomotor metrics (see Table 1 for full results, Fig. 2 for sample heatmaps created from fixations recorded in both conditions, and Fig. S1 in Supplemental Materials for plots). Specifically, we found that in the Discontinuous condition, the number of fixations increased, average fixation duration decreased, average inter-fixation distance increased, inter-observer consistency increased, and the probability of blinking decreased. The two metrics for which the differences between conditions failed to reach the threshold of statistical significance were first-saccade latency and heatmap entropy.

## 2.3. Discussion

In Experiment 1 we tested the hypothesis that eye movements performed while viewing a scene are influenced by knowledge acquired earlier (prior knowledge). To this end, we analyzed gaze patterns of participants viewing identical images (the critical frames) in two conditions: Continuous, in which events depicted in the critical frames naturally followed from the events shown in the ‘context’ frames presented before, and Discontinuous, in which there was no relationship between the critical frames and the context frames. We found that the conditions differed for five out of seven oculomotor characteristics that we analyzed. The directions of these differences suggested that, in the Continuous condition, participants explored the critical frames less extensively than in the Discontinuous condition.

**Table 1**  
Results of Experiment 1.

Metric	Effect direction	Beta and SE for intercept	Beta and SE for condition	$\chi^2(1)$	p-value
Number of fixations	↗	5.25 (0.15)	0.41 (0.04)	86.07	< 0.001 ***
Fixation duration (in milliseconds)	↘	308.41 (11.4)	-20.86 (3.8)	29.95	< 0.001 ***
Inter-fixation distance (in degrees of visual angle)	↗	4.35(0.18)	0.32 (0.05)	45.35	< 0.001 ***
Inter-observer consistency	↗	0.79 (0.01)	0.02 (< 0.01)	22.81	< 0.001 ***
Probability of blinking	↘	-1.05 (0.2)	-0.32 (0.08)	15.31	< 0.001 ***
First-saccade latency (in milliseconds)	-	323.23 (16.45)	-8.78 (5.05)	3.02	0.082 n. s.
Heatmap entropy	-	0.54 (0.01)	0.01 (< 0.01)	2.36	0.124 n. s.

Note. Asterisks in the last column indicate statistical significance and follow a standard convention. This convention is used in all tables in this article.

However, a potential concern regarding Experiment 1 is that each critical frame was followed by the attention-check question. Specifically, while the occurrence of the question could be anticipated in the Discontinuous condition – because it always appeared after a frame belonging to a different film than the previous frames in a sequence – it could not be anticipated in the Continuous condition, since in this condition all frames in a sequence belonged to the same film. Therefore, participants viewing critical frames in the Discontinuous condition could be changing their viewing strategy solely because they anticipated a question about the scene content afterwards. To exclude the possibility that this anticipation of the question fully explains the change in visual exploration, we conducted two more experiments. One is reported below as Experiment 2 (note that it was preregistered as Experiment 3), while the other experiment (preregistered as Experiment 2) is reported in Supplemental Materials as Experiment S1.

### 3. Experiment 2

Experiment 2 used the same the design as Experiment 1 except that the attention-check questions could appear after *any* frame, including the context frames. This ensured that the critical frames in the Discontinuous condition were no longer predictive of the occurrence of questions. This new design necessitated using different questions than in Experiment 1 (see below). The difference, if it had any noticeable effect at all, affected both conditions equally and therefore was unlikely to affect the differences between them (which were our primary output of

interest).

#### 3.1. Method

**Participants and sample size.** We collected data from 50 participants (two more than planned; 29 women, mean age: 20.22 years). 44 were undergraduate students of Northeastern University. The remaining 6 belonged to the participant pool from which the participants for Experiment 1 were recruited. All participants declared having normal or corrected to normal vision. For the participants recruited at Northeastern University, this was confirmed using standard tests of visual acuity (ETDRS, Ferris & Bailey, 1996), color vision (H-R-R, Hardy, Rand, & Rittler, 1954), and stereo acuity (Titmus, Ohlsson et al., 2001). All excluded participants (4 cases) were replaced; see *Excluding participants* section in Supplemental Materials for details.

**Stimuli.** We used the same 80 frame-sequences as in Experiment 1 and a novel set of attention-check questions. Specifically, we created 16 ‘generic’ multiple choice attention-check questions that could be asked after any frame and still have one unambiguously correct answer (for example: ‘Is this scene likely to be in a city?’; ‘yes’/‘no’). Their full list is available in the Supplemental Materials.

