Temporal eye movement strategies during naturalistic viewing

Helena X. Wang
Department of Psychology and Center for Neural Science, New York University, USA

Jeremy Freeman
Department of Psychology and Center for Neural Science, New York University, USA

Elisha P. Merriam
Department of Psychology and Center for Neural Science, New York University, USA

Uri Hasson
Department of Psychology and the Neuroscience Institute, Princeton University, USA

David J. Heeger
Department of Psychology and Center for Neural Science, New York University, USA

The deployment of eye movements to complex spatiotemporal stimuli likely involves a variety of cognitive factors. However, eye movements to movies are surprisingly reliable both within and across observers. We exploited and manipulated that reliability to characterize observers' temporal viewing strategies while they viewed naturalistic movies. Introducing cuts and scrambling the temporal order of the resulting clips systematically changed eye movement reliability. We developed a computational model that exhibited this behavior and provided an excellent fit to the measured eye movement reliability. The model assumed that observers searched for, found, and tracked a point of interest and that this process reset when there was a cut. The model did not require that eye movements depend on temporal context in any other way, and it managed to describe eye movements consistently across different observers and two movie sequences. Thus, we found no evidence for the integration of information over long time scales (greater than a second). The results are consistent with the idea that observers employ a simple tracking strategy even while viewing complex, engaging naturalistic stimuli.

Keywords: eye movement, naturalistic viewing, point of interest

Citation: 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, 1–27, http://www.journalofvision.org/content/12/1/16, doi:10.1167/12.1.16.

Introduction

The human visual system relies on rapid eye movements to foveate regions of interest in a visual scene. Static images such as photographs and line drawings have long been used to infer a large number of stimulus- and task-dependent factors that drive eye movements (Buswell, 1935; Mannan, Ruddock, & Wooding, 1995; Noton & Stark, 1971; Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch, 2005; Reinagel & Zador, 1999; Tatler, Baddeley, & Gilchrist, 2005; Yarbus, 1967). The use of dynamic, naturalistic stimuli has extended that work to reveal how the time course of eye movements depends on the temporal evolution of visual events. The dominant computational framework for studying gaze allocation for both static and dynamic stimuli begins with the characterization of local image properties at fixated locations (Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Parkhurst et al., 2002; Parkhurst & Niebur, 2003; Peters et al., 2005; Rajashekar, van der Linde, Bovik, & Cormack, 2007; Reinagel & Zador, 1999; Tatler, Baddeley et al., 2005). Low-level visual features such as intensity, color, orientation, and motion contrast are computed at each location and combined to yield a master, scalar "saliency map" that predicts the conspicuity of a given location in a scene (e.g., Itti & Baldi, 2005; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985; Parkhurst et al., 2002; Peters et al., 2005). Observers are more likely to fixate locations of high salience. Such bottom-up models provide a biologically grounded and principled approach for relating gaze locations to stimulus features.

Many factors not predicted by bottom-up saliency also contribute to the spatiotemporal deployment of eye movements. Eye movements depend on the instructions and ongoing goals of a task (Ballard & Hayhoe, 2009; Buswell, 1935; Land, 2009; Land & Hayhoe, 2001; Rothkopf, Ballard, & Hayhoe, 2007; Yarbus, 1967), prior

doi: 10.1167/12.1.16 Received March 10, 2011; published January 19, 2012 ISSN 1534-7362 © ARVO

expectations and knowledge about semantic and spatial relationships among objects in a scene (Henderson, Weeks, & Hollingworth, 1999; Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006), and social cues such as faces and gaze directions (Birmingham, Bischof, & Kingstone, 2008; Friesen & Kingstone, 1998; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010). Many of these top-down factors contribute to idiosyncratic eye movement patterns in individual observers, reflecting differences in their task strategy and prior knowledge (Buswell, 1935; Noton & Stark, 1971; Yarbus, 1967). Alternatively, idiosyncrasies may reflect individual differences in oculomotor control and execution (Andrews & Coppola, 1999). As such, the allocation of eye movements in complex scenes likely reflects a collection of processes of varying time scales, from early sensory processing to recognition and memory.

We sought to examine the relationship between the temporal properties of a naturalistic scene and the temporal dynamics of eye movements. Rather than try to determine which features in such stimuli drive eye movements, we asked how eye movements depended on the integration of visual information (of any kind) across time. In spite of their complexity, some temporally dynamic stimuli (e.g., well-produced films) evoke similar eye movements across repeated viewings and across different observers (Carmi & Itti, 2006a; Goldstein, Woods, & Peli, 2007; Hasson, Landesman et al., 2008; Hasson, Malach, & Heeger, 2010; Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Shepherd et al., 2010; Tosi, Mecacci, & Pasquali, 1997). This repeated-viewing and inter-subject reliability represents a substantial level of control over the observer's viewing behavior and can be quantified without specifying or modeling the salient image features that attract eye movements.

The content of a movie spans multiple time scales that may influence reliable viewing behavior. There are moment-to-moment changes in the visual stimulus. However, there are also properties that span longer time scales. For example, understanding the narrative of a film requires integrating information over time. Some or all of such long time-scale features might or might not contribute to the reliability of eye movements. Like eye movements, brain activity during movie viewing is highly reliable within and across observers (Hasson et al., 2010; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004), but the activity in some brain areas is less reliable when the temporal sequence of the film is disrupted (Hasson, Yang et al., 2008), implying that the activity in those brain areas depends on the accumulation of information over long time scales.

Here, we used movie scenes that evoked highly reliable eye movements across observers to measure, manipulate, and model the reliable component of eye movements. Our goal was to determine if the reliability of eye movements is affected by disrupting the temporal sequence of a stimulus, and if so, whether eye movement reliability necessarily depends on information in the stimulus that is presented over long time scales. We manipulated the temporal continuity of a scene from a feature film by dividing the scene into clips of various durations and presenting them in scrambled order. We measured eye movements to the temporally scrambled version of the scene and compared them with eye movements to the same clips when they were presented in the original intact order. The original scene was shot as a single take without any cuts, and the scrambling manipulation introduced sharp discontinuities in the spatiotemporal structure of the stimulus. Scrambling systematically disrupted the reliability of eye movements, in a manner that depended on the temporal scale of scrambling.

We developed a simple computational model to account for these data. To capture the reliable component of eye movements (i.e., the variability in eye position over time that was shared across observers), the model assumed that the observer tracked a point of interest on the screen while viewing the intact scene. We approximated that point of interest as the median of the measured eye movement time courses across observers for the intact scene. The model made no assumptions about the processes underlying the high reliability, which could consist entirely of bottom-up features, entirely of top-down factors, or of a combination of the two. When an observer viewed the temporally scrambled version of the movie, the model assumed that the observer searched for, found, and tracked the same point of interest after each cut. As such, the model did not require any dependence of eye movement reliability on temporal context, and the model predicted that the dependence of eye movement reliability on temporal scrambling simply reflected time needed to find the point of interest following each cut. The model provided an excellent fit, with a small number of parameters, to eye movement measurements across multiple observers and for scenes from two very different movies. Therefore, we found no evidence that the integration of information over longer time scales (greater than about 1 s) influenced eye movements in any way that contributed to their reliability.

Materials and methods

Observers

Twelve observers, aged between 24 and 47, with normal or corrected-to-normal vision participated in the study. Observers provided written informed consent, and the experimental protocol was approved by the University Committee on Activities involving Human Subjects at New York University.

Stimuli and experimental procedure

Stimuli for the main experiment were derived from a 6-min scene from the motion picture *Children of Men* (Universal Pictures, 2006). The experiment was also conducted with a 3-min scene from the film *Russian Ark* (the State Hermitage Museum, 2002). Both scenes were shot as single takes without any cuts.

The scene from *Children of Men* was subdivided into short clips, each of equal duration. This process was repeated for five different durations (0.5 s, 1 s, 2 s, 5 s, and 30 s), which we refer to as "scramble durations." We pooled all of these clips together, randomly shuffled their order, and concatenated them, resulting in a 30-min movie composed of interleaved clips of varying lengths, with cuts at the transition between clips (Figure 1A). Randomly interleaving clips of different durations prevented anticipatory eye movements to predictable cut onsets, which might have occurred if observers viewed separate sequences containing clips of the same scramble duration. We refer to the scrambled movie as "interleaved" and the original 6-min movie as "intact." The same manipulation was applied to the *Russian Ark* scene to make an interleaved movie of 15 min.

Eleven observers participated in the *Children of Men* experiment. Three of these observers, and one additional observer, participated in the *Russian Ark* experiment. Some observers had seen *Children of Men* before the experiment, but there were no qualitative differences in the results between those observers and the observers who had not seen the movie. None of the observers had seen *Russian Ark* before the experiment. For each experiment, each observer viewed the intact movie twice and the interleaved movie once (*Children of Men* shown in two consecutive parts, ~15 min at a time; *Russian Ark* shown in whole). For all data reported for the main experiments, the observer viewed the intact movie first, then the interleaved movie, then the intact movie again. To verify that our conclusions did not rely on this ordering of conditions, we collected data from two additional observers who had not seen *Children of Men* before the experiment. These observers viewed the interleaved scene of *Children of Men* twice (on two separate days) before finally viewing the intact scene.

Gaze positions were measured (500 Hz, monocular) with an infrared (Eyelink 2000, SR Research) eye tracker. A 9-point calibration was performed at the start of each movie presentation. All movies were 24 frames/s and presented using the Psychtoolbox (Brainard, 1997; Pelli, 1997) in MATLAB (Mathworks) on a 22″ flat screen CRT monitor (Hewlett-Packard p1230; resolution of 1152 × 870) at a distance of 57 cm. The monitor provided approximately 39° × 30° of viewing angle. The *Children of Men* stimuli were shown at 1037 × 560 resolution (35.5° × 19.5° of viewing angle) and the *Russian Ark* stimuli were shown at 1037 × 585 resolution (35.5° × 20.4° of viewing angle). All stimuli were shown without sound, so as to avoid potential artifacts from temporally scrambling the soundtrack and to specifically identify eye movements induced by a visual stimulus (rather than a

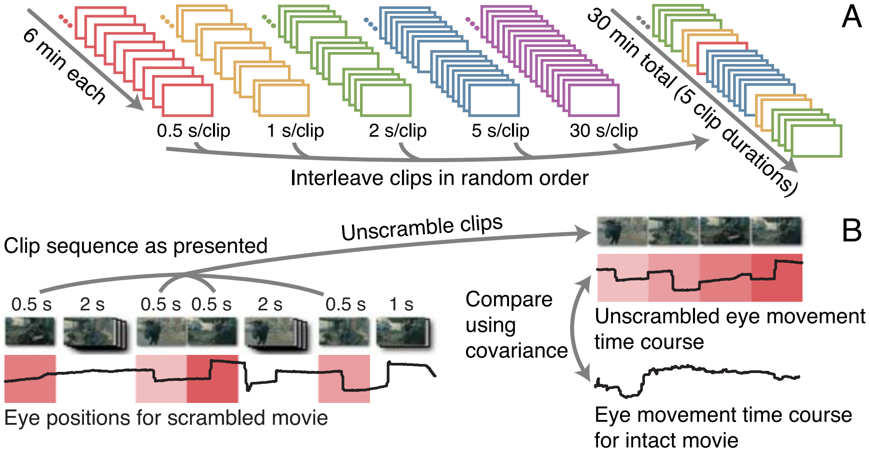
A
6 min each
0.5 s/clip
1 s/clip
2 s/clip
5 s/clip
30 s/clip
30 min total (5 clip durations)
Interleave clips in random order
B
Clip sequence as presented
Unscramble clips
0.5 s
2 s
0.5 s
0.5 s
2 s
0.5 s
1 s
Eye positions for scrambled movie
Compare using covariance
Unscrambled eye movement time course
Eye movement time course for intact movie

Figure 1. Scrambling manipulation and unscrambling analysis. (A) Construction of interleaved movie stimulus. A continuous 6-min scene from the film *Children of Men* was divided evenly into short clips at each of five durations: 0.5 s, 1 s, 2 s, 5 s, and 30 s. In the cartoon, each rectangular box depicts a movie sequence of 0.5 s long, so that a group of two boxes represents a 1-s clip, a group of four boxes represents a 2-s clip, and so on. Clips of all durations were interleaved in random order to create a 30-min movie. (B) Unscrambling analysis. For each clip duration (shown here for 0.5 s), eye movement time courses (horizontal and vertical) were extracted and rearranged to match the order of the corresponding clips in the intact movie. Covariance was computed between the unscrambled eye movement time courses and those for the intact movie.

combined audiovisual stimulus). Both scenes evoked highly reliable eye movements despite the lack of sound.

Data preprocessing

Eye positions were recorded in screen coordinates and normalized by the resolution of the movie, such that both horizontal and vertical eye positions varied between values of 0 and 1 irrespective of screen dimension or stimulus. A value of 0 corresponded to the leftmost (horizontally) or uppermost (vertically) edge of the movie, and a value of 1 corresponded to the rightmost (horizontally) or the bottommost (vertically) edge. Data points were discarded if eye positions were offscreen, or if there was signal loss (e.g., if the eye tracker reported failing to locate the pupil center because of eye blink, eyelash occlusion, or other recording artifacts). Spline interpolation was used to fill in these discarded time points, which accounted for 8.9% ± 3.5% of time points (mean ± standard deviation across $n = 15$ observers, combining data from both movies). One observer for the *Children of Men* experiment was excluded from further analysis because the variance of his eye positions (both horizontal and vertical) for the interleaved movie and for one of the intact movie measurements were two standard deviations below that of the rest of the observers. Thus, all subsequent analyses for the main experiments were based on data from 10 observers for the *Children of Men* experiment and 4 observers for the *Russian Ark* experiment.

Saccades were detected and parsed using the Eyelink (SR Research) saccade detection algorithm. The following detection thresholds were used: eye movement amplitude > 0.1°, velocity > 30°/s, and acceleration > 8000°/s^2. The configuration was relatively conservative (hence, insensitive to noise) and ignored most microsaccades. On average, 1.7 ± 0.3 saccades/s (mean ± standard deviation across $n = 10$ observers) were detected for the *Children of Men* experiment, and 2.1 ± 0.3 saccades/s were detected for the *Russian Ark* experiment ($n = 4$ observers).

Covariance analysis

Reliability of eye movements was quantified in two ways. First, we measured the covariance between eye positions for the intact movie and eye positions for the same content when presented within the interleaved movie (as explained in the following paragraphs). Second, we measured the squared difference between eye positions for the intact movie and eye positions for the same content when presented within the interleaved movie and used that to estimate how well the intact eye positions predicted the unscrambled interleaved eye positions as a function of time (as explained below under Eye position error, variance in eye position, and fractional explained variance section).

