Journal of Experimental Psychology: General

Predictive Looking and Predictive Looking Errors in Everyday Activities

Xing Su, Matthew A. Bezdek, Tan T. Nguyen, Christopher S. Hall, Jeffrey M. Zacks

Funded by:

- Office of Naval Research

Abstract

People spontaneously segment continuous streams of experience into events. Some models of event comprehension propose that segmentation is triggered by increases in prediction error, but evidence directly measuring prediction errors during ongoing comprehension is lacking. Here, we measured prediction continuously using eye-tracking. College-aged participants watched movies of everyday activity while their eyes were tracked, and gaze location was used to predict the future location of the actor's hands. Gaze location predicted hand location up to nine seconds into the future. Gaze-based prediction error (lower prediction accuracy) was compared with outputs from a computational model of event comprehension; gaze-based prediction error correlated better with model error than model uncertainty. Gaze-based prediction error was correlated with behavioral event segmentation, as predicted by the theoretical models. These results suggest that predictive looking gives a valid assay of online prediction error, and that prediction error is associated with event segmentation.

Keywords: Event Segmentation Theory; Prediction Error; Event Perception; Eye-tracking

Public Significance Statement

When comprehending everyday events such as checking out at the store, people anticipate what will happen next—for example, predicting that after the checker scans the last item the terminal will prompt for payment. Such predictions can improve comprehension and action, for example by guiding the eyes to the payment terminal. By combining eye tracking with computational modeling, this research found that when predictions lead to errors, this can signal that one event has ended and a new event has begun. These connections between prediction, attention, and the structure of events in experience can inform domains including human-robot interaction, education, and entertainment.

Introduction

To guide behavior in complex, naturalistic activity it is important to anticipate how the situation will evolve—to predict the near future (Clark, 2013). It is also important to abstract from the complex surface dynamics of activity to represent the deeper regularities in events (Richmond & Zacks, 2017). Studying these processes in humans without interrupting them has been a major challenge to research on predictive event comprehension mechanisms. Here, we present a method for using the eyes as a window to the brain's prediction engines. We show, first, that people look predictively during ongoing activity comprehension and, second, that errors in predictive looking are associated with the segmentation of continuous activity into meaningful events.

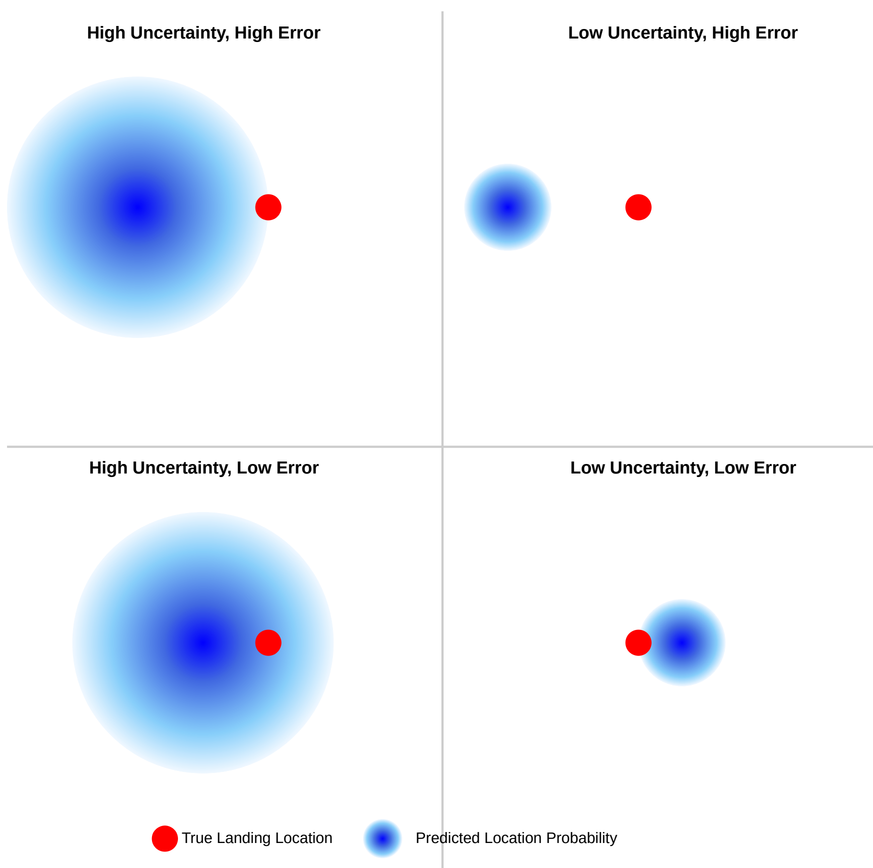
The Role of Prediction and Prediction Quality in Cognition