Next, we paired the questions with our sequences (by a ‘sequence’ we mean a critical frame and two equinumerous sets of context frames, used in our Continuous and Discontinuous conditions). First, of the 60 sequences not used in Experiment S1, we randomly selected 48 sequences and paired each of them with one question, that could randomly occur after any context frame. Therefore, each of the 16 unique questions was used with three different sequences (out of the 60). Note that although the position of that frame in the context was the same in both conditions (e.g. the question appeared after a third frame), the actual image (frame) differed between the Continuous and Discontinuous conditions. Second, we randomly selected 16 out of the remaining 20 sequences and paired each with a unique generic question. In these sequences, the question was always presented at the very end, after a critical frame. In total, we created 112 unique frame-question pairs ( $16 \times 3 \times 2 + 16$ ) which were the same across all participants. We validated these pairs in Validation Experiment V2.1, described in Supplemental Materials.

**Procedure and design.** Experiment 2 differed from Experiment 1 regarding the content of the questions and their placement within the procedure: here, they could also appear after context frames. Additionally, to make the procedure more engaging, we shortened the duration of the gaze-compliant fixation dot preceding all frames from 750 ms to 375 ms. Apart from these two changes, Experiment 2 was identical to Experiment 1.

**Apparatus.** The data were collected in two laboratories: at Queen Mary University of London, where the data for Experiment 1 were collected, and at Northeastern University in Boston. The setups in both laboratories were equivalent and used the same eye tracker model.

#### 3.2. Results

Data pre-processing and analysis are the same as in Experiment 1.



**Fig. 2.** Sample critical frame and the heatmaps of fixations registered on it in different conditions in Experiment 1. Note. Each heatmap illustrates fixations of 24 unique participants overlaid on the critical frame and smoothed using a Gaussian kernel – see details in Supplemental Material. Pixel values in both heatmaps were jointly normalized, so their color values are directly comparable. The frame was taken from the film Topaz (Hitchcock, 1969).

The results of Experiment 2 (see Table 2 and Fig. S2 in in Supplemental Materials) mirror the results of Experiment 1 for all metrics apart from two: the probability of blinking and first-saccade latency. Specifically, we found statistically significant differences between conditions for the number of fixations, average fixation duration, average inter-fixation distance, inter-observer consistency and first-saccade latency but not for the probability of blinking and heatmap entropy. Importantly, except for the entropy, all the metrics showed the same directions of differences between our conditions in both Experiments, irrespective of the statistical significance of these differences.

We analyzed aggregated data from both Experiments to check if they indeed differed qualitatively regarding the probability of blinking and first-saccade latency. To this end, we extended the (G) LMMs used for these metrics by including a main effect of experiment number (1, coded as 0, vs. 2, coded as 1) and an interaction between the experiment number and the experimental condition. Next, we used likelihood-ratio tests to check if these models provided better fit to the data than models without each of these effects. Whenever such test indicated that the initial model is better, this meant that the removed effect contributed to the quality of fit – this is how we determined the statistical significance of each effect. For the probability of blinking, all fixed effects were statistically significant (intercept:  $\beta = -1.06$ ,  $SE = 0.19$ ; condition:  $\beta = -0.32$ ;  $SE = 0.08$ ,  $\chi^2(2) = 15.78$ ,  $p < 0.001$ ; experiment:  $\beta = 0.41$ ,  $SE = 0.26$ ,  $\chi^2(2) = 10.34$ ,  $p = 0.006$ ; interaction:  $\beta = 0.27$ ,  $SE = 0.11$ ,  $\chi^2(1) = 5.72$ ,  $p = 0.017$ ), indicating that the drop in the probability of blinking between the Continuous and Discontinuous conditions was larger in Experiment 1 than in Experiment 2. For the first-saccade latency, we did not find statistically significant effects other than a decrease in the Discontinuous condition (intercept:  $\beta = 323.03$ ,  $SE = 14.18$ ; condition:  $\beta = -8.47$ ,  $SE = 5.21$ ;  $\chi^2(2) = 7.17$ ,  $p = 0.028$ ; experiment:  $\beta = -26.72$ ,  $SE = 18.06$ ,  $\chi^2(2) = 2.64$ ,  $p = 0.268$ ; interaction:  $\beta = -2.53$ ,  $SE = 7.34$ ,  $\chi^2(2) = 0.12$ ,  $p = 0.730$ ).