Reliability of eye movements was quantified with covariance and cross-covariance. For each observer and scramble duration, eye movement time courses for the interleaved movie were reassembled to match the temporal sequence of the intact movie. As an example, for a scramble duration of 0.5 s, excerpts from the eye movement recordings (both horizontal and vertical) corresponding to all 0.5-s clips in the interleaved movie were rearranged and assembled to match the temporal order of the same clips in the intact movie (Figure 1B). We refer to this as the “unscrambled eye movement time course.” The unscrambled eye movement time course and the eye movement time course for the intact movie contained eye positions in response to the same visual content. However, in one case (intact), the clips had been presented in their original order, and in the other case (interleaved), the clips had been presented in a random sequence. The same procedure was performed for each of the other scramble durations.

For each observer and each scramble duration, cross-covariance functions (Figure 2) were computed between the unscrambled eye movement time courses and the eye movements for the intact movie (separately for horizontal and vertical). Cross-covariance is the sliding inner product of two mean-subtracted signals. It is expressed as a function of the time lag between the two time courses. For two discrete signals g and h, the sample cross-covariance is defined as

$$\hat{C}_{g,h}[k] = \frac{1}{N}\sum_{\delta}(g[\delta] - \mu_g)(h[\delta + k] - \mu_h), \qquad (1)$$

where k is the time lag between the two signals, N is the number of samples, and μ_g and μ_h are the sample means of the two signals. Both g and h were zero-padded so that the sum was always over N samples. For some analyses, we used each observer's own eye movements for the intact movie, but in other analyses, we used the median (across observers) eye movement time course for the intact movie. The median eye movement time course was computed by aligning all eye movement time courses (2 repeats of each intact movie per observer; 20 in total for *Children of Men* stimuli, and 8 for *Russian Ark*) to the same set of sampling time points and taking the median at each time point. The covariance of two time courses is the value of the cross-covariance function at a time lag of $k = 0$. Covariance is often normalized by the product of the standard deviation of the time courses, yielding the familiar Pearson's correlation coefficient. We observed that eye position variances were not constant across scramble durations (see Eye movement reliability decreased with shorter scramble durations section; Figures 3E and 3F). Trying to account for how variances depend on scramble duration would have made the model intractable. Our principal analysis, therefore, was to compute unnormalized covariance.

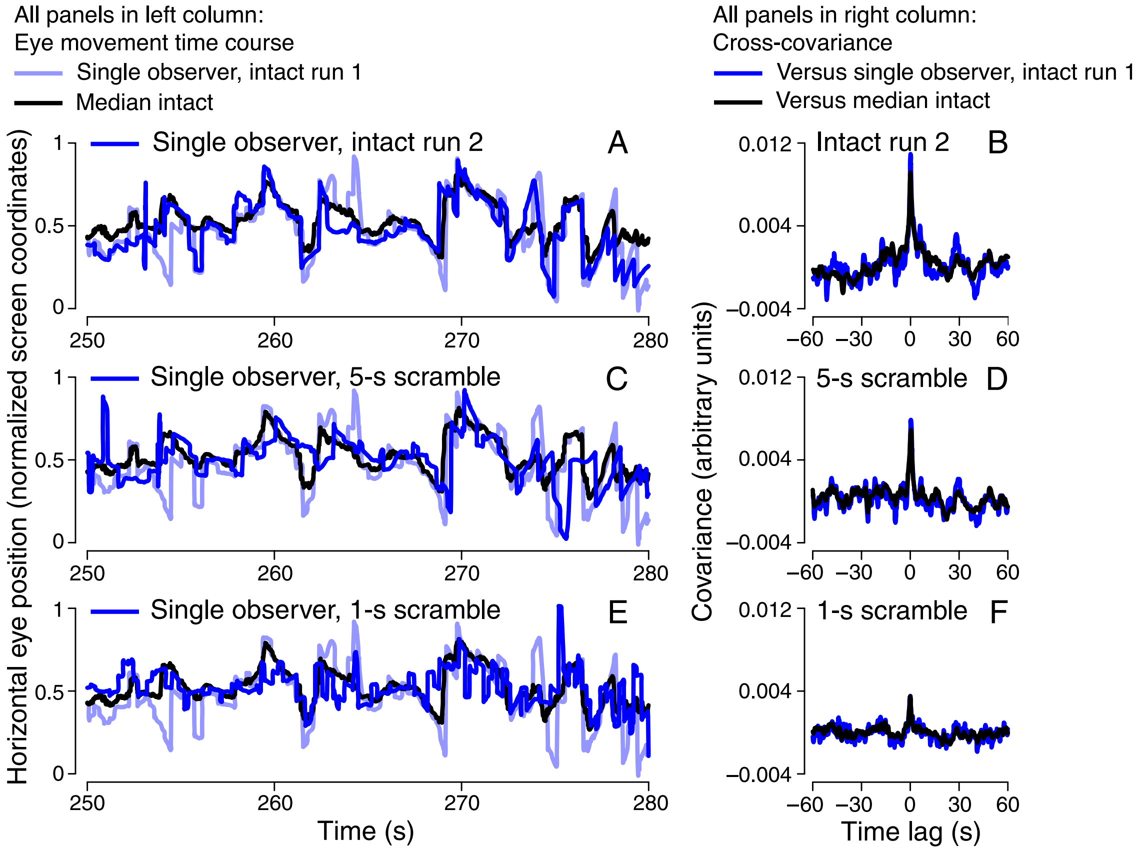

Figure 2. Examples of eye movement cross-covariance for different scramble durations. (A) Eye movement time courses for the intact movie. Dark and light blue, sample eye movements from a single observer for two separate presentations of the intact movie. Black, median across the other observers for the intact movie (n = 9). Eye positions are normalized by the extent of the video in each dimension, so 0 corresponds to the leftmost edge of the video and 1 corresponds to the rightmost. Only horizontal eye positions are shown (in this and the other panels), but the results for vertical eye positions were similar. (B) Cross-covariance of eye movements for the intact movie. Blue, cross-covariance between eye movement time courses for two presentations of the intact movie from a single observer. Black, cross-covariance between eye movements from a single observer for a single presentation of the intact movie and the median across the other observers for the intact movie (n = 9). The peak at a time lag of 0 s shows that eye movement time courses were highly correlated and time locked to the stimulus. (C) Eye movements for the 5-s scramble duration. Dark blue, unscrambled eye movements from a single observer for the 5-s scramble duration (see Figure 1B). Light blue, eye movements from the same observer for a single presentation of the intact movie. Black, median across all observers for the intact movie (n = 10). Light blue is replotted from A. (D) Cross-covariance for the 5-s scramble duration. Blue, cross-covariance between unscrambled eye movement time courses from a single observer for the 5-s scramble duration and eye movements from the same observer for a single presentation of the intact movie. Black, cross-covariance between unscrambled eye movements from the same observer for the 5-s scramble duration and median across all observers for the intact movie (n = 10). (E, F) Same as (C) and (D) for the 1-s scramble duration. Covariances are lower for the shorter scramble duration.

Except where noted, covariances were computed between individual observers' unscrambled eye movements for the interleaved movie and the median eye movement time course (across all observers) for the intact movie. In some analyses, we also compared the eye movement time course for the intact movie from an individual observer with the median from the other observers (Figures 3A–3D, dashed lines). In that case, the median excluded the data of the individual observer to avoid any statistical bias.

A phase randomization test was used to assess whether the covariance between two eye movement time courses was statistically significant. Specifically, we took the discrete Fourier transform of one of the time courses, randomly permuted its Fourier phase without changing the amplitude, inverted the Fourier transform, and recomputed the covariance between the resulting time course and the other time course. This procedure was repeated 1000 times to yield a null distribution for the covariance between the two time courses. We determined a p value as the fraction of the null distribution that was as large or larger than the covariance observed without randomization.

Model

We developed a simple model to account for the reliability of eye movements to complex movie stimuli and tested the predictions of the model. We begin with the key assumptions and intuitions behind the model. See Appendix A for a detailed derivation.

The model assumes that for any particular movie stimulus, there is a hypothetical "point of interest" that follows a particular trajectory over time. This point of interest can be thought of as a target or the correct place to look on the screen at any point in time. The observer is assumed to behave as follows. Starting from the beginning of any stimulus presentation (the start of the movie or immediately after a cut), each saccade made by the observer had a fixed probability of finding the point of interest. That probability depended on several unknown factors, including both the stimulus and the observer. Before finding the target, eye movements were assumed to be uncorrelated with the location of the point of interest. After finding the target, the observer locked on and tracked the point of interest. So long as the observer had locked on to the point of interest, covariance between the observer's eye movement time course and the point of interest trajectory was maximal (limited only by measurement noise and by the observer's internal cognitive and motor variability). The model made no assumptions about the statistical nature of eye movements before the observer locked on but only required that they were uncorrelated with the point of interest during that period. In fact, our data show that the variance of eye movements evolved systematically as a function of time after a cut, including a tendency for observers to fixate near the center of the screen following a cut (Figure 6A). A variation of the model assumes that while the point of interest was being tracked, there was a certain probability at any time that the observer abandoned the point of interest and made exploratory eye movements to look for another point of interest. This process of exploration could be exactly the same as that which happens immediately following a cut. Adding this exploratory process to the model affected only the maximal covariance and is accounted for in the derivation (see Appendix A) as one possible source of noise.

How is this model affected by our scrambling manipulation? Scrambling the temporal order of the movie introduced artificial cuts that are not present in the intact movie. We assumed that the tracking process was reset after each cut. When clip durations were short (and the number of such clips is large), the observer reset (and needed to rediscover the point of interest) more frequently. A large proportion of the eye positions, over the course of the entire stimulus, were uncorrelated with the point of interest, simply because the observer needed time to find the point of interest following each cut and, consequently, spent more time not looking at the right place. Therefore, the eye movement reliability was lower (low covariance) for shorter scramble durations.

We derived a closed-form expression for the model (see Appendix A). For each scramble duration, covariance with the point of interest depended solely on the proportion of time during which the observer was locked on or not locked on to the point of interest. In the derivation, the model assumed that the observer made a series of independent saccades after each cut and that there was a fixed probability λ (Figure 5C) that the observer would find and lock on to the point of interest after each saccade. Assuming that the saccades were statistically independent after each cut made the model analytically tractable, but violations of this independence assumption would not have qualitatively changed the predictions of the model (see Integration of visual information across fixations during search section). We also defined Q_{H} and Q_{V} to be the "maximal" covariances attainable (in horizontal and vertical eye positions, respectively) between an intact eye movement time course and the trajectory of the point of interest. For a particular scramble duration, we expressed the predicted covariance between an unscrambled eye movement time course, S^d, and the point of interest time course, S, as a function of λ, Q_{H}, and Q_{V} (see Equation A12 in Appendix A).

Model fitting

We used the median eye movement time course (across all observers, $n = 10$) for the intact movie as an estimate for the point of interest, which served as a prediction for the unscrambled eye movement time courses for the interleaved movie. Covariances were computed between individual observers' unscrambled eye movements, S^d, and the median time course, S (see Covariance analysis section above). We fit the model to the data by finding parameters that minimized the squared error between the predicted covariance (from Equation A12 in Appendix A) and measured covariance. First, we estimated the parameters for the inter-saccade interval distribution of each observer. Specifically, parameters μ and σ (Equation A1 in Appendix A) were determined by fitting a lognormal distribution (using maximum likelihood, *lognfit* function in MATLAB) to each observer's inter-saccade intervals for the intact movie. Fitted values of μ and σ did not vary substantially across observers. Second, with μ and σ fixed for each observer, we then estimated the parameters that best accounted for the covariance values for that observer. The covariance was computed between that observer's unscrambled eye movement time course, S^d, for each scramble duration d, and the point of interest time course, S, separately for horizontal and vertical eye positions. We used the median eye movements for the intact movie as an estimate of S because it is robust to outliers; using the mean eye movement time course produced similar results. The fit was performed simultaneously for all scramble durations and simultaneously for horizontal and vertical eye positions. We accounted for individual variation in maximal covariance in both horizontal and vertical eye positions with the free parameters Q_H and Q_V. A constrained non-linear optimization routine (*fmincon* function) in MATLAB was used to numerically solve for the values of the three free parameters (λ, Q_H, and Q_V) that minimized the squared error between the predicted and measured covariances (10 data points). In the fit, λ was constrained to be between 0 and 1 and Q_H and Q_V to be greater than 0. Hence, there were a total of 5 free parameters: μ and σ were fit to the inter-saccade interval distributions, and λ, Q_H, and Q_V were fit to the measured covariances.

Bootstrapping was used to obtain confidence intervals for the parameter λ (Figure 5C). For each observer, eye movement epochs of 30 s were randomly sampled with replacement from the eye movement time course for the intact movie and concatenated to obtain an eye movement time course with length equivalent to the length of the original scene (6 min for *Children of Men* and 3 min for *Russian Ark*). Corresponding epochs were extracted from the five unscrambled eye movement time courses for the interleaved movie, such that for each 30-s epoch, eye positions for the 30-s scramble duration were derived from a single clip, and eye positions for the remaining four scramble durations were derived from clips that had been unscrambled to match the content of that 30-s clip. After each resampling, covariances were recomputed and the fit was performed to reestimate λ. This procedure was repeated 1000 times, and the 2.5th and 97.5th percentiles of the resulting distribution of λ values provided a 95% confidence interval (equivalent to two standard deviations if the distributions were normally distributed).

Goodness of fit was assessed with cross-validation. Half of all 30-s clips from the intact movie (and corresponding clips from the interleaved movie) were used to compute covariances and estimate model parameters (training). We then used all the fitted parameters (λ, Q_H, and Q_V) to predict covariances on the remaining half of data and the fitted parameter λ was used to compare model predictions with actual covariances for the other half of the data (testing). The cross-validation was unstable in individual observers due to the occasional occurrence (for some training and testing splits) of large differences in asymptotic covariances Q_H and Q_V between the training and testing data. We therefore performed this analysis only after concatenating data across all observers, which stabilized estimates of maximal reliability. This procedure was performed 1000 times to obtain a 95% confidence interval on the goodness-of-fit measure r^2 (coefficient of determination or percentage variance explained by the fit) for the combined data.

Eye position error, variance in eye position, and fractional explained variance

Another implication of the model is that the point of interest should serve as a poor predictor of a measured eye movement time course immediately following a cut but become better shortly after when eye movements converge on to the point of interest. To test this prediction, we examined how the squared difference between the measured eye movement time courses and the point of interest (given by the median eye movement time course across observers) evolved as a function of time after a cut. This difference (the measured "eye position error") should start high and drop to a baseline (Figure 6B). At any particular time point, a large eye position error might have suggested that the observer was unlikely to have locked on to the point of interest, and a smaller eye position error might have suggested that the observer was more likely to have locked on. The magnitude of eye position error, therefore, might have been proportional to the fraction of time (across all clips) that the observer was not locked on to the point of interest. The eye position error was, however, confounded by changes in the eye position variance, which evolved systematically as a function of time after a cut (Figure 6A).