Prediction is crucial for numerous psychological processes. In scene perception, predictions of where goal-relevant objects are likely to be found determine where people look and attend (e.g., Diaz et al., 2013; Henderson, 2017; Itti & Koch, 2001; Malcolm, 2010; Mann et al., 2013; Rothkopf et al., 2016). In learning, prediction errors drive updating (e.g., Glimcher, 2011; Maia, 2009). These empirical findings align with the predictive brain hypothesis that the brain is not primarily reactive to external stimuli, but rather is predominantly a proactive system that formulates hypotheses, anticipates the consequences of the agents' actions, and constructs expectations (Bubic, 2010; Friston et al., 2006; Maldonato & Dell'Orco, 2012).

As the environment changes, the quality of predictions can fluctuate. Two psychologically relevant indexes exist to measure this: prediction error and prediction uncertainty. Prediction error is the difference between a model's predicted outcome and the actual observed outcome. For example, if a system is predicting where a ball will land, prediction error can be operationalized as the distance between the predicted and actual locations. It has been shown to be important for providing feedback, guiding action control, decision-making, and learning, with robust neurological evidence (for a review see Corlett et al., 2022). Prediction uncertainty is the spread of predictions across potential outcomes. For example, a system might predict a ball's landing location precisely (low uncertainty) or predict a larger region in which the ball is likely to land (high uncertainty). Uncertainty has been found to be important for cognitive control (Mushtaq et al., 2011) and proposed as an important signal for updating event models (Baldwin & Kosie, 2021; Kuperberg, 2021). Figure 1 illustrates the difference between prediction errors and prediction uncertainty.

Figure 1

Prediction Error V.S Prediction Uncertainty



Note. Illustration of the relationship between prediction uncertainty and prediction error. Panels show hypothetical situations in which a system predicts where a ball will land (blue) and then observes the actual landing (red), illustrating four possible combinations of prediction uncertainty and prediction error.

Event segmentation

To predict how everyday activities unfold, people build mental models of current situations—called *event models*—which they use to simulate what might happen next. These event models capture the underlying structure of activities, abstracting away from surface-level complexities. As activities progress, these models need timely updates to maintain accurate predictions. This updating process results in the segmentation of continuous activities into meaningful units. This cognitive process is called event segmentation.

Event segmentation is important for comprehension, memory, and action planning (for a review, see Zacks, 2020). In controlled laboratory settings, with little training, participants can reliability and consistently identify event boundaries while observing the activity (Zacks et al., 2001; Newtson, 1980 and Speer et al., 2003). Furthermore, when asked to identify events at different grains, people's segmentation pattern revealed a hierarchical structure of events where shorter, fine-grained events are nested within longer, coarse-grained events (Zacks et al., 2001). Neuroimaging studies showed that during movie viewing, cortical patterns shift systematically and align with participants' marked event boundaries, suggesting that event boundaries are a normal aspect of ongoing perception (Baldassano et al., 2017; Geerligs et al., 2021).

Theories of event segmentation have proposed different control mechanisms for event segmentation. According to Event Segmentation Theory (EST; Zacks et al., 2007), comprehension mechanisms maintain *working event models*, which are representations of a current event that integrate current sensory and perceptual information with information from episodic memory and from semantic knowledge. Comprehension uses the current working event model to continuously generate predictions about what will occur next. When prediction errors increase, the system abandons the current working event model and constructs a new one; this is associated with the subjective experience of an *event boundary*. In contrast, Shin & DuBrow's (2021) *inference-based event segmentation* account proposes that event boundaries are triggered by increases in prediction uncertainty rather than prediction error. Similarly, Kuperberg's (2021) *hierarchical generative framework* model proposes roles for both prediction error and prediction uncertainty. Both prediction error and prediction uncertainty represent lapses in prediction quality, and they are likely correlated during natural comprehension.

Computational and Behavior Evidence for Prediction Error Driven Segmentation

The hypothesis that lapses in prediction quality cause segmentation is supported by both behavioral studies and computational simulations. For example, by pausing a video and directly asking participants for their predictions, Zacks et al. (2011) found that people's predictions of what would happen five seconds later were more prone to error around event boundaries compared to the middle of events. One limitation of this method is that pausing the movie intermittently inevitably disrupts people's ongoing perception. Additionally, this paradigm allows for only intermittent measurement of prediction errors.