#### 4. Discussion

Our study investigated how prior knowledge about recent events affects gaze control. Participants viewed the same film frames, called the critical frames, while possessing prior knowledge either about events leading to situations depicted in these frames (in the Continuous condition), or about unrelated events (in the Discontinuous condition). We analyzed seven characteristics of eye movements registered on these frames and found differences between conditions for six of them (the number of fixations, average fixation duration, average inter-fixation

distance, inter-observer consistency, the probability of blinking, and first-saccade latency). These differences, although small, indicated that in the Continuous condition, visual processing was more focused on key elements, while in the Discontinuous, it was more exploratory. Our results therefore demonstrate the role of prior knowledge about the recent past in oculomotor control. Importantly, given that this knowledge is rapidly acquired (as our results show) and usually, albeit not always, short lived (Hannula, 2010; Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002), it is distinct from a general knowledge about the world which, for example, allows for identifying inconsistent objects within scenes, such as an octopus in a farmyard (Loftus & Mackworth, 1978; see also Smith, Loschky, & Bailey, 2021).

When comparing participants' gaze behavior between conditions, in the Discontinuous condition we observed a slight increase in the number of fixations and in the inter-fixation distance, as well as a small decrease in fixation duration. These changes in gaze behavior – apart from indicating that observers explored the scenes somewhat more – might indicate a shift in the mode of visual processing from focal (focused on examining restricted image regions in detail) to ambient (focused on grasping the overall layout of a scene). The distinction between these two processing modes (or at least viewing modes) has a history dating back to 1947 – see (Pannasch, Helmert, Roth, Herbold, & Walter, 2008) for a brief summary. Although the precise neural underpinnings of ambient and focal processing are still being debated (Marsman, 2013; Mills et al., 2017), the distinction itself has been demonstrated using tasks such as simulated driving (Velichkovsky, Rothert, Kopf, Dornhöfer, & Joos, 2002), solving Rubik's cube (Guo, Helmert, Graupner, & Pannasch, 2022), film viewing (Eisenberg & Zacks, 2016), and viewing scenes of different kinds (Pannasch et al., 2008; Unema, Pannasch, Joos, & Velichkovsky, 2005), as well as with computational modelling (Follet, Fontaine, & Meur, 2011; Le Meur & Fons, 2020). One well established behavioral finding related to this distinction is that when individuals view static scenes, their gaze behavior changes from ambient to focal over time (note that this is in opposite direction to the change we observed between our Continuous and Discontinuous conditions). For example, Pannasch et al. (2008) found that the first two seconds of scene viewing are characterized by more ambient processing than the seconds from four to six. The effects we found, although smaller, were obtained for images viewed for two seconds. Therefore, they extend the findings of Pannasch and colleagues by suggesting that the initial balance between ambient and focal processing is influenced by prior knowledge. Specifically, the initial processing is more focal if the prior knowledge is relevant to the viewed situation (see also Eisenberg & Zacks, 2016). The decrease of the first-saccade latencies we also observed in the Discontinuous condition – amounting to shorter first fixations – indicates that the effect we investigated emerges soon after stimulus onset. This result dovetails with the observation that a process happening within a similar time window, namely, gist extraction, also can be affected by prior knowledge provided to participants in a similar fashion as in our experiment (McLean, Nuthmann, Renoult, & Malcolm, 2023; see also Smith & Loschky, 2019).

Interestingly, heatmap entropy – a metric previously used to quantify the tendency to explore an image (Gameiro et al., 2017; Shiferaw, Downey, & Crewther, 2019) – did not differ between our experimental conditions. Entropy value for a heatmap depends on its spread, which, in turn, depends on fixated image regions. The lack of changes in entropy suggests that in both conditions, largely the same regions of the critical frames were fixated. This was probably because our frames, including the critical ones, usually contained only several key regions (e.g., faces) that attracted the majority of fixations irrespective of the condition. The existence of such regions, together with the observed increase in the number of fixations, likely explains another effect we found in the Discontinuous condition: the slight increase in inter-observer consistency. It is likely that in this condition more participants fixated all the key image regions, thus decreasing the overall variability among fixation patterns. Both these results highlight caveats in using film frames as