To isolate the component of the eye position error that reflected only the probability of locking on to the point of interest (or the fraction of time that an observer was locked on), we computed what we call the "fractional explained variance." This quantity estimated the fraction of eye position error explained by the point of interest relative to that expected under the assumption of no correlation between the point of interest and the unscrambled eye movements. We computed the fractional explained variance in eye position in the following manner:

(1) The squared error in eye position, $G(t) = \mathrm{E}[(S^d(t) - S(t))^2]$, was computed for each observer (Figure 6B), where $S^d(t)$ was the unscrambled eye movement time course for clip duration d from the interleaved movie, $S(t)$ was the median eye movement time course for the intact movie, and t ranged from 0 to d for each S^d of a particular duration d (i.e., from the beginning to the end of each clip). $G(t)$ was computed by averaging across all clips from all scramble durations for that observer, aligned to each cut. $G(t)$ computed separately for each scramble duration d yielded similar curves, therefore justifying averaging across durations, resulting in more averaging for smaller values of t.

(2) The variance of the unscrambled eye movement time courses, $v_{Sd}(t)$, was estimated as a function of time after a cut (Figure 6A). The sample mean eye position time course, $E[S^d(t)]$, averaged across clips, was ~0.5 for both horizontal and vertical dimensions (center of the screen) at any time t. The variance $v_{Sd}(t)$, therefore, reflected the fact that eye positions tended to cluster near the center of the screen shortly after a cut and then gradually expand outward over time (Figure 6A). $v_{Sd}(t)$ was computed separately for each observer across all clips from all scramble durations for that observer; all observers showed the same tendency.

(3) The maximal position error, $G_0(t)$, was computed as the sum of the variance (across clips) of the unscrambled eye movements, $v_{Sd}(t)$, and the variance (over time) of the median eye movement time course (see Fractional explained variance: Derivations section in Appendix A). Intuitively, $G_0(t)$ reflected how eye position error would have evolved over time after a cut if the unscrambled eye movements never locked on to the point of interest. $G_0(t)$ was not constant over time, as would be expected if the variance of $S^d(t)$ was stationary, confirming that it would have been inappropriate to use $G(t)$ by itself to infer the temporal dependence of $S^d(t)$ on the point of interest.

(4) For each observer, we then computed fractional explained variance as $1 - G(t)/G_0(t)$ (Figure 6C), which could be interpreted as an estimate for the probability (across clips) that the unscrambled eye position was locked on to the point of interest (the median eye position) as a function of time after a cut (see Appendix A for derivation).

We also simulated fractional explained variance using the model described above (Figure 6C, inset; see Appendix A for details).

Results

Eye movements to intact movie were reliable both within and across observers

Replicating previous results (Goldstein et al., 2007; Hasson, Landesman et al., 2008; Hasson, Yang et al., 2008; Shepherd et al., 2010; Tosi et al., 1997), we found that movies evoked reliable eye movements. We tracked eye position in 10 observers while they watched a 6-min scene from the feature film *Children of Men*. Each observer viewed the scene twice. The movie stimulus evoked reliable eye movements across repeated presentations within an individual observer and across observers (Figure 2A). We quantified the degree of reliability using cross-covariance (see Covariance analysis section), separately for horizontal and vertical eye movements. For each observer, cross-covariance was computed between eye movement time courses for two presentations of the intact movie (for that observer) and between eye movements for one presentation of the intact movie (for that observer) and the median eye movement time course across the other 9 observers (Figure 2B). In both cases, cross-covariance was maximal at a time lag of zero, suggesting that correlated changes in eye position were time locked to stimulus events. The width of the peak indicated the temporal precision of the time locking. The magnitude of the peak at a time lag of zero (i.e., the covariance) provided a measure of the reliability of eye movements for that observer, given instrument noise and the observer's internal cognitive and motor variability across repeated measurements. The cross-covariance for time lags far from zero provided a qualitative baseline for spurious covariance due to chance. In general, covariance was high (well above the baseline for all observers, and highly statistically significant: $p < 0.001$ for all observers, phase randomization test; see Covariance analysis section).

Eye movement reliability decreased with shorter scramble durations

We parametrically disrupted the temporal continuity of the movie by scrambling the original scene at different time scales (0.5 s, 1 s, 2 s, 5 s, and 30 s). The original scene was divided into clips with each of these "scramble durations," and the clips were randomly ordered and reassembled into one long interleaved movie (see Stimuli and experimental procedure section; Figure 1). Eye positions were recorded while observers viewed this interleaved movie. For each observer, an eye movement time course corresponding to each scramble duration was extracted from the measurements for the interleaved movie, unscrambled (i.e., reordered to match the order of the intact movie), and compared with the eye movements for the intact movie. If temporal scrambling affected the reliability of eye movements, then the covariances should have been smaller.

Eye movements were less reliable for shorter scramble durations (Figures 2, 3A, and 3B). Covariances between unscrambled eye movements and eye movements for the intact movie (either the observer's own or the median across observers) were smaller for shorter scramble durations, as indicated by the lower peaks in the cross-covariance (Figures 2B, 2D, and 2F). Covariance was statistically above baseline even for the shortest scramble duration ($p < 0.025$ for the 0.5-s scramble duration for all observers in horizontal eye position and for 8 out of 10 observers in vertical eye position; $p < 0.025$ for all other scramble durations for all observers in both horizontal and vertical; phase randomization test, see Covariance analysis section). Covariance increased monotonically with scramble duration, for each of the 10 observers (Figures 3A and 3B). Covariances were computed by comparing a single observer's unscrambled eye movements with the median

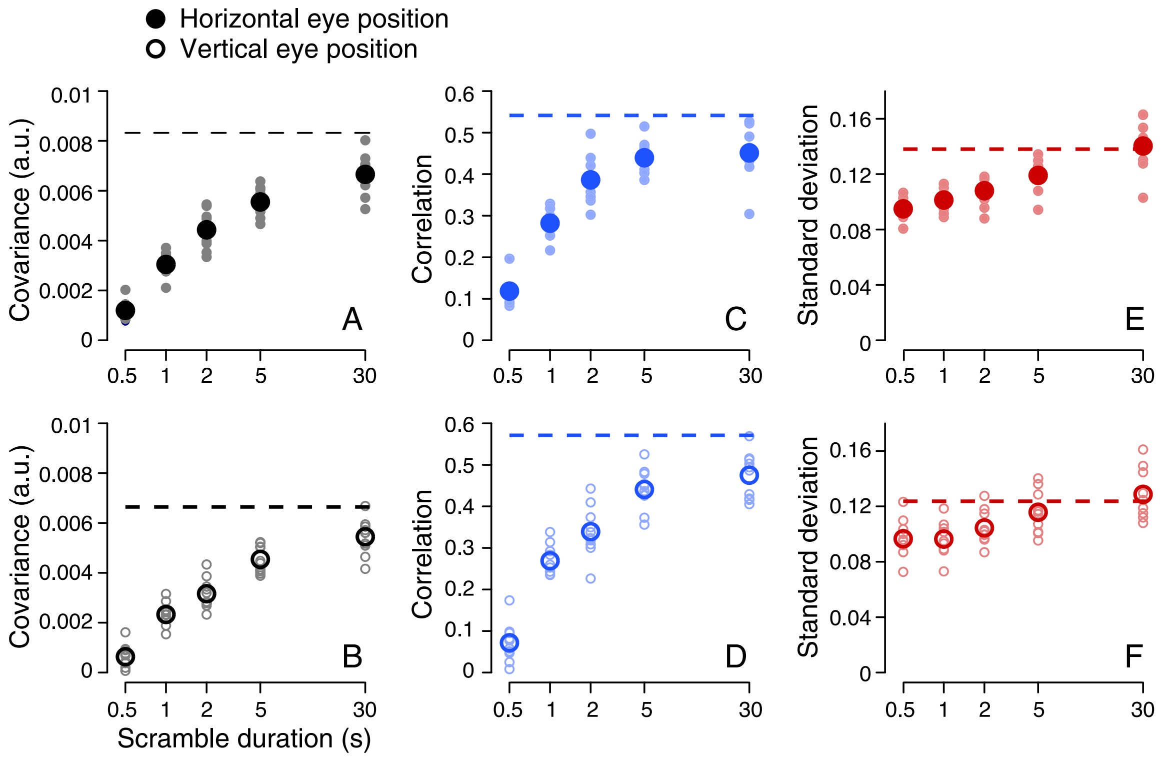

Figure 3. Eye movement reliability increases with scramble duration. Top row: Horizontal eye movements. Bottom row: Vertical eye movements. (A, B) Covariance as a function of scramble duration. Small symbols, covariances between the unscrambled eye movement time courses from a single observer for each scramble duration and median eye movements across observers for the intact movie. Large symbols, average covariances across observers (*n* = 10). Dashed lines, average covariances between the eye movements for a single presentation of the intact movie from each observer and the median eye movement time course across the other observers. Covariance was computed separately for each observer and averaged between the two intact movie presentations for each observer and across observers (*n* = 10); a.u., arbitrary units. (C, D) Correlation as a function of scramble duration. Same format as A and B. Correlations increased with scramble duration similarly to covariances (panels A and B). However, the correlation between two time courses equals their covariance divided by the product of the standard deviations, so the correlations depended both on the covariances and the standard deviations (panels E and F). (E, F) Standard deviation of eye movement time courses for each scramble duration. Small symbols, standard deviations of the unscrambled eye movement time courses for a single observer. Large symbols, average standard deviations across observers (*n* = 10). Dashed lines indicate the standard deviations for the intact time courses. At shorter scramble durations, eye positions tended to be clustered and did not span the full range of screen coordinates, yielding smaller standard deviations.

intact eye movements across observers. Covariances computed by comparing eye movements within an observer were similar. The covariance between the eye movements for two presentations of the intact movie indicated maximal reliability attainable for an observer, in the absence of scrambling. This covariance was computed, for each observer, between each intact eye movement time course from the individual observer (two per observer) and the median time course across the other observers. Covariances were then averaged between the two estimates per observer and across all observers (Figures 3A and 3B, dashed lines). The fact that all observers were similarly affected by the scrambling manipulation suggests a behavioral strategy or computation that was common across observers.

We used covariance rather than the more familiar correlation to quantify reliability, because correlation is covariance normalized by variance and conflates changes in covariance and variance. The correlation coefficient could have increased either because covariance increased or because variance decreased. The variance in eye positions showed a dependence on scramble duration, increasing monotonically with longer scramble durations (Figures 3E and 3F). We attribute this to the fact that variance decreased sharply right after each cut and then increased gradually over seconds; thus, overall variance was smaller for shorter cuts (see also Figure 6A and Discussion section). A potential problem with reporting covariance is that its magnitude is not intuitively interpretable (in the way that the correlation coefficient is). However, we plotted covariance for each scramble duration alongside the covariance for the intact movie. This “maximal” covariance serves as a reference point. While covariance is reported in our primary analyses, correlations were computed in a complementary analysis (Figures 3C and 3D); the pattern of results was qualitatively similar, with a maximal correlation coefficient of about 0.5–0.6, but the correlations would have been more difficult to model and interpret because the variances depended on scramble duration (Figures 3E and 3F).

Eye movement reliability did not depend on repeated viewing or presentation order

Observers might have employed different strategies depending on whether they had seen the scene more than once, resulting in systematically different eye movement time courses for the repeated presentations of the movie clips. To assess this possibility, we computed covariances separately for the first and second viewings of the intact movie, by comparing eye movements from the initial presentation from one observer to the median eye movement time course across the initial presentations of the other 9 observers and doing the same for the second presentation of the movie for each observer. In addition, we also computed the covariances between the individual eye movements from each observer's first presentation and the median eye movements from the other 9 observers' second presentations, and vice versa. This yielded four sets of covariances for assessing inter-subject eye movement reliability for the two presentations of the intact movie. We found no evidence that the covariances differed between any pair of these four sets ($p > 0.1$ for all 6 comparisons; randomization test, whether or not corrected for multiple comparisons). This suggests that eye movement reliability did not depend significantly on repeated viewings of the same scene.

In addition, to verify that the covariance values for the unscrambled time courses did not rely on the ordering of conditions, we collected data from two additional observers who viewed the interleaved movie first (see Stimuli and experimental procedure section). For each of these observers, we computed covariances by comparing the unscrambled eye movements to the median eye movements for the intact scene across the previous 10 observers. We performed this procedure separately for unscrambled eye movements corresponding to each viewing of the interleaved scene (two presentations per observer). We found no evidence for a difference in covariance values for these observers compared to those obtained for the original observers, who viewed the scenes in a different order ($p > 0.05$ for horizontal and vertical covariance values in all scramble durations; randomization test, corrected for multiple comparisons). Furthermore, for both of the additional observers, covariance values were qualitatively similar across the two repeated presentations of the interleaved scene, validating our earlier observation that reliability measurements did not depend substantially on the order of presentation or the experience of prior presentations.

A simple model accounted for the increase in eye movement reliability with scramble duration

The cinematically composed movie scene evoked reliable eye movements within and across viewers. Temporal scrambling systematically disrupted eye movement reliability. This might seem to imply that eye movements depended on temporal context. For example, perhaps observers accumulated information about the content of a clip over several seconds to make a decision about where to look next. However, is this kind of temporal context (and its disruption) necessary to explain the effect of scrambling on covariance?

We considered an alternative, simpler model in which observers tracked a point of interest on the screen, and eye movements depended on temporal context only insofar as the tracking process began anew at the beginning of each clip immediately following each cut. The point of interest provided a simple descriptive model to capture the reliable component of eye movements (i.e., the variability in eye position over time that was shared across observers). The model made no assumptions about the factors underlying the point of interest (i.e., bottom-up or top-down). It only required that the point of interest in a given stimulus frame was the same regardless of the temporal context in which the stimulus was presented (i.e., same when it was presented within an intact movie or within the different scramble durations of the interleaved movie). According to this model, eye movements for the intact movie were reliable because observers tended to track the same point of interest. Furthermore, according to the model, eye movements were uncorrelated immediately following a cut because it took time for observers to find a point of interest. With more cut transitions (and shorter clip durations), the search for a point of interest reoccurred with greater frequency. Consequently, eye movements for shorter scramble durations were less reliable, according to this model, simply because observers spent a greater percentage of time searching for a point of interest. Is this simple tracking model sufficient to explain the data?