Computational models of human cognitive processes offer a means to tease apart correlated prediction error and uncertainty signals. The Structured Event Memory (SEM) model (Franklin et al., 2020; Nguyen et al., 2024) uses a hybrid architecture to mimic human event comprehension process. SEM represents working event models using recurrent neural networks and models event-to-event transitions using approximate Bayesian inference. On each timepoint during training and testing, the model is presented with an input “scene vector” and uses the currently activated neural network to predict the state of that scene vector on the next time step. In the simulations reported by Nguyen et al.2024, the scene vectors were abstracted recordings of actors performing everyday activities, which captured the motion of the actor and their interaction with objects. Over training, each network adjusts its weights (using backpropagation), learning the sequential dependencies in the scene vector recordings. To model how a cognitive system might learn about different types of events, SEM uses approximate Bayesian latent state inference to select which neural network to use on each timestep. This process evaluates the posterior likelihood of each neural network given the current sequence of scene vectors and selects the most likely network. It also has the option to initialize a new network if that has a higher likelihood than any of the current networks; thus, SEM builds a library of event type representations over learning. In its generic form, SEM performs inference on every time step. Nguyen et al.2024 also simulated two variants that imposed a gating mechanism on the inference process, to model a control mechanism to gate event model updating. These variants trigger event-to-event transition, or segmentation based on increases in either prediction error (pe-SEM) or prediction uncertainty (uncertainty-SEM). After training, both variants successfully predicted activity progression and formed segments and categories corresponding to human judgments, despite not being explicitly trained for segmentation or categorization. Notably, uncertainty-SEM's performance tracked human segmentation and categorization more closely than pe-SEM.

Gaze Reveals Human’s Predictive Processes

A limitation of the explicit prediction task used by Zacks et al., (2011) is that it interrupts ongoing perceptual processing. Indirect measures such as eye tracking offer the promise to assay prediction quality without disrupting comprehension mechanisms. Humans look predictively to locations that are likely to be important in the near future, and this tendency has been used to study predictive processing in ongoing behavior and in populations who cannot verbally report their predictions. For example, researchers have assessed infants' learning by measuring anticipatory eye movements directed towards locations where stimuli are expected based on learned sequences (e.g., Romberg & Saffran, 2013). In sports, athletes demonstrate predictive behavior by looking ahead to the expected position of a ball rather than simply tracking its current location (e.g., Hayhoe et al., 2012). Within complex scenes, people’s focus of attention is guided by predictions about the most meaningful and task-relevant information present (e.g., Henderson, 2017).

Eye tracking can also measure the dynamics of human prediction during the viewing of naturalistic activities. Eisenberg and colleagues (2018) developed the Predictive Looking at Action Task (PLAT) to investigate the time course of predictability during video watching. Predictive looking was quantified as the duration participants fixated on the object to be touched during successive 500-ms time windows in the three seconds preceding the actual contact. The results revealed that viewers predict and direct their gaze ahead to objects about to be contacted. Moreover, around event boundaries, predictive looking was delayed compared to predictive looking in the middle of an event. This suggests that near an event boundary, predicting future actions becomes more challenging, leading to increased looking times just before contact. One limitation of the PLAT is that prediction error is only measured in the time window preceding an object contact, leading to sparse sampling of prediction error. Moreover, prediction uncertainty confounds with prediction error, as an increase in prediction uncertainty could also delay predictive looking near event boundaries. Consequently, it remains unclear whether individuals engage in continuous predictive looking throughout ongoing activities and whether predictive looking patterns can distinguish error from uncertainty.

The Current Study

In this study, we exploited viewers' predictive looking behavior to investigate the temporal dynamics of predictions during event comprehension. Our first hypothesis was that viewers would look predictively during passive viewing of everyday activities. Our second hypothesis was that predictive looking errors would be more strongly associated with SEM's prediction errors compared to uncertainty; if so, this would provide evidence for the specificity of the prediction error and prediction uncertainty estimates from the model. Although uncertainty is also a measure of prediction quality, it captures a distinct aspect of prediction from error. Third, we hypothesized that these prediction errors would be associated with viewers' identification of event boundaries, as predicted by Event Segmentation Theory (EST).

To test these hypotheses, we tracked participants' eye movements while they viewed movies of actors performing everyday activities. Building upon prior research indicating that individuals