**Table 2**  
Results of Experiment 2.

Metric	Effect direction	Beta and SE for intercept	Beta and SE for condition	$\chi^2(1)$	p-value
Number of fixations	↗	5.19 (0.13)	0.22 (0.05)	22.70	< 0.001 ***
Fixation duration (in milliseconds)	↘	298.18 (9.67)	-12.47 (3.88)	10.31	0.001 **
Inter-fixation distance (in degrees of visual angle)	↗	4.41 (0.18)	0.2 (0.05)	14.62	< 0.001 ***
Inter-observer consistency	↗	0.74 (0.02)	0.03 (< 0.01)	34.32	< 0.001 ***
Probability of blinking	-	-0.64 (0.18)	-0.06 (0.07)	0.59	0.442 n. s.
First-saccade latency (in milliseconds)	↘	296.51 (11.75)	-11.54 (5.28)	4.77	0.029 *
Heatmap entropy	-	0.56 (0.01)	-0.01 (< 0.01)	1.79	0.181 n. s.

stimuli.

The last metric we analyzed was the probability of blinking. It decreased in the Discontinuous condition in both Experiments, suggesting heightened attention to the critical frames (Ranti, Jones, Klin, & Shultz, 2020; see also Nakano, Kato, Morito, Itoi, & Kitazawa, 2013). This decrease was larger in Experiment 1, where the critical frames in the Discontinuous condition were reliable predictors of the questions. The anticipation of questions likely strengthened the increase in attentiveness elicited by the critical frame alone, which was marked by the further decrease in the probability of blinking.

While our effects are interesting from a theoretical perspective, it is worth considering their practical significance too. Funder and Ozer (2019) proposed that when assessing effects in psychological experiments, there are at least two important factors to take into account. The first factor is the absolute size of the effect. The effects we found were small. There are, however, two caveats to consider here. First, they in part depended on the specifics of our stimuli, because at least some of the oculomotor characteristics we measured depend on stimuli content (Le Meur & Fons, 2020; Pannasch et al., 2008) and size (Gameiro et al., 2017; Marsman, 2013; Otero-Millan, Macknik, Langston, & Martinez-Conde, 2013). Second, establishing the relevance of effects found in laboratory-based eye tracking studies to real-world situations is challenging (Henderson, 2017; Tatler et al., 2011), and, therefore, it is hard to judge if the effects we observed would be correspondingly small in more ecologically valid settings. Nevertheless, even under the assumption that our effects are small, they could still have a practical significance because of the second factor proposed by Funder and Ozer (2019) – the frequency of the real-world situations in which a phenomenon of interest occurs. We investigated the influence of prior knowledge that is accumulated whenever humans view unfolding events on a behavior that is ubiquitous, namely, eye movements. Given how common these two phenomena are, even minor effects related to them are worthy of attention.

An intuitive explanation of our findings is that participants were surprised by the events shown in the critical frames only in the Discontinuous condition. This explanation has some merit, and we refer to the definitions of surprise to interpret our results. Although these definitions vary (Barto, Mirolli, & Baldassarre, 2013; Modirshanechi, Brea, & Gerstner, 2022; Reisenzein, Horstmann, & Schützwohl, 2019), many of them highlight two mutually non-exclusive cases when an event can be surprising. First, the event can be inconsistent with the individual's current mental model of the world. In our study, the context frames in both conditions provided participants with knowledge enabling them to build such mental models (Loschky et al., 2020; see also Smith, 2012): situations presented in them were changing in a sensible fashion. For example, a man sitting in a car in one frame was shown next to the car in a next frame, which indicated that he had left the vehicle. In the Discontinuous condition, the events in the critical frames were breaking the logical continuity of what was happening in the context frames. Therefore, these events were discrepant with the mental models that were based on the knowledge accumulated while viewing the context frames and were therefore surprising in that sense.

The second case in which an event can be called surprising is when it disproves previously generated predictions. The mental models described above can likely be sources of predictions about future events. For illustration, consider that subsequent frames often depicted the same characters in the same settings, so it was possible to predict the re-appearance of a character in subsequent frames. Moreover, the general familiarization with our stimuli could lead to expectations about what sort of events are likely to occur, and which are not. For example, the appearance of a man in a long, black coat was quite likely while the appearance of a giraffe was not. Therefore, each critical frame had the potential to either confirm predictions (to some acceptable degree) or disprove them, and whenever the latter was the case, the frame was surprising.