We derived an implementation of the model and fit it to the measurements. The model depended on the distribution of saccade latencies (i.e., the inter-saccade interval distribution). The intervals at which an observer made saccades during a movie were well characterized by a lognormal distribution (Figure 4A). Parameters for that distribution were estimated from the data and were assumed to be invariant throughout the experiment for each observer. The model assumed that the observer made a series of independent saccades following each cut and that there was a fixed probability λ that the fixation following each saccade would lock on to the point of interest. Thus, the probability that the observer locks on at a given time after a cut is a weighted sum: The first term is the probability that the first saccade occurs at that time

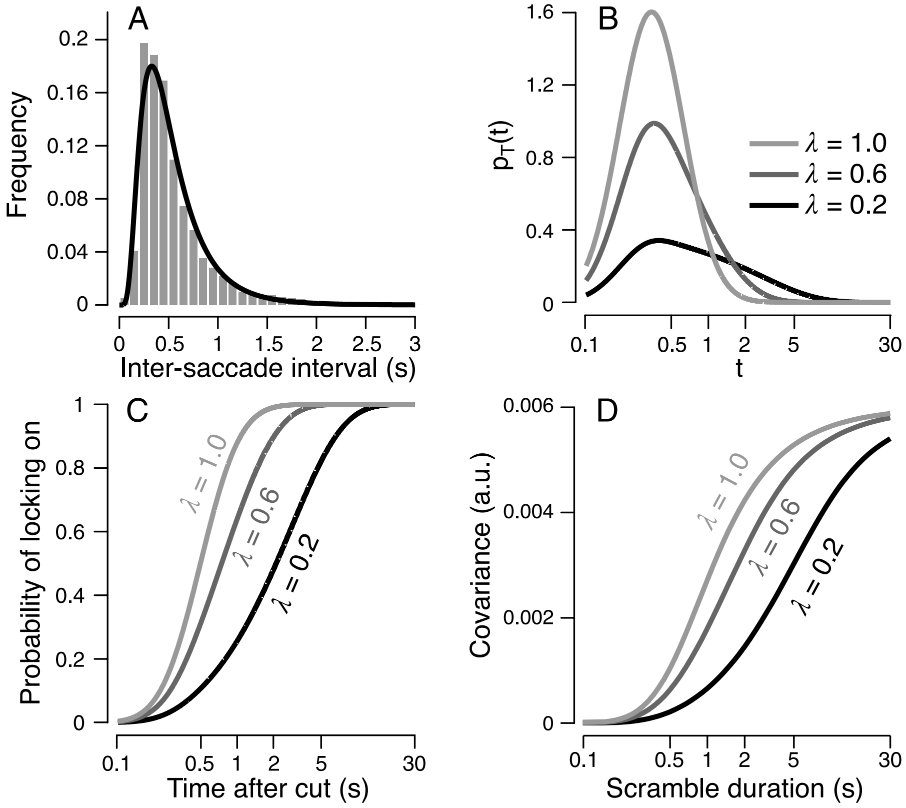

Figure 4. Model. (A) Inter-saccade intervals are well described by a lognormal distribution. Gray, histogram of inter-saccade intervals from a single observer for the two presentations of the intact movie. Black, best-fitting lognormal distribution. (B) Probability density function $p_T(t)$ for a continuous random variable T that describes the amount of time it takes to find a hypothetical "point of interest" after a cut (see Model section and Appendix A). The parameter λ determines the probability of finding the point of interest following a saccade. When λ = 1, the probability of fixating the point of interest after the first saccade is 1, so $p_T(t)$ is just the lognormal distribution (panel A). When $\lambda < 1$, the probability of finding the point of interest after each saccade is lower so the shape of $p_T(t)$ changes to have larger probabilities associated with later saccades after a cut. (C) Cumulative probability distribution $p_T(t)$ that describes the probability of having fixated the point of interest at a particular time after a cut. Over time following a cut, the probability increases to 1, but it does so more quickly (with a steeper slope) for larger values of λ. (D) Covariance as a function of scramble duration as predicted by the model, for different values of λ; a.u., arbitrary units.

and finds the point of interest, the second is the probability that the second saccade occurs at that time and finds the point of interest (given that the previous saccade did not), and so on. The mean of this probability distribution corresponds to the average time that it takes for an observer to find the point of interest following a cut. For small values of λ, it becomes increasingly likely that the point of interest will be found only after a long period of time (Figure 4B). The probability of having locked on to the point of interest within any particular time after a cut likewise depends on λ (Figure 4C). The function rises more slowly for a smaller λ, because it takes more time to accumulate probability of having locked on.

According to this model, eye movement reliability (covariance) depends systematically on λ, the probability of finding the point of interest at each fixation following a saccade (Figure 4D). We assumed that, while the observer locked on to some true point of interest, covariance with that point of interest was maximal. However, until he or she locked on, covariance was 0. Under this assumption, covariance over the course of the movie was proportional to the relative amount of time during which the observer was locked on (see Equation A9 in Appendix A). For example, when scramble durations were long, an observer spent most of the time locked on, and covariance was nearly maximal. However, when scramble durations were short (and there were many cuts), the observer spent less time locked on and more time searching for points of interest, so covariance was smaller. By such reasoning, we derived a closed-form expression for the covariance expected at different scramble durations (see Equation A12 in Appendix A). The relationship depends only on the frequency of saccades, the maximal obtainable covariance, and the free parameter λ, which describes the probability that an observer found the point of interest at each fixation following a saccade. Values for these parameters were found by numerically minimizing the squared error between the observed covariances and the predicted covariances. Parameterizing saccade times using a log-normal distribution yielded a closed-form solution, but the qualitative predictions of the model did not depend on the specific form of the saccade time distribution.

We fit the model to the data by finding values of λ and maximal covariances (horizontal and vertical) that best predicted the measured covariances across all scramble durations. The median (across observers) eye movement time course for the intact movie provided an estimate for the point of interest trajectory, and covariance was computed, for each observer, between each of the unscrambled eye movement time courses and this point of interest. The model fit the data well; when fit to the covariances combined across observers (see Modeling fitting section), r^2 was 0.88 (cross-validated 2.5th–97.5th percentiles = 0.61–0.98). The fitted value of λ was 0.79, corresponding to an expected time of 0.73 s (bootstrapped 2.5th–97.5th percentiles = 0.63–0.83 s) within which observers were able to find and lock on to the point of interest. The model was also separately fit to the data for each individual observer, and again accounted for most of the variance in the data from each observer (Figures 5A and 5B). Fitted values of λ for individual observers were between 0.5 and 1 (mean λ = 0.82 across 10 observers, Figure 5C), corresponding to an expected time of 0.75 ± 0.16 s (mean ± standard deviation, n = 10) for locking on to the point of interest. Although there may have been systematic individual differences in λ, our data did not have sufficient sensitivity or statistical power to explore it; values of λ varied somewhat across observers, but the confidence intervals for the most part overlapped. We fit the model to data from the two additional observers who viewed the

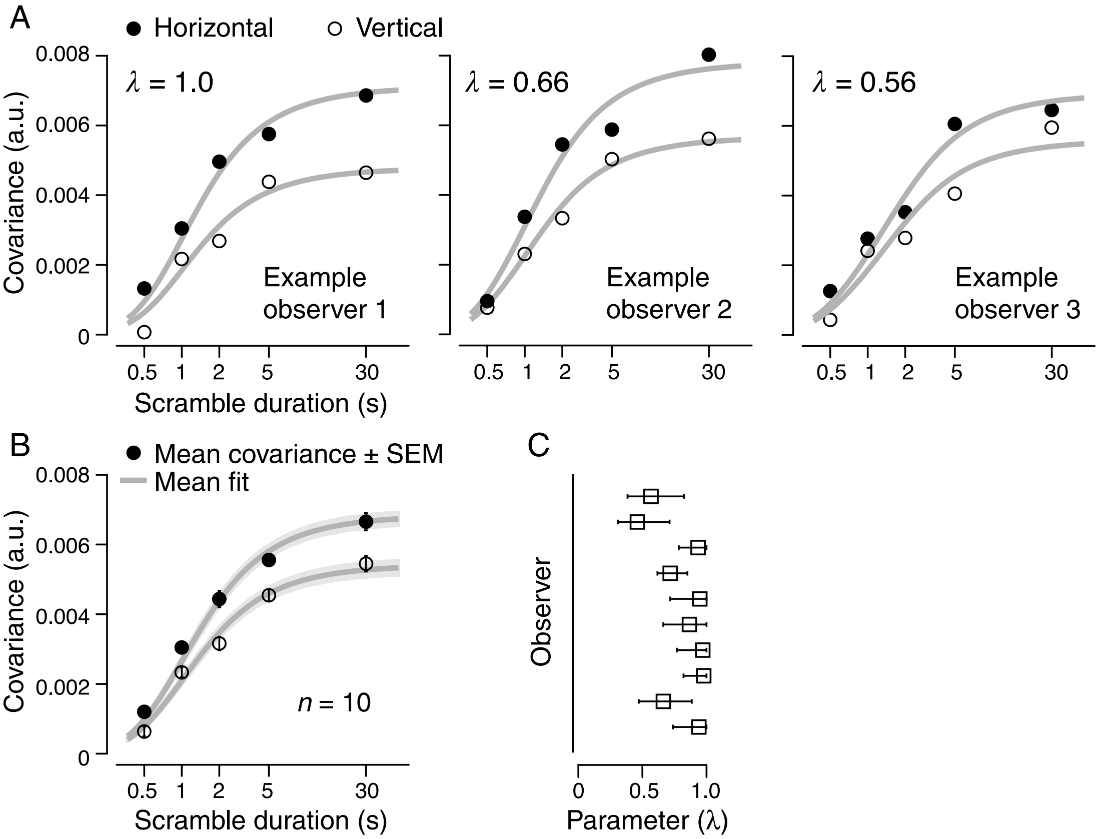

Figure 5. Model fits. (A) Eye movement reliability as a function of scramble duration for individual observers. Circles, covariances between the unscrambled eye movement time courses from a single observer for each scramble duration and median eye movements across observers for the intact movie. Filled and open circles, covariances for horizontal and vertical eye movements, respectively; a.u., arbitrary units. Gray curves, best fit of the model. The median eye movement time course for the intact movie was used as a proxy for the "point of interest" trajectory in the model (see Figure 4). The three free parameters were: λ, probability of locking onto the point of interest on each fixation after a saccade; Q_H and Q_V, the asymptotic covariances for horizontal and vertical eye movements. (B) Eye movement reliability averaged across observers (n = 10). Filled and open circles, average covariances for horizontal and vertical eye movements, respectively (replotted from Figures 3A and 3B, large symbols). Error bars, *SEM* across observers. Model was fit to each individual observer and individual fits were averaged. Gray curves, mean fit. Light gray shaded area, confidence interval on the mean fit, computed by taking the standard error across individual fits. (C) Best-fitting value of the parameter λ, which corresponded to the probability that each fixation locked onto a point of interest. Error bars, 95% confidence intervals obtained through bootstrapping (see Materials and methods section).

scenes in a different order (see Stimuli and experimental procedure section). The model provided a good fit for those observers as well, with fitted values of λ = 0.82, 0.73 (additional observer 1, two separate presentations of the interleaved scene) and 0.56, 0.99 (observer 2), comparable to those obtained for the original 10 observers (p = 0.35, randomization test).

While the estimate of λ was sensitive to the parameter estimates of the saccade latencies used during the fit, the expected time to find the point of interest (computed from a combination of the fitted value of λ and inter-saccade interval parameters) did not depend on the specific parameterization of saccade times. Compared to the overall distribution of saccade latencies, inter-saccade intervals tended to be shorter immediately after a cut. Therefore, our use of saccade parameters derived from eye positions from the overall intact scene was an oversimplification. Using shorter inter-saccade intervals for the fit yielded smaller values of λ than reported above. However, we verified that the overall expected time to find the point of interest remained the same. This means that when saccade latencies were shorter, the probability of finding the point of interest after each saccade was consequently lower, resulting in a greater number of saccades to reach the point of interest.

A complementary analysis confirmed predictions of the model

To further confirm the appropriateness of the model, we validated the time needed to lock on to the point of interest (determined by fitted values of λ and saccade latencies) using a complementary and independent analysis (Figure 6). Deviations between unscrambled eye movements and the estimated point of interest trajectory ("eye position error") were computed as a function of time after each cut (Figure 6B). Eye position error started high right after a cut, decreased sharply, but showed a gradual increase over time. However, the eye position error at any time point after a cut depended not only on the difference between the unscrambled eye position and the estimated point of interest but also on the variance of unscrambled eye movements (Figure 6A), which showed a similar decrease and then increase over time. To isolate the component of the eye position error that was independent of eye movement variance, we first computed the maximal eye position error, which reflected how eye position error would have evolved if the unscrambled eye movements never locked on to the point of interest (error was always maximal). This maximal eye position error was computed from the variance (across clips) of the unscrambled eye movements (Figure 6A) and the variance (over time) of the point of interest (see Fractional explained variance: Derivations section in Appendix A). One minus the ratio between the measured and maximal eye position errors

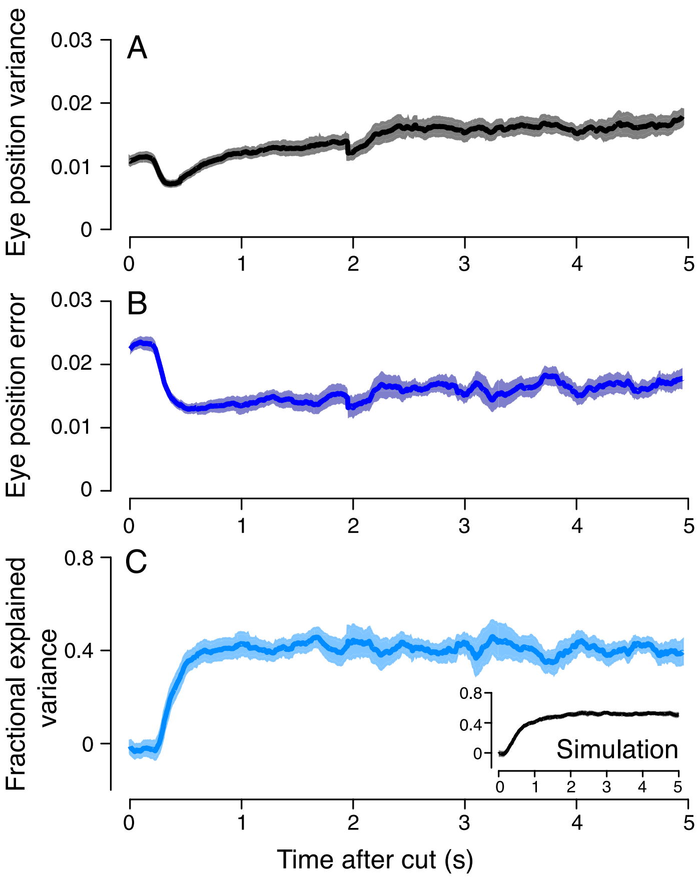

Figure 6. Reliability of eye movements over time. (A) Variance in eye position as a function of time after a cut. Variance was computed at each time point, across all clips, separately for each observer. Black curve, mean across observers (n = 10). Shaded area, *SEM* across observers. Results are shown (in this and the other panels) for horizontal eye movements; those for vertical eye movements were similar. Data points (in this and the other panels) shortly after a cut were averaged across more clips than later time points. (B) Eye position error as a function of time after a cut. Purple curve, the squared position difference between the measured time courses and the median across observers, averaged across all clips and averaged across observers. Shaded area, *SEM* across observers. (C) Fractional explained variance as a function of time after a cut (see Eye position error, variance in eye position, and fractional explained variance section). Light blue curve, mean across observers (n = 10). Shaded area, *SEM* across observers. This represents how well the dynamics of the median eye movement time course accounted for the dynamics of the unscrambled time courses, irrespective of the variance in eye position (panel A). Values near zero indicate that the median eye movement time course did not account for the unscrambled time courses, and a value of 1 indicates that the median matched the unscrambled time courses completely. Inset: Simulated fractional explained variance (see Simulating fractional explained variance section in Appendix A). Shaded region, *SEM* across simulations for individual observers.

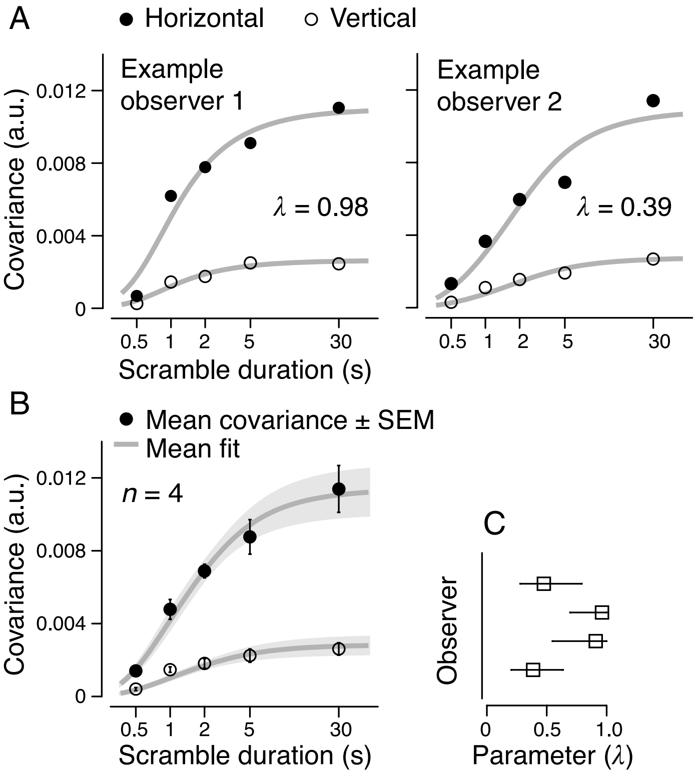

Figure 7. Data and model fits for the *Russian Ark* movie. (A) Covariance as a function of scramble duration for two sample observers viewing stimuli from the *Russian Ark* movie. Same conventions as in Figure 5A. (B) Average covariance (n = 4 observers) and model fits. Same conventions as in Figure 5B. (C) Bootstrapped best-fitting values of the λ parameter. Same conventions as in Figure 5C.

("fractional explained variance," Figure 6C) indicated how well the trajectory of the point of interest predicted the measured eye movements, independent of the eye movement variance (see Fractional explained variance: Derivations section in Appendix A). The fractional explained variance started to increase about 0.2 s after cuts and flattened out after 0.5–0.8 s. This shows that the point of interest predicted the unscrambled eye movements poorly right after a cut but did better given enough time, consistent with the model's prediction that eye movements start out uncorrelated with the point of interest and then converge. Time courses of fractional explained variance computed separately for each scramble duration were nearly identical, consistent with the model's assumption that convergence on to the point of interest, on average, depended only on the amount of time the observer had to view the clip after a cut. The results are also consistent with the idea that the search-and-track process reset following each cut.

Model simulations using the fitted values of λ and the best-fitting lognormal parameters of saccade latencies (μ and σ; see Equation A1 in Appendix A) showed a similar fractional explained variance (Figure 6C, inset), which also started to increase at 0.2 s after cuts and achieved asymptote around 1 s. The simulated fractional variance showed a more gradual rise, which might be due to our imperfect assumption that each saccade was independent and had a fixed probability of finding the point of interest (see Integration of visual information across fixations during search section). Despite this difference, the probability of finding the point of interest averaged over the initial few fixations was similar for both the measurement and the simulation, consistent with the predictions of our model. This analysis also revealed the fine-grained temporal dynamics of locking on, an aspect of the results (and the model) not fully captured by the covariance analysis.

The variance of eye position dipped shortly after a cut and gradually increased over a period of several seconds (Figure 6A). The mean eye position remained close to the center of the screen (data not shown), so we interpret the change in variance as a tendency for eye positions to converge to the center of the screen right after a cut. This is consistent with the evidence that observers tend to orient toward the center of the screen after stimulus onset (Parkhurst et al., 2002; Tatler, 2007; Tseng, Carmi, Cameron, Munoz, & Itti, 2009) and keep their eyes concentrated near the center during rapid scene cuts (Tosi et al., 1997). Some of the change in variance, over time after a cut, might also reflect an increased tendency to make exploratory eye movements to look for another point of interest long after a cut. This time-dependent change in variance also explains why the variance of eye positions depended on scramble duration. This change in variance, however, did not affect the average reliability of eye movements as measured by covariance. We teased apart the effect of eye position variance from eye movement reliability, by computing the fractional explained variance, or the proportion of total variance at any time that may be accounted for by the point of interest (Figure 6C). Nonetheless, the non-stationary variance of eye positions reveals an interesting aspect of the data not captured within the scope of the model.

Model accounted for the eye movement reliability of a second movie

To test whether the model generalized across stimuli, we tested a smaller group of observers on a second movie with a very different pace and cinematography (*Russian Ark*, 2002). Eye movements for this movie showed a similar relationship between scramble duration and covariance (Figure 7), confirming that the dependence of eye movement reliability on scramble duration was not specific to the choice of film. The model fit the eye movements well (r^2 = 0.91, cross-validated 2.5th–97.5th percentiles = 0.66–0.98), yielding values of λ qualitatively similar to those estimated with the other movie (compare Figures 5C and 7C) and an expected time of 0.85 ± 0.40 s (mean ± standard deviation, $n = 4$) within which observers were able to find the point of interest. It remains to be tested whether the model would yield substantially different results for other classes of movies. A starting assumption of the model is that the unperturbed eye movements are reliable (high covariance between eye movements for the intact movie). Since reliability depends on the content of a movie (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Hasson, Landesman et al., 2008; Shepherd et al., 2010), e.g., differences in the degree to which the stimulus engages an observer, it is possible that we would observe different results for a substantially different choice of film (e.g., a static scene without action or movement or a scene with many cuts). Nonetheless, the fact that our model provided a good fit for two very different movie stimuli is consistent with its content-free nature and suggests some degree of generalizability.

Discussion

Engaging movies evoke highly consistent and reproducible eye movements (Goldstein et al., 2007; Hasson, Landesman et al., 2008; Hasson et al., 2010; Hasson, Yang et al., 2008; Orban, 2008; Tosi et al., 1997). We exploited this high reliability of eye movements and parametrically varied the temporal structure of two movie stimuli by scrambling them at different temporal scales. Our scrambling manipulation preserved the frame-by-frame features in the original stimulus, while disrupting the temporal relationships portrayed by the scene in a content-independent manner. Eye movements for the intact scene were compared to those obtained for the same content presented within a scrambled context, which allowed us to assess the extent to which eye movements depended on the instantaneous properties of a scene versus its temporal context. Reliability of eye movements decreased with shorter scrambling durations, in a manner that was consistent across multiple observers and two movies. We characterized the effect of scrambling with a simple model in which eye movement reliability arises from observers tracking a relevant point of interest on the screen and in which the tracking process reset with every cut. Fits from the model for the two movies yielded parameters that corresponded to an expected time of ~0.8 s, within which observers were able to find and lock on to the point of interest; this value was independently verified in a separate complementary analysis. The explanatory power of this simple model suggests that the temporal accumulation of information over time periods exceeding a second is not needed to explain our data. That is, a simple, memory-less model captured the reliability of eye movements to complex, dynamic scenes with different degrees of temporal scrambling.

Spatial factors that drive eye movements

Early work on eye movements using still images, like photographs and line drawings, found that certain locations consistently attracted an observer's fixation during free viewing (Buswell, 1935; Yarbus, 1967). Since then, many studies have shown that fixated locations tend to differ from non-fixated locations in a number of low-level statistics, such as local intensity, color, and orientation (Krieger et al., 2000; Mannan et al., 1995; Parkhurst & Niebur, 2003; Rajashekar et al., 2007; Reinagel & Zador, 1999; Tatler, Baddeley et al., 2005). These findings have led to computational models that predict fixation locations by extracting the "saliency" (conspicuity) of local features in a scene (Itti & Koch, 2001; Koch & Ullman, 1985; Parkhurst et al., 2002; Peters et al., 2005; Tatler, Baddeley et al., 2005). The computations embodied in such bottom-up models connect elegantly with known aspects of neural processing in cortical visual areas. They provide a quantitative and principled approach for relating eye movement behavior to a stimulus.

Eye movements are also influenced by many other cognitive factors not predicted by feature saliency. For example, in the presence of a task, eye movements depend on the task demands and the observer's internal goals (Buswell, 1935; Hayhoe & Ballard, 2005; Land, 2009; Land & Hayhoe, 2001; Noton & Stark, 1971; Rothkopf et al., 2007; Turano, Geruschat, & Baker, 2003; Yarbus, 1967). Contextual knowledge based on the co-occurrence of objects (e.g., a plate on a dining table) and semantic content of the scene can facilitate the selection of attentional targets and bias gaze strategy (Eckstein, Drescher, & Shimozaki, 2006; Henderson et al., 1999; Neider & Zelinsky, 2006; Torralba et al., 2006). In fact, it has been argued that bottom-up saliency does not necessarily drive eye movements causally, as the local image statistics underlying saliency are also correlated with higher level scene content (such as semantic informativeness; Einhauser & Konig, 2003; Einhäuser, Spain, & Perona, 2008; Henderson, Brockmole, Castelhano, & Mack, 2007). Several models for predicting fixation locations incorporate both bottom-up and top-down elements. In some implementations, saliency maps are selectively modulated by information that reflect top-down control or prior expectations (e.g., about the location or features of a target object), based on knowledge of a task or an understanding of scene gist (Navalpakkam & Itti, 2005; Oliva, Torralba, Castelhano, & Henderson, 2003; Peters & Itti, 2007; Torralba et al., 2006). In other implementations, a probabilistic model learns preattentive targets from scene statistics, therefore combining both bottom-up saliency and top-down biases (Butko, Zhang, Cottrell, &

Movellan, 2008; Kanan, Tong, Zhang, & Cottrell, 2009; Yamada & Cottrell, 1995; Zhang, Tong, & Cottrell, 2009). Alternatively, some models integrate both saliency and top-down information at the level of object representation (Sun, Fisher, Wang, & Gomes, 2008; Wischnewski, Belardinelli, Schneider, & Steil, 2010), reflecting the hypothesis that the "proto-object" (i.e., the position and a cluster of features relevant to an object) represents the basic unit for prioritizing attention (Einhäuser et al., 2008; Hollingworth & Henderson, 2002; Scholl, 2001). The combination of bottom-up and top-down information outperforms purely bottom-up models when fixations are of immediate behavioral relevance, such as during search tasks (Kanan et al., 2009; Navalpakkam & Itti, 2005; Oliva et al., 2003; Torralba et al., 2006) or tasks involving interactive viewing (e.g., video game playing; Peters & Itti, 2007). Finally, socially relevant cues not predicted by saliency models, such as faces, gaze direction, and body movement, also serve as powerful predictors of eye movements (Birmingham et al., 2008; Friesen & Kingstone, 1998; Shepherd et al., 2010).

Temporal factors that drive eye movements

A variety of dynamic stimulus types and approaches have been used to explore how viewing behavior depends on continuously changing visual information (e.g., Butko et al., 2008; Carmi & Itti, 2006a, 2006b; Dorr et al., 2010; Goldstein et al., 2007; Hasson, Landesman et al., 2008; Itti, 2005; Itti & Baldi, 2005, 2009; Le Meur, Le Callet, & Barba, 2007; Rothkopf et al., 2007; Shepherd et al., 2010; Wischnewski et al., 2010). Some of this work has focused on the perception and representation of scene information across time. Complex, dynamic scenes often contain editorial cuts, such as viewpoint switches; changes in scene content across these cuts or the cuts themselves may go unnoticed by the observer (Bordwell & Thompson, 2001; Levin & Simons, 1997, 2000; Reisz & Millar, 1953; Smith & Henderson, 2008). Extending research on change detection and memory representation of static scenes (e.g., Grimes, 1996; Henderson & Hollingworth, 2003; Hollingworth & Henderson, 2002; Irwin & Zelinsky, 2002; McConkie & Currie, 1996; Melcher, 2006; Melcher & Kowler, 2001; O'Regan, 1992; Rensink, O'Regan, & Clark, 1997; Tatler, Gilchrist, & Land, 2005; Tatler, Gilchrist, & Rusted, 2003), many studies have examined the perceptual and memorial consequences of changes in a dynamic scene (Angelone, Levin, & Simons, 2003; Garsoffky, Huff, & Schwan, 2007; Garsoffky, Schwan, & Hesse, 2002; Kraft, 1986; Levin & Simons, 1997, 2000) and their interactions with eye movements (d'Ydewalle, Desmet, & Van Rensbergen, 1998; d'Ydewalle & Vanderbeeken, 1990; Germeys & d'Ydewalle, 2007; Hirose, Kennedy, & Tatler, 2010; Smith & Henderson, 2008). For example, Hirose et al. (2010) found that eye movement behavior reflected observers' differential sensitivity to object property and location changes across viewpoint switches. Smith and Henderson (2008) found that undetected editorial cuts ("edit blindness") in feature films appeared to depend mainly on inattentional blindness induced by the content of the new shot rather than coincidence with periods of perceptual insensitivity induced by saccades or blinks.