It is worth noting that the predictions generated in our procedure

were likely less specific than those investigated in many previous studies focusing specifically on predictions (Ekman, Kusch, & de Lange, 2023; Kok, Failing, & de Lange, 2014; Kok, Jehee, & de Lange, 2012) and related phenomena, such as surprise (Horstmann, 2015) and expectations (Amit, Abeles, Carrasco, & Yuval-Greenberg, 2019). Typically, these studies use a limited number of stimuli, and each stimulus is often paired with a specific cue that precedes it. For example, a seminal study by Kok et al. (2012) used gratings that were tilted either to the right or to the left and the tilt directions were paired with different auditory cues preceding grating presentation, leading to specific predictions about the upcoming visual input. In contrast, in our study the breadth of possible visual inputs was much broader which likely resulted in less specific predictions. In fact, we speculate that these predictions were even less specific than the ones typically generated in everyday situations. This is because in everyday situations, many changes in an individual's visual input result from self-motion (which is under the individual's control, e.g. during walking) and in consequence can be predicted more precisely than changes resulting from the motion of a camera (like in our study). Therefore, our experimental approach is complementary to the one used by Kok and colleagues. This is because these two approaches focus on predictions that are, respectively, more and less specific than the ones formed in everyday situations.

Prior knowledge about the recent past, which underpins the mental models that facilitate prediction generation, must be stored in memory. The relationship between oculomotor control and different types of memory is intricate and while we demonstrate one way in which it manifests itself, there are many more (for reviews, see Hannula, 2010; Henderson, 2017; Van der Stigchel & Hollingworth, 2018; Wynn, Shen, & Ryan, 2019). For example, semantically inconsistent objects in scenes (like an octopus in a farmyard) are looked at differently than consistent objects that are matched for visual features and scene location (Coco et al., 2020; Öhlschläger & Vö, 2017). While this effect relies on a general knowledge about the world stored in the long term memory (for example, the knowledge that octopuses do not belong farmyards), we observed effects that are driven by knowledge that has been acquired only recently. While it is likely that both these forms of knowledge are commonly used in conjunction to guide human gaze, the exact nature of their interaction remains poorly understood.

Another interesting question related to the relationship between memory and eye movements stems from the fact that our results mirror the results from two recent studies investigating this relationship (Cronin, Peacock, & Henderson, 2020; Walter & Bex, 2021). These studies investigated how viewing scenes under high memory load differs from viewing scenes under low load and found differences in gaze behavior that echo the differences we observed between the Continuous and Discontinuous conditions. Given this similarity in results, it is tempting to consider if they can be driven by the same mechanism. Specifically, it is possible that the balance between the ambient and focal viewing modes is influenced by the current memory load – the higher the load, the less ambient the viewing (see Eisenberg & Zacks, 2016 for a similar suggestion). Although our study – unlike the two aforementioned ones – did not manipulate memory load directly, it still could be changing in it. This is because the critical frames in the Discontinuous condition could signal – already at the early stages of processing – that the currently upheld mental model is no longer relevant and can be removed from memory. This, in turn, might have decreased the memory load and lead to more exploratory behavior.

In our experiments, constructing mental models was one of the consequences of accumulating knowledge about the unfolding events. These events were conveyed by the semantic content of the frames. In typical scenes such as film frames, this content is inseparable from the visual features of stimuli (Pedziwiatr, von dem Hagen, & Teufel, 2023) – for example, human silhouettes indicate the locations of specific characters. Therefore, it is necessary to consider our findings in the context of an interplay between the frame-to-frame changes in visual features and prior-knowledge accumulation. In everyday situations, as well as in

dynamic stimuli such as movie clips, visual features often change continuously because of the movement of scene elements such as objects or people. Motion strongly attracts eye movements (Itti, 2005; Mital, Smith, Hill, & Henderson, 2011; Roth, Rolfs, Hellwich, & Obermayer, 2023), and it has been suggested that in the presence of (object) motion, the knowledge accumulation is redundant for eye movements control (Wang, Freeman, Merriam, Hasson, & Heeger, 2012). Our stimuli were static and thereby rendered tracking any elements in a continuous fashion impossible (see a recent study by Hutson et al. (2022) for a similar approach). Nevertheless, our stimuli still did contain visual features that could be tracked and compared between the frames. Such tracking, however, is unlikely to fully explain our results. This is because several studies have demonstrated that even for dynamic stimuli in which continuous, motion-based tracking is possible, the narrative structure and semantic content of stimuli – which likely cannot be processed based on visual features alone (Pedziwiatr et al., 2023) – impact oculomotor control (Goettker, Pidaparthy, Braun, Elder, & Gegenfurtner, 2021; Kirkorian & Anderson, 2018; Loschky, Larson, Magliano, & Smith, 2015). Here, when such tracking was not possible, the reliance on the high-level factors must have been even stronger.