Computational models of eye movements have also been extended to explain how eye movements depend on temporal features within dynamic scenes under free viewing (e.g., Itti, 2005; Itti & Baldi, 2005, 2009; Kienzle, Schölkopf, Wichmann, & Franz, 2007; Le Meur et al., 2007; Peters & Itti, 2007; Vig, Dorr, & Barth, 2009; Wischnewski et al., 2010; Zhang et al., 2009). Spatiotemporal versions of saliency models reveal that motion contrast and temporal novelty serve as strong predictors for locations of eye movements (e.g., Itti, 2005; Itti & Baldi, 2005, 2009). Motivated by the importance of temporal salience on eye movements, two studies investigated how eye movements depended on temporal continuity of a scene by comparing eye movements for a continuous movie and sequences of static frames from the same movie. 't Hart et al. (2009) recorded eye movements during free exploration of indoor and outdoor environments and compared them to those during head-fixed replays of the same visual input (either dynamic or static versions) in the laboratory. They found that eye movements during continuous replay movies predicted real-world gaze positions better than those during shuffled sequences of 1-s still frames and better than those predicted by a static model saliency map. This confirmed that temporal continuity played an important and consistent role in influencing eye movements during different types of dynamic visual inputs. Furthermore, static model saliency yielded better predictions of eye positions during continuous replay movies than did eye positions during 1-s still frames, suggesting that a consequence of temporal continuity was a larger dependence of eye movements on bottom-up spatial information. In addition, similar to what we found for eye positions during short scramble durations of movie clips, 't Hart et al. found that eye position for the still frames showed a stronger spatial bias toward the stimulus center (Buswell, 1935; Parkhurst et al., 2002; Tatler, 2007; Tseng et al., 2009), which contributed substantially to inter-observer consistency. Another study used Normalized Scanpath Saliency (Peters et al., 2005) as a metric to quantify inter-observer consistency in eye movements (Dorr et al., 2010). Their measure of consistency was quite different from the covariance-based measure of reliability that we used. They found that the time course of inter-observer consistency differed substantially between the continuous and static frame versions of homemade natural movies (e.g., a busy roundabout intersection with moving cars). Consistency of eye movements for static frames (sampled at a regular interval from the continuous scene and shown for 3 s at a time) was high immediately after each frame transition but dropped sharply until the next frame onset. However, in the continuous version of the movie, consistency peaked immediately after movie onset and remained at a modest level throughout the rest of the presentation. Like 't Hart et al., Dorr et al. (2010) also found that much of their inter-observer consistency was dominated by the tendency to fixate the center of the screen after each onset, independent of the specific visual stimulus.

Other studies have examined how editorial visual disruptions in dynamic scenes impact the temporal dynamics of eye movements. Vig et al. (2009) quantified the time delay between visual events in video clips and the responding eye movements during free viewing, by cross-correlating saliency maps and spatiotemporal fixation maps and identifying the time shift at which the two maps showed maximal correlation. They found that the lag was near zero for a database of dynamic natural scenes shot with a static camera (e.g., populated streets and parks) but much longer (133 ms) for a separate database with video clips containing editorial transitions, such as camera movements (pan, tilt, zoom), special effects (fade, dissolve, wipe), and jump cuts. They reasoned that whereas eye movements are usually slightly anticipatory (e.g., looking ahead of the movement) for continuous scenes, the presence of cuts and other editing techniques introduce temporal discontinuities that interrupt that anticipation. Finally, Carmi and Itti (2006a, 2006b) examined the evolution of eye movements over time following rapid-transition jump cuts in dynamic scenes. They found that eye movements were well predicted by a saliency model shortly after a cut, but prediction accuracy diminished over a period of 2.5 s across several fixations. They explained these results in terms of a competition between bottom-up processes and top-down processes that depended on "perceptual memory," which we interpret as including any process that integrates information across time.

Like Carmi and Itti (2006a, 2006b), our study also explored how the factors driving eye movements evolve over time. We took advantage of the fact that a class of dynamic stimuli—high-production films—elicits highly reliable eye movements across observers. We manipulated and modeled that reliability to draw conclusions about observers' viewing behavior, specifically, how it depended on temporal context. This provided a complementary approach for studying the temporal dependence of eye movements without explicitly modeling their governing factors or predicting them directly as a function of the stimulus. While we found no evidence that eye movement reliability depended on visual information accumulated over time, our model is agnostic as to whether such information represents low-level or high-level cues. As discussed above, many high-level processes besides saliency, such as contextual and social cues, can also guide eye movements on a fast temporal scale. Furthermore, by design we modeled only the reliable component of eye movements, namely, the point of interest that captured the variability in eye position over time that was shared across observers. Any deviation from the point of interest (considered "noise" in our model) likely reflected sources of variability other than measurement noise, which could include idiosyncratic viewing strategies that may or may not depend on temporal context, as well as systematic tendencies to fixate certain locations on the screen as a function of time elapsed after a cut (see Variance of eye movements as a function of time section below). Therefore, it remains an open question how the declining impact of feature saliency on eye movements after a cut, as found by Carmi and Itti, relates to the time course of eye movement reliability as found by our study.

Integration of visual information across fixations during search

Our model assumed that observers began tracking a point of interest after some delay following a cut but was agnostic with respect to what happened in the few hundred milliseconds before observers found the point of interest. The predictions of the model only required that, during this time, eye movements were uncorrelated with the point of interest trajectory. Specifically, we assumed that each saccade before the observer locked on had a fixed probability of finding the point of interest (parameter λ in Equation A3, see Appendix A and Figure 5C). We could not, however, exclude the possibility that this probability increased across fixations during the period before locking on (i.e., that information was accumulated across fixations about the likely location of the point of interest). Such a framework in which the observer uses prior information to search for relevant points of interest bears some resemblance to visual search (Treisman & Gelade, 1980). Human behavior during search has been modeled by assuming that the observer chooses where to look to maximize information about the location of the target (Najemnik & Geisler, 2005). Accordingly, visual information is integrated across fixations and updated iteratively. There is also empirical evidence for the accrual of visual information across the first two fixations during search (Caspi, Beutter, & Eckstein, 2004), within the time frame that observers typically find the point of interest for our movie stimuli. Note that visual search models indicate that human performance does not significantly depend on information integrated beyond a relatively short time scale of two fixations (Najemnik & Geisler, 2005). If the observer indeed integrates information across fixations to optimally locate the point of interest, the probability of locking on should increase with every fixation. In that case, the fitted values of λ (Figures 5C and 7C) can be thought of as an average probability of finding the point of interest over those fixations. However, this would not change the model's prediction of the average time required to find the point of interest after a cut. As such, this elaboration would not affect the model's prediction for how covariance with the point of interest depends on scramble duration. Indeed, the fractional explained variance analysis revealed that the probability (across clips) of locking on rose more sharply than predicted by the model (Figure 6C), possibly suggesting integration of visual information within the first couple of saccades (e.g., the first saccade had a lower probably of locking on than predicted, and the second saccade had a higher probability).

Variance of eye movements as a function of time

We used covariance (instead of correlation) to quantify the reliability of eye movements, and therefore, our model did not account for or depend on the variance of eye movements. It is well established that fixation locations on static images become more variable across observers with prolonged viewing (Henderson & Hollingworth, 1999; Mannan et al., 1995; Tatler, Baddeley et al., 2005), but such time-dependent increase in variability is less pronounced for dynamic videos (Dorr et al., 2010), presumably due to the impact of continuous temporal change on attentional selection (Itti, 2005; Yantis & Jonides, 1984). We calculated eye position variance across clips (rather than across observers) as a function of time after a cut and found that the distribution of eye positions depended systematically on the time elapsed since the cut. A portion of this variance might contribute to changes in interobserver variability as a function of time. Consistent with previous studies (Dorr et al., 2010; Tatler, 2007; Tosi et al., 1997; Tseng et al., 2009), we also found that the distribution of eye positions fell near the center of the screen right after a cut but gradually spread to include positions away from the center over a period of several seconds thereafter (Figure 6A). Factors underlying the change in variance may include the well-documented center bias immediately after stimulus onset (Buswell, 1935; Dorr et al., 2010; Parkhurst et al., 2002; Tatler, 2007; Tosi et al., 1997; Tseng et al., 2009), as well as a tendency to explore and look for new points of interest with prolonged viewing. The change in variance therefore revealed a separate, but nonetheless intriguing, aspect of the data not encompassed by the scope of the model.

Relationship to narrative continuity and editorial cuts

We investigated how the spatiotemporal continuity of a movie scene contributed to the reliability of eye movements, but such spatiotemporal continuity was not necessarily equivalent to narrative continuity, which is linked to the comprehension of event relationships in a scene. For example, Hasson, Yang et al. (2008) found that presenting a film backward in time preserved most of its spatiotemporal continuity while severely disrupting its narrative continuity and compromising observers' comprehension. However, the reliability of eye movements was similar for both forward and backward movies, suggesting that narrative continuity (or comprehension) was not necessary for reliable eye movements. In other instances, narrative continuity (achieved through editing techniques) may help mask spatiotemporal discontinuity and therefore enhance the reliability of eye movements.

In our experiments, we specifically used scenes that were shot as single takes without any cuts. In most feature films or TV commercials, cuts typically occur every 2–10 s, though their average frequency varies by era and by genre (Bordwell & Thompson, 2001; MacLaclan & Logan, 1993; Salt, 1992). The types of cuts intentionally placed in films ("editorial cuts") differ from the cuts produced by our scrambling manipulation, which were sharp spatiotemporal discontinuities in the scene. Most editorial cuts adhere to the conventions of film editing so as to maintain narrative continuity (Bordwell & Thompson, 2001; d'Ydewalle et al., 1998; d'Ydewalle & Vanderbeeken, 1990; Hochberg & Brooks, 1978; Kraft, 1987; Reisz & Millar, 1953; Salt, 1992). For example, viewpoints typically stay on the same side of the "axis of action," so as to preserve the left–right relationship between two characters in a scene or a character's direction of movement across cuts ("180 rule"). These techniques maintain the psychological continuity of the scene by exploiting the observer's inferences of event and spatial relationships (d'Ydewalle et al., 1998; d'Ydewalle & Vanderbeeken, 1990; Frith & Robson, 1975; Germeys & d'Ydewalle, 2007; Hochberg & Brooks, 1978; Kraft, 1987; Levin & Simons, 2000). In fact, changes across editorial cuts or the editorial cuts themselves often go unnoticed by the observer (Bordwell & Thompson, 2001; Levin & Simons, 1997, 2000; Reisz & Millar, 1953; Smith & Henderson, 2008). A well-produced film with editorial cuts can evoke highly reliable eye movements (Hasson, Landesman et al., 2008; Hasson, Yang et al., 2008), with correlations comparable to what we found for our intact scenes (approximately 0.5). Thus, although we employed single-shot scenes because of their high temporal continuity, well-designed editorial cuts likely help preserve the temporal continuity of a scene, and our experimental results would likely generalize to well-produced movie scenes containing such cuts.

Nonetheless, the effectiveness of editorial cuts depends greatly on their composition, and different types of editorial cuts may differentially affect the perceived continuity of the scene as well as the observer's viewing behavior (d'Ydewalle et al., 1998; d'Ydewalle & Vanderbeeken, 1990; Dmytryk, 1986; Germeys & d'Ydewalle, 2007; May, Dean, & Barnard, 2003; Smith & Henderson, 2008).

For example, Smith and Henderson (2008) found that an observer was less likely to notice a cut if it stayed within the same scene and coincided with a sudden onset of visual motion. Failure to notice changes across the type of transitions common in editorial cuts (e.g., viewpoint change) is linked to inattentional blindness and change blindness (Levin & Simons, 1997, 2000; Mack & Rock, 1998; Rensink et al., 1997); the retention of information across these transitions and how that interacts with eye movements are areas of active study (e.g., Hirose et al., 2010; Smith & Henderson, 2008).

In our experiments, scrambling served as an experimental manipulation for varying the temporal structure of the movie scenes. We specifically employed single-shot scenes to ensure high temporal continuity in the unscrambled stimulus. Our artificial jump cuts introduced substantial visual disruption to the temporal structure, in a manner that was independent of the underlying content of the scene. There are two alternative manipulations that we could have used for scrambling, but both would have limited our experimental control. One possibility would have been to employ a conventional scene with existing editorial cuts and scramble the temporal order of that scene by introducing new cuts. The resulting interleaved movie would contain both the original cuts and the ones we inserted. If all the editorial cuts in the scene were well designed, they should only minimally impact eye movements, and we would expect to obtain similar results to those obtained for a single-shot scene. However, as discussed above, the cognitive effects exerted by editorial cuts can vary greatly depending on their type and the filmmaker's style, therefore introducing additional visual disruptiveness outside of our experimental control. A second possibility would have been to employ a conventional scene with existing editorial cuts and shuffle only those cuts. However, the distribution of clip length would depend on the film and again lie outside of our experimental control, making it impossible to precisely manipulate the scramble duration or to apply the same manipulation to different movies. Furthermore, the type of visual disruptiveness introduced by shuffling only editorial cuts may differ systematically from that introduced by inserting jump cuts, as editorial cuts often depict breakpoints marking a shift in action or perceptual events (Carroll & Bever, 1976; Schwan, Garsoffky, & Hesse, 2000). Thus, inserting artificial jump cuts as we have done allowed us to take control over the visual disruption and scramble duration of our stimuli, independent of the choice and content of the movie scene.

While we focused on a specific set of scrambling manipulations applied to continuous single-shot video sequences, the derived model parsimoniously captured the data set and represents a general (and thereby testable) hypothesis for how eye movement reliability depends on temporal context for naturalistic, dynamic stimuli. How well the model can account for eye movements for broader sets of stimuli, such as films with less editorial structure, remains a question for further study.

Appendix A

Model derivation

We derive the relationship between eye movement covariance and scramble duration. The mathematical notations used in the derivation and their descriptions are listed in Table A1. We begin by finding an expression for the probability that an observer will fixate the point of interest as a function of time, t. The intervals at which an observer makes saccades are well described by a lognormal distribution, with parameters that can be estimated directly from our data (Figure 4A). The lognormal probability density function with parameters μ and σ is defined as

$$f(t|\mu,\sigma) = \frac{1}{t\sigma\sqrt{2\pi}} e^{\frac{-(\ln t-\mu)^2}{2\sigma^2}}. \tag{A1}$$

Assuming consecutive inter-saccade intervals are independent, the time of the jth saccade is the sum of j random variables, each with a probability density distribution $f(t|\mu, \sigma)$. We define the probability density distribution for the time of the jth saccade as $z_j(t)$. The pdf $z_j(t)$ is the convolution of lognormal pdf $f(t)$ with itself $j - 1$ times. For $j > 1$, this expression has no closed form, so we used an approximation. The convolution of $j - 1$ identical lognormal functions f with parameters μ and σ is commonly approximated by another lognormal distribution, $f(t|\mu_j, \sigma_j)$, where

$$\sigma_j = \sqrt{\ln\left(\frac{e^{\sigma^2}-1}{j}+1\right)}, \tag{A2}$$
$$\mu_j = \ln(je^{\mu}) + \frac{\sigma^2}{2} - \frac{\sigma_j^2}{2}.$$

In making this approximation, the first and second moments (i.e., the mean and variance) of $f(t|\mu_j, \sigma_j)$ were matched to j times those of $f(t|\mu, \sigma)$ (Fenton–Wilkinson method; Fenton, 1960). We verified in simulation that this approximation was accurate to within 1% error for the range of μ, σ, and j used in our calculations.

On each fixation following a saccade, the observer has a fixed probability λ of finding and locking onto the point of interest. Thus, the probability of finding the point of interest precisely on the jth fixation is $\lambda(1 - \lambda)^{j-1}$. This is

Notation	Type	Definition
L	Constant	Duration of the original scene
d	Constant	Duration of each clip used to evenly divide up the scene
S	Time course	Point of interest time course (median eye movement time course for intact movie)
S^d	Time course	Unscrambled eye movement time course for scramble duration d
$f(\mu, \sigma)$	Function	Lognormal probability density function describing inter-saccade intervals; depends on μ and σ
z_j	Function	Probability density function describing the time of the jth saccade
p_T	Function	Probability density function describing the likelihood of finding a point of interest as a function of time after a cut onset; depends on f and λ
P_T	Function	Cumulative density function of p_T
C	Function	Covariance between two eye movement time courses
T	Variable	A continuous random variable with pdf p_T
t	Variable	Time after a cut (a value for random variable T with pdf p_T; $\Pr(t < T < t + dt) = p_T\,(t)dt$ for an infinitely small interval dt)
τ	Variable	A random variable describing the amount of time an observer is not locked onto the point of interest within a cut; depends on T and d
μ, σ	Parameters	Parameters governing the shape of the lognormal pdf f
μ_j, σ_j	Parameters	Parameters governing the shape of the lognormal pdf for the time of the jth saccade (an approximation for $z_j(t)$)
λ	Parameter	Probability of finding and locking onto a point of interest on each fixation after a saccade
Q	Parameter	Maximal covariance between an intact eye movement time course and the trajectory of the point of interest, as limited by noise
N	Empirical measure	Total number of samples in the time courses (corresponding to a total time duration of L)
N_0	Empirical measure	Number of samples (out of N) that the observer is not locked on to the point of interest
S_0	Time course	Random point of interest with the same distribution as entries in S, i.e., with mean and variance μ_S and v_S
μ_{Sd}	Function	Mean of unscrambled eye movements, i.e., mean of S^d (stationary with respect to time after a cut)
μ_S	Function	Mean of the point of interest, i.e., mean of S (stationary with respect to time after a cut)
$v_{Sd}(t)$	Function	Variance of unscrambled eye movements after a cut, i.e., variance of $S^d(t)$
V_S	Function	Variance of the point of interest, i.e., variance of S (stationary with respect to time after a cut)
$G(t)$	Function	Eye position error between the unscrambled time courses and the point of interest after a cut; $G(t) = E[(S^d(t) - S(t))^2]$
$G_0(t)$	Function	Maximal eye position error or the eye position error expected between the unscrambled time courses and an uncorrelated, random point of interest after a cut; $G_0(t) = E[(S^d(t) - S_0)^2] = v_{Sd}(t) + v_S$

Table A1. Notation for derivations in Appendix A.

the probability of finding the point of interest on the jth fixation times the probability of not finding it on all previous fixations.

We introduce the probability density function $p_T(t)$ for a continuous random variable T, which describes the probability of finding a point of interest over time after a cut:

$$p_T(t) = \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} z_j(t) \approx \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} f(t|\mu_j, \sigma_j), \tag{A3}$$

where $f(t|\mu_j, \sigma_j)$ is the lognormal approximation for $z_j(t)$, the probability density distribution for the time of the jth saccade, and parameters μ_j and σ_j are related to μ and σ as in Equations A2. Note that for $j = 1$, $z_1(t) = f(t|\mu, \sigma)$. Consistent with standard probability notation, lowercase t denotes a specific value for the random variable T:

$$\Pr[a \le T \le b] = \int_a^b p_T(t)dt. \tag{A4}$$

The approximated form of function $p_T(t)$ in Equation A3 is a sum of a series of lognormal distributions. Each lognormal distribution describes the time of an individual saccade, and each distribution is weighted by the probability of finding the point of interest following that saccade. When $\lambda = 1$, the observer always fixates the point of interest after the first saccade, and $p_T(t)$ is equal to a lognormal distribution describing the time of that saccade (all terms in the summation where $j > 1$ equal 0). For smaller λ, more saccades are required to find the point of interest, and the shape of $p_T(t)$ changes to have larger probabilities associated with later saccades (Figure 4B).

The cumulative distribution associated with the density $p_T(t)$ is

$$P_T(t) = \int_0^t p_T(x)dx. \tag{A5}$$

$P_T(t)$ increases to 1 as t increases ($P_T(t) \rightarrow 1$ as $t \rightarrow \infty$), which means that the probability of finding the point of interest converges to 1 as the amount of time allotted to find it increases (Figure 4C). For larger values of λ, the slope is steeper, i.e., $P_T(t)$ converges to 1 more quickly. When there is only a finite amount of time in a clip (i.e., t is bounded), and especially when λ is small, $P_T(t)$ may still be far from 1 even when t achieves its maximum value. That is, there is a non-trivial probability that the observer will not have found the point of interest before the end of the clip.

To predict covariance, we derive an expression for the average amount of time during which the observer is locked on (or not locked on) to the point of interest. An intact scene of length L is divided evenly into clips, each of length d, whose order may be randomly scrambled. Over the entire movie, there are L/d clips in total. For each clip, the above probability distributions are used to estimate the average value of a random variable τ, which describes the time during which the observer is not locked on to the point of interest for that clip. For any given clip, the maximal value that τ can take is d. When τ is less than d, the value of τ depends on the value of the random variable T with probability density function given by Equation A3. Thus, a natural choice is to define the variable τ piecewise:

$$\tau = \begin{cases} T, & T < d \\ d, & T \geq d. \end{cases} \tag{A6}$$

By the law of total expectation, the expected value of τ is given by

$$\mathrm{E}(\tau) = \Pr(T < d)\mathrm{E}(\tau | T < d) + \Pr(T \geq d)\mathrm{E}(\tau | T \geq d), \tag{A7}$$

where T is the random variable with density $p_T(t)$ and cumulative distribution $P_T(t)$ as given above. The two expectations on the right-hand side are

$$\begin{array}{lll} \mathrm{E}(\tau | T < d) = \mathrm{E}(T | T < d) = \dfrac{\int_0^d t p_T(t)dt}{P_T(d)} & \text{if observer locks on before end of cut} & \\ \mathrm{E}(\tau | T \geq d) \ = d & \text{if observer does not lock on} & . \end{array} \tag{A8}$$

Suppose $S(t)$ is the "correct" eye position (as a function of time t) corresponding to the point of interest in the intact movie, and $S^d(t)$ is the unscrambled eye movement time course for scramble duration d. The expected duration of $S^d(t)$ that is not correlated with $S(t)$ is therefore $\mathrm{E}(\tau)$ summed over all L/d cuts: $(L/d)\mathrm{E}(\tau)$.

We define Q to be the covariance between an intact eye movement time course made by the observer and the trajectory of the point of interest. If the observer's eye positions matched the location of the point of interest perfectly when locked on (i.e., there was no noise or variability), Q would simply be the variance of $S(t)$. The actual value of Q depends on both the measurement noise and the observer's cognitive and motor variability (including exploratory eye movements to look for a new point of interest; see Model section). We interpret Q as the maximal covariance attainable for that observer.

We further assume that the covariance between the two eye movement time courses (intact and unscrambled) is proportional to the maximal covariance, Q, times 1 minus the fraction of time during which the eye movements are uncorrelated (see Linearity assumption section below). By this assumption, the covariance C between $S^d(t)$ and $S(t)$ is given by

$$C(S^d, S) = Q\left(1 - \frac{\frac{L}{d}\mathrm{E}(\tau)}{L}\right). \tag{A9}$$

Substituting in the conditional probabilities from Equation A7 for $E(\tau)$, we obtain

$$C(S^d, S) = Q\left(1 - \frac{\frac{L}{d}(\Pr(T < d)\mathrm{E}(\tau | T < d) + \Pr(T \geq d)\mathrm{E}(\tau | T \geq d))}{L}\right). \tag{A10}$$

Since $\Pr(T \geq d) = 1 - \Pr(T < d)$, and $\Pr(T < d)$ is the cumulative distribution $P_T(t)$ evaluated at d, substituting in Equation A8 and canceling out L yields

$$C(S^d, S) = Q\left(1 - \frac{P_T(d)\dfrac{\int_0^d t p_T(t)dt}{P_T(d)} + d(1 - P_T(d))}{d}\right). \tag{A11}$$

Simplifying (canceling d in the second term and the two 1s and rearranging terms) gives

$$C\left(S^d, S\right) = Q\left(P_T(d) - \frac{\int_0^d tp_T(t)dt}{d}\right), \tag{A12}$$

where $p_T(t)$ is the pdf defined in Equation A3 and its cumulative distribution $P_T(t)$ may be computed through numerical integration. Note that this derivation is independent of the specific parameterization of saccade times. Any distributional form for saccade times can be plugged in to the equations for $p_T(t)$ and $P_T(t)$ to obtain an expression for predicted covariance. We used the lognormal distribution, which we observed to be a good description of the inter-saccade interval distribution (Figure 4A).

Linearity assumption

In the above derivation, we assumed that the covariance between two eye movement time courses (intact and unscrambled) was proportional to the maximal covariance, Q, times 1 minus the fraction of time during which the eye movements were uncorrelated (Equation A9). Here, we provide mathematical intuition for why this relationship holds and show that it is a reasonable assumption for our data.

The sample covariance between two measured eye movement time courses S^d and S is computed as

$$\hat{C}\left(S^d, S\right) = \frac{1}{N}\sum_{k=1}^{N}\left(S^d(k) - \mu_{S^d}\right)\left(S(k) - \mu_S\right), \tag{A13}$$

where μ_{Sd} and μ_S are the sample means of S^d and S, index k indicates individual measurement samples, and N is the total number of samples in the time courses (corresponding to a total time duration of L).

In Equation A9, the covariance is expressed as proportional to the expected time during which the observer is not locked on to the point of interest. We want to show that according to our model, the empirical covariance expressed in Equation A13 can be approximated with Equation A9. To do so, we show that the form of Equation A13 simplifies greatly when considering the case in which the two signals are maximally correlated for only a subset of samples (i.e., the time points during which the observer is locked on). Specifically, assume that the observer is not locked on to the point of interest for N_0 measurement samples (uncorrelated) and is locked on for $N - N_0$ samples (with maximal covariance). Additionally, assume that the individual samples of S^d and S in Equation A13 are independent and that the sample mean μ_{Sd} does not change as a function of N_0. It follows that for any value of N_0, the product $(S^d(k) - \mu_{Sd})(S(k) - \mu_S)$ summed over N_0 out of the N terms will be approximately 0 (because S^d and S are uncorrelated for those terms), and the remaining $N - N_0$ samples will constitute $1 - N_0/N$ of the maximal covariance Q. Thus, for a finite sample,

$$\hat{C}\left(S^d, S\right) \approx Q\left(1 - \frac{N_0}{N}\right), \tag{A14}$$

and equality holds in the limit of infinite samples. The right-hand side of Equation A14 is just a discrete time version of Equation A9: The number of samples N_0 corresponds to the time $(L/d)\mathrm{E}(\tau)$ during which the observer is not locked on, and the total number of samples corresponds to the total time L. Thus, if Equation A14 holds for our data, it validates the assumption of the model as expressed in Equation A9.

We used simulations to verify that the relationship in Equation A14 holds when applied to the eye movement data measured in our experiments. Although our data did not strictly adhere to the above assumptions (for example, neighboring sample points of eye positions tended to be correlated), the simulation results showed that the linear relationship nonetheless provided a good approximation, i.e., violation of these assumptions had only a negligible effect on linearity. The simulations further suggested that the linearity assumption was more accurate for shorter scramble durations. However, deviations from linearity were small even at the longest scramble duration.

For Equation A14 to hold, only the sample mean and not the sample variance of S^d needs to be independent of N_0 (number of samples for which S^d and S are uncorrelated). In fact, the variance in eye position was smaller for shorter scramble durations (Figures 3E and 3F). However, the sample mean of S^d was approximately invariant (near the center of the screen) for unscrambled eye movement time course, as assumed in the derivation of Equation A14.

Fractional explained variance: Derivations

The deviation expected between an observer's eye position and the point of interest, under the assumption that the two are uncorrelated ("maximal eye position error"), is denoted $G_0(t)$. This value is expressed as $G_0(t) = \mathrm{E}[(S^d(t) - S_0(t))^2]$, where t is time after a cut, $S^d(t)$ is unscrambled eye movements for scramble duration d, and $S_0(t)$ is a random point of interest from the same distribution as the actual point of interest $S(t)$ but not correlated with $S^d(t)$. We show here that $G_0(t)$ is equal to the summed variances of the two underlying variables, $S^d(t)$ and $S_0(t)$.

For a moment, assume that both $S^d(t)$ and $S_0(t)$ are normally distributed at time t; $S^d(t)$ has variance $v_{Sd}(t)$ and mean μ_{Sd}, and $S_0(t)$ has variance v_S and mean μ_S. Furthermore, $\mu_{Sd} = \mu_S$ for all time points t. Note that treating v_S and μ_S as stationary with respect to t is reasonable because $S_0(t)$ has the same mean and variance as the point of interest $S(t)$; we would not expect the statistics of S to change as a function of time t after a cut from the manipulations of the interleaved movie.