Taken together, we demonstrated that prior knowledge about recent past-events modulates the degree of ongoing oculomotor exploration during static-image viewing. Given that this type of knowledge is accumulated whenever individuals observe unfolding events, we contend that it is necessary to include it among factors routinely guiding gaze.

#### CRedit authorship contribution statement

**Marek A. Pedziwiatr:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sophie Heer:** Conceptualization, Data curation, Investigation, Methodology, Writing – review & editing. **Antoine Coutrot:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Peter Bex:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing. **Isabelle Mareschal:** Conceptualization, Funding acquisition, Methodology, Supervision, Project administration, Resources, Writing – review & editing.

#### Declaration of Competing Interest

None.

#### Data availability

Please see *Open practices statement* section.

#### Acknowledgments

This research was funded by a Leverhulme Trust grant (RPG-2020-024) to IM, PB, and AC.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2023.105544>.

#### References

- Amit, R., Abeles, D., Carrasco, M., & Yuval-Greenberg, S. (2019). Oculomotor inhibition reflects temporal expectations. *NeuroImage*, 184, 279–292. <https://doi.org/10.1016/j.neuroimage.2018.09.026>
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16(8), 437–443. <https://doi.org/10.1016/j.tics.2012.06.010>

- Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00907>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 251–264. <https://doi.org/10.18637/jss.v067.i01>
- Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, 4(1). <https://doi.org/10.1177/2515245920960351>, 2515245920960351.
- Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related brain potentials during semantic integration of object–scene information. *Journal of Cognitive Neuroscience*, 32(4), 571–589. [https://doi.org/10.1162/jocn\\_a.01504](https://doi.org/10.1162/jocn_a.01504)
- Cronin, D. A., Peacock, C. E., & Henderson, J. M. (2020). Visual and verbal working memory loads interfere with scene-viewing. *Attention, Perception, & Psychophysics*, 82(6), 2814–2820. <https://doi.org/10.3758/s13414-020-02076-1>
- Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10), 28. <https://doi.org/10.1167/10.10.28>
- Eisenberg, M. L., & Zacks, J. M. (2016). Ambient and focal visual processing of naturalistic activity. *Journal of Vision*, 16(2), 5. <https://doi.org/10.1167/16.2.5>
- Ekman, M., Kusch, S., & de Lange, F. P. (2023). Successor-like representation guides the prediction of future events in human visual cortex and hippocampus. *eLife*, 12, Article e78904. <https://doi.org/10.7554/eLife.78904>
- End, A., & Gamer, M. (2019). Task instructions can accelerate the early preference for social features in naturalistic scenes. *Royal Society Open Science*, 6(3). <https://doi.org/10.1098/rsos.180596>
- Fairhall, S. L., Albi, A., & Melcher, D. (2014). Temporal integration windows for naturalistic visual sequences. *PLoS One*, 9(7), Article e102248. <https://doi.org/10.1371/journal.pone.0102248>
- Ferris, F. L., & Bailey, I. (1996). Standardizing the measurement of visual acuity for clinical research studies. *Ophthalmology*, 103(1), 181–182. [https://doi.org/10.1016/S0161-6420\(96\)30742-2](https://doi.org/10.1016/S0161-6420(96)30742-2)
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, 17(5), 738–743. <https://doi.org/10.1038/nn.3689>
- Follet, B., Fontaine, B., & Meur, O. L. (2011). *New Insights into Ambient and Focal Visual Fixations using an Automatic Classification Algorithm*. 2 pp. 592–610.
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Gameiro, R. R., Kaspar, K., König, S. U., Nordholt, S., & König, P. (2017). Exploration and exploitation in natural viewing behavior. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-02526-1>
- Goettker, A., Pidaparthy, H., Braun, D. I., Elder, J. H., & Gegenfurtner, K. R. (2021). Ice hockey spectators use contextual cues to guide predictive eye movements. *Current Biology*, 31(16), R991–R992. <https://doi.org/10.1016/j.cub.2021.06.087>
- Guo, Y., Helmert, J. R., Graupner, S.-T., & Pannasch, S. (2022). Eye movement patterns in complex tasks: characteristics of ambient and focal processing. *PLoS One*, 17(11), Article e0277099. <https://doi.org/10.1371/journal.pone.0277099>
- Hannula, D. E. (2010). Worth a glance: Using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, 4. <https://doi.org/10.3389/fnhum.2010.00166>
- Hardy, L. H., Rand, G., & Rittler, M. C. (1954). H-R-R polychromatic plates\*. *Journal of the Optical Society of America*, 44(7), 509. <https://doi.org/10.1364/JOSA.44.000509>
- Harrison, L., Bestmann, S., Rosa, M. J., Penny, W., & Green, G. G. R. (2011). Time scales of representation in the human brain: Weighing past information to predict future events. *Frontiers in Human Neuroscience*, 5. <https://doi.org/10.3389/fnhum.2011.00037>
- Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, 21(1), 15–23. <https://doi.org/10.1016/j.tics.2016.11.003>
- Hitchcock, A. (1969). *Topaz [Motion Picture]*. Universal Pictures.
- Hogendoorn, H. (2022). Perception in real-time: Predicting the present, reconstructing the past. *Trends in Cognitive Sciences*, 26(2), 128–141. <https://doi.org/10.1016/j.tics.2021.11.003>
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 113–136. <https://doi.org/10.1037/0096-1523.28.1.113>
- Horstmann, G. (2015). The surprise–attention link: A review. *Annals of the New York Academy of Sciences*, 1339(1), 106–115. <https://doi.org/10.1111/nyas.12679>
- Hutson, J. P., Chandran, P., Magliano, J. P., Smith, T. J., & Loschky, L. C. (2022). Narrative comprehension guides eye movements in the absence of motion. *Cognitive Science*, 46(5). <https://doi.org/10.1111/cogs.13131>
- Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, 64(6), 882–895. <https://doi.org/10.3758/BF03196793>
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12(6), 1093–1123. <https://doi.org/10.1080/13506280444000661>
- Kirkorian, H. L., & Anderson, D. R. (2018). Effect of sequential video shot comprehensibility on attentional synchrony: A comparison of children and adults. *Proceedings of the National Academy of Sciences*, 115(40), 9867–9874. <https://doi.org/10.1073/pnas.1611606114>
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, 36(14).
- Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, 26(7), 1546–1554. [https://doi.org/10.1162/jocn\\_a.00562](https://doi.org/10.1162/jocn_a.00562)



- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270. <https://doi.org/10.1016/j.neuron.2012.04.034>
- Le Meur, O., & Fong, P.-A. (2020). Predicting image influence on visual saliency distribution: The focal and ambient dichotomy. In , 1–5. *ACM Symposium on Eye Tracking Research and Applications*. <https://doi.org/10.1145/3379156.3391362>
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4(4), 565–572. <https://doi.org/10.1037/0096-1523.4.4.565>
- Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLoS One*, 10(11), Article e0142474. <https://doi.org/10.1371/journal.pone.0142474>
- Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2020). The scene perception & event comprehension theory (SPECT) applied to visual narratives. *Topics in Cognitive Science*, 12(1), 311–351. <https://doi.org/10.1111/tops.12455>
- Lyu, M., Choe, K. W., Kardan, O., Kotabe, H. P., Henderson, J. M., & Berman, M. G. (2020). Overt attentional correlates of memorability of scene images and their relationships to scene semantics. *Journal of Vision*, 20(9), 2. <https://doi.org/10.1167/jov.20.9.2>
- Marsman. (2013). Linking cortical visual processing to viewing behaviour using fMRI. *Frontiers in Systems Neuroscience*. <https://doi.org/10.3389/fnsys.2013.00109>
- McLean, D., Nuthmann, A., Renoult, L., & Malcolm, G. L. (2023). Expectation-based gist facilitation: Rapid scene understanding and the role of top-down information. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0001363>
- Mills, M., Alwatban, M., Hage, B., Barney, E., Truemper, E. J., Bashford, G. R., & Dodd, M. D. (2017). Cerebral hemodynamics during scene viewing: Hemispheric lateralization predicts temporal gaze behavior associated with distinct modes of visual processing. *Journal of Experimental Psychology: Human Perception and Performance*, 43(7), 1291–1302. <https://doi.org/10.1037/xhp0000357>
- Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3(1), 5–24. <https://doi.org/10.1007/s12559-010-9074-z>
- Modirshanechi, A., Brea, J., & Gerstner, W. (2022). A taxonomy of surprise definitions. *Journal of Mathematical Psychology*, 110, Article 102712. <https://doi.org/10.1016/j.jmp.2022.102712>
- Nakano, T., Kato, M., Morito, Y., Itoi, S., & Kitazawa, S. (2013). Blink-related momentary activation of the default mode network while viewing videos. *Proceedings of the National Academy of Sciences*, 110(2), 702–706. <https://doi.org/10.1073/pnas.1214804110>
- Öhlschläger, S., & Vö, M. L.-H. (2017). SCEGRAM: An image database for semantic and syntactic inconsistencies in scenes. *Behavior Research Methods*, 49(5), 1780–1791. <https://doi.org/10.3758/s13428-016-0820-3>
- Ohlsson, J., Villarreal, G., Abrahamsson, M., Cavazos, H., Sjöström, A., & Sjöstrand, J. (2001). Screening merits of the Lang II, Frisby, Randot, Titmus, and TNO stereo tests. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 5(5), 316–322. <https://doi.org/10.1067/mpa.2001.118669>
- Otero-Millan, J., Macknik, S. L., Langston, R. E., & Martinez-Conde, S. (2013). An oculomotor continuum from exploration to fixation. *Proceedings of the National Academy of Sciences*, 110(15), 6175–6180. <https://doi.org/10.1073/pnas.1222715110>
- Pannasch, S., Helmer, J. R., Roth, K., Herbold, A.-K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2). <https://doi.org/10.16910/jemr.2.2.4>
- Pedziwiatr, M. A., von dem Hagen, E., & Teufel, C. (2023). Knowledge-driven perceptual organization reshapes information sampling via eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 49(3), 408–427. <https://doi.org/10.1037/xhp0001080>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>.
- Ranti, C., Jones, W., Klin, A., & Shultz, S. (2020). Blink rate patterns provide a reliable measure of individual engagement with scene content. *Scientific Reports*, 10(1), 8267. <https://doi.org/10.1038/s41598-020-64999-x>
- Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science*, 11(1), 50–74. <https://doi.org/10.1111/tops.12292>
- Roth, N., Rolfs, M., Hellwich, O., & Obermayer, K. (2023). *Objects guide human gaze behavior in dynamic real-world scenes* [preprint]. *Neuroscience*. <https://doi.org/10.1101/2023.03.14.532608>
- Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews*, 96, 353–366. <https://doi.org/10.1016/j.neubiorev.2018.12.007>
- Smith, M. E., & Loschky, L. C. (2019). The influence of sequential predictions on scene-gist recognition. *Journal of Vision*, 19(12), 14. <https://doi.org/10.1167/19.12.14>
- Smith, M. E., Loschky, L. C., & Bailey, H. R. (2021). Knowledge guides attention to goal-relevant information in older adults. *Cognitive Research: Principles and Implications*, 6(1), 56. <https://doi.org/10.1186/s41235-021-00321-1>
- Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, 6(1), 1–27. <https://doi.org/10.3167/proj.2012.060102>
- Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting saliency. *Journal of Vision*, 11(5), 5. <https://doi.org/10.1167/11.5.5>
- Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 366(1564), 596–610. <https://doi.org/10.1098/rstb.2010.0188>
- Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3), 473–494. <https://doi.org/10.1080/13506280444000409>
- Van der Stigchel, S., & Hollingworth, A. (2018). Visuospatial working memory as a fundamental component of the eye movement system. *Current Directions in Psychological Science*, 27(2), 136–143. <https://doi.org/10.1177/0963721417741710>
- Velichkovsky, B. M., Rother, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 145–156. [https://doi.org/10.1016/S1369-8478\(02\)00013-X](https://doi.org/10.1016/S1369-8478(02)00013-X)
- Walter, K., & Bex, P. (2021). Cognitive load influences oculomotor behavior in natural scenes. *Scientific Reports*, 11(1), 1–12. <https://doi.org/10.1038/s41598-021-91845-5>
- Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision*, 12(1), 16. <https://doi.org/10.1167/12.1.16>
- Wurtz, R. H., Joiner, W. M., & Berman, R. A. (2011). Neuronal mechanisms for visual stability: Progress and problems. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 366(1564), 492–503. <https://doi.org/10.1098/rstb.2010.0186>
- Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye movements actively reinstate spatiotemporal mnemonic content. *Vision*, 3(2). <https://doi.org/10.3390/vision3020021>