Let $S^{d\prime}(t) = S^d(t) - \mu_{Sd}$ and $S_0{}'(t) = S_0(t) - \mu_S$, and we can substitute the variables in the expression of $G_0(t)$ with their mean-subtracted versions:

$$G_0(t) = \mathrm{E}[(S^{d\prime}(t) - S_0'(t))^2],$$

$$= \mathrm{E}[(S^{d\prime}(t))^2] - 2\mathrm{E}[S^{d\prime}(t)S_0'(t)] + \mathrm{E}[S_0'(t)^2]. \tag{A15}$$

The first term and last terms are simply $v_{Sd}(t)$ and v_S, respectively. The cross term $2\mathrm{E}[S^{d\prime}(t)S_0'(t)] \approx 0$ because $S^{d\prime}(t)$ and $S_0'(t)$ are uncorrelated, zero mean, and normally distributed. Therefore,

$$G_0(t) = v_{S^d}(t) + v_S. \tag{A16}$$

Note that this shows that the trajectory of $G_0(t)$ depends on the trajectory of the unscrambled eye position variance $v_{Sd}(t)$; if $v_{Sd}(t)$ was constant irrespective of time after a cut, then the maximal eye position error $G_0(t)$ would also be constant.

In our data, $S^d(t)$ was computed by aligning the unscrambled eye movements for a particular scramble duration d to each cut in that scramble duration. $S^d(t)$ computed separately for each d yielded similar curves as a function of t. Therefore, at each time point t, $v_{Sd}(t)$ may be estimated using the variance of $S^d(t)$ across all clips (n = 1344 clips from all 5 scramble durations for t = 0–0.5 s; n = 624 clips from the 4 longest scramble durations for t = 0.5–1 s; and so on). We used the median eye movements across observers for the intact movie as an estimate for the point of interest $S(t)$. The variance of the median time course $S(t)$ across time provided an estimate for the variance v_S, which was equivalent to computing the variance across clips for each t under the assumption that variance was stationary with respect to time. We verified in our data that our assumptions for the derivation were reasonable, i.e., that $S^d(t)$ (across clips) and $S(t)$ (across time) were well approximated as Gaussian and that $\mu_{Sd} \approx \mu_S$ for all t (i.e., the mean eye position across all clips was the same for the unscrambled and median intact time courses and near the center of the screen). Furthermore, simulations of $G_0(t)$, computed with randomly permuted values of $S(t)$ as $S_0(t)$, yielded values close to $v_{Sd}(t) + v_S$, as predicted by the derivation.

To isolate the component of the eye position explained by the point of interest, we computed "fractional explained variance" as $1 - G(t)/G_0(t)$, where $G(t)$ was the measured position error (see Eye position error, variance in eye position, and fractional explained variance section) and $G_0(t)$ was the maximal position error (computed using Equation A16). Here, we show that this quantity can be thought of as an approximate empirical estimate for the fraction of the time the observer was not locked on to a random point of interest (or locked on to the actual point of interest) as a function of time. Suppose at each time point t, the observer has a probability of $P_E(t)$ locking on to the point of interest. When the observer is locked on, $\mathrm{E}[(S^d(t) - S(t))^2] \approx 0$. When the observer is not locked on ($1 - P_E(t)$ of the time), $S(t)$ will be random with respect to $S^d(t)$, so $\mathrm{E}[(S^d(t) - S(t))^2] \approx \mathrm{E}[(S^d(t) - S_0(t))^2]$. Therefore, at any time point t,

$$G(t) = P_E(t)\mathrm{E}\left[(S^d(t) - S(t))^2\right]\big|_{\text{locked on}}$$

$$+ (1 - P_E(t))\mathrm{E}[(S^d(t) - S(t))^2]\big|_{\text{not locked on}}$$

$$\approx (1 - P_E(t))\mathrm{E}[(S^d(t) - S_0(t))^2]. \tag{A17}$$

Recall that the maximal eye position error $G_0(t) = \mathrm{E}[(S^d(t) - S_0(t))^2]$. Therefore, $1 - G(\mathrm{t})/G_0(t) \approx 1 - (1 - P_E(t)) = P_E(t)$. Consequently, the quantity $1 - G(t)/G_0(t)$ approximates $P_E(t)$ and corresponds to the probability (across clips) at time t after a cut that the unscrambled time course was locked on to point of interest (median intact eye position across observers).

Simulating fractional explained variance

The model was developed to explain the covariance measurements as a function of scramble duration, but we used it also to simulate eye position error and the corresponding fractional explained variance. For each observer, we simulated the eye position error, $G(t)$, for t up to 5 s, by generating artificial epochs of an unscrambled eye movement time course, $S^d(t)$, and comparing these epochs of $S^d(t)$ to the corresponding portions of the median eye movement time course, $S(t)$. All samples of simulated $S^d(t)$ were drawn from a measured eye movement time course for the intact movie (out of two repeats per each observer). To simulate the fact that each epoch of $S^d(t)$ contained samples that were uncorrelated with $S(t)$ right after a cut, we determined a random time Δ in each epoch after which the observer was presumed to lock on. Specifically, for $\Delta < t \leq 5$, samples of $S(t)$ corresponded to the same segment of the movie as those in the median time course $S(t)$, such that the covariance between $S^d(t)$ and $S(t)$ was maximal (as determined by that observer). For $t \leq \Delta$, samples of $S^d(t)$ were set to those from a random portion of the intact time course, such that $S^d(t)$ still contained actual positions on the screen but unrelated to $S(t)$. The value of Δ was determined by the model. Specifically, it was drawn according to the distribution of a random variable that described the time during which an observer was not locked on to the point of interest for a clip (τ in Equation A6). This random variable was determined using the fit parameter λ and lognormal parameters of saccade latencies for that observer (Equation A3), subject to the constraint $\Delta \leq 5$ s ($d = 5$ in Equation A6). The maximal position error $G_0(t)$ was computed using Equation A16, in which $v_{Sd}(t)$ was the variance of the simulated $S^d(t)$ (or the variance of the intact time courses), which was constant over time. We then computed the fractional explained variance $1 - G(t)/G_0(t)$. The simulation was performed independently 1000 times for each observer, and the value $1 - G(t)/G_0(t)$, averaged across simulations, yielded the model's prediction for the fractional explained variance (Figure 6C, inset).

Acknowledgments

We are grateful to M. Landy for helpful discussions. This work was supported by NIH Grant R21-DA024423 (D.J.H.) and NSF Graduate Student Fellowship (J.F.).

Commercial relationships: none.

Corresponding author: Helena X. Wang.
Email: helena.wang@nyu.edu.
Address: Center for Neural Science, New York University, 4 Washington Place, Room 809, New York, NY 10003, USA.

References

Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research, 39,* 2947–2953.

Angelone, B. L., Levin, D. T., & Simons, D. J. (2003). The relationship between change detection and recognition of centrally attended objects in motion pictures. *Perception, 32,* 947–962.

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition, 17,* 1185–1204.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16,* 341–355.

Bordwell, D., & Thompson, K. (2001). *Film art: An introduction* (6th ed.). New York: McGraw-Hill.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Butko, N. J., Zhang, L., Cottrell, G. W., & Movellan, J. R. (2008). Visual saliency model for robot cameras. *Proceedings of the 2008 International Conference on Robotics and Automation (ICRA),* 2398–2403.

Carmi, R., & Itti, L. (2006a). The role of memory in guiding attention during natural vision. *Journal of Vision, 6*(9):4, 898–914, http://www.journalofvision.org/content/6/9/4, doi:10.1167/6.9.4. [PubMed] [Article]

Carmi, R., & Itti, L. (2006b). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46,* 4333–4345.

Carroll, J. M., & Bever, T. G. (1976). Segmentation in cinema perception. *Science, 191,* 1053–1055.

Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the United States of America, 101,* 13086–13090.

Dmytryk, E. (1986). *On filmmaking*. London: Focal Press.

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, 1–17, http://www.journalofvision.org/content/10/10/28, doi:10.1167/10.10.28. [PubMed] [Article]

d'Ydewalle, G., Desmet, G., & Van Rensbergen, J. (1998). Film perception: The processing of film cuts. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 357–367). Oxford, UK: Elsevier.

d'Ydewalle, G., & Vanderbeeken, M. (1990). Perceptual and cognitive processing of editing rules in film. In R. Groner, G. d'Ydewalle, & R. Parhani (Eds.), *From eye to mind: Information acquisition in perception, search, and reading* (pp. 129–139). Amsterdam, The Netherlands: Elsevier.

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17,* 973–980.

Einhauser, W., & Konig, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience, 17,* 1089–1097.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1–26, http://www.journalofvision.org/content/8/14/18, doi:10.1167/8.14.18. [PubMed] [Article]

Fenton, L. F. (1960). The sum of log-normal probability distributions in scatter transmission systems. *IRE Transactions on Communications Systems, 8,* 57–67.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5,* 490–495.

Frith, U., & Robson, J. E. (1975). Perceiving the language of films. *Perception, 4,* 97–103.

Garsoffky, B., Huff, M., & Schwan, S. (2007). Changing viewpoints during dynamic events. *Perception, 36,* 366–374.

Garsoffky, B., Schwan, S., & Hesse, F. W. (2002). Viewpoint dependency in the recognition of dynamic scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1035–1050.

Germeys, F., & d'Ydewalle, G. (2007). The psychology of film: Perceiving beyond the cut. *Psychological Research, 71,* 458–466.

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine, 37,* 957–964.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. A. Akins (Ed.), *Perception, Vancouver studies in cognitive science* (vol. 5, pp. 89–110). New York: Oxford University Press.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. (2008). Neurocinematics: The neuroscience of film. *Projections, 2,* 1–26.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences, 14,* 40–48.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science, 303,* 1634–1640.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience, 28,* 2539–2550.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9,* 188–194.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Image salience versus cognitive control of eye movements in real-world scenes: Evidence from visual search. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movement research: Insights into mind and brain* (pp. 537–562). Oxford, UK: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–271.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics, 65,* 58–71.

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,* 210–228.

Hirose, Y., Kennedy, A., & Tatler, B. W. (2010). Perception and memory across viewpoint changes in moving images. *Journal of Vision, 10*(4):2, 1–19, http://www.journalofvision.org/content/10/4/2, doi:10.1167/10.4.2. [PubMed] [Article]

Hochberg, J., & Brooks, V. (1978). Film cutting and visual momentum. In J. W. Senders, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and the higher psychological functions* (pp. 293–313). Hillsdale, NJ: Erlbaum.

Hollingworth, A., & Henderson, J. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 113–136.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics, 64,* 882–895.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12,* 1093–1123.

Itti, L., & Baldi, P. (2005). A principled approach to detecting surprising events in video. *Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1,* 631–637.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49,* 1295–1306.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Kanan, C., Tong, M., Zhang, L., & Cottrell, G. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition, 17,* 979–1003.

Kienzle, W., Schölkopf, B., Wichmann, F. A., & Franz, M. O. (2007). How to find interesting locations in video: A spatiotemporal interest point detector learned from human eye movements. *Proceedings of the 29th DAGM Conference on Pattern Recognition, LNCS 4713,* 405–414.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Kraft, R. N. (1986). The role of cutting in the evaluation and retention of film. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 155–163.

Kraft, R. N. (1987). Rules and strategies of visual narratives. *Perceptual and Motor Skills, 64,* 3–14.

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision, 13,* 201–214.

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience, 26,* 51–62.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41,* 3559–3565.

Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research, 47,* 2483–2498.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4,* 501–506.

Levin, D. T., & Simons, D. J. (2000). Perceiving stability in a changing world: Combining shots and integrating views in motion pictures and the real world. *Media Psychology, 2,* 357–380.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

MacLaclan, J., & Logan, M. (1993). Camera shot length in TV commercials and their memorability and persuasiveness. *Journal of Advertising Research, 33,* 57–61.

Mannan, S., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9,* 363–386.

May, J., Dean, M., & Barnard, P. (2003). Using film cutting techniques in interface design. *Human-Computer Interaction, 18,* 325–372.

McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 563–581.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision, 6*(1):2, 8–17, http://www.journalofvision.org/content/6/1/2, doi:10.1167/6.1.2. [PubMed] [Article]

Melcher, D., & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research, 41,* 3597–3611.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434,* 387–391.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45,* 205–231.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46,* 614–621.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11,* 929–942.

Oliva, A., Torralba, A., Castelhano, M., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *Proceedings of the 2003 International Conference on Image Processing (ICIP), 1,* I-253–I-256.

Orban, G. A. (2008). Higher order visual processing in macaque extrastriate cortex. *Physiological Reviews, 88,* 59–89.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46,* 461–488.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107–123.

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16,* 125–154.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Peters, R. J., & Itti, L. (2007). Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 1–8.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45,* 2397–2416.

Rajashekar, U., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2007). Foveated analysis of image features at fixations. *Vision Research, 47,* 3160–3172.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network, 10,* 341–350.

Reisz, K., & Millar, G. (1953). *Technique of film editing*. London: Focal Press.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*(14):16, 1–20, http://www.journalofvision.org/content/7/14/16, doi:10.1167/7.14.16. [PubMed] [Article]

Salt, B. (1992). *Film style and technology: History and analysis* (2nd ed.). London: Starword.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46.

Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory & Cognition, 28,* 214–223.

Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human–monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Current Biology, 20,* 649–656.

Smith, T. J., & Henderson, J. M. (2008). The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research, 2,* 1–17.

Sun, Y., Fisher, R., Wang, F., & Gomes, H. M. (2008). A computer vision model for visual-object-based attention and eye movements. *Computer Vision and Image Understanding, 112,* 126–142.

Tatler, B., Gilchrist, I., & Land, M. (2005). Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology, 58A,* 931–960.

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, 1–17, http://www.journalofvision.org/content/7/14/4, doi:10.1167/7.14.4. [PubMed] [Article]

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45,* 643–659.

Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception, 32,* 579–592.

’t Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., et al. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition, 17,* 1132–1158.

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,* 766–786.

Tosi, V., Mecacci, L., & Pasquali, E. (1997). Scanning eye movements made when viewing film: Preliminary observations. *International Journal of Neuroscience, 92,* 47–52.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology, 12,* 97–136.

Tseng, P. H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision, 9*(7):4, 1–16, http://www.journalofvision.org/content/9/7/4, doi:10.1167/9.7.4. [PubMed] [Article]

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research, 43,* 333–346.

Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. *Spatial Vision, 22,* 397–408.

Wischnewski, M., Belardinelli, A., Schneider, W. X., & Steil, J. J. (2010). Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. *Cognitive Computation, 2,* 326–343.

Yamada, K., & Cottrell, G. W. (1995). A model of scan paths applied to face recognition. *Proceedings of the 17th Annual Conference of the Cognitive Science Society,* 55–60.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621.

Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.

Zhang, L., Tong, M. H., & Cottrell, G. W. (2009). SUNDAy: Saliency using natural statistics for dynamic analysis of scenes. *Proceedings of the 31st Annual Conference of the Cognitive Science Society,* 2944–2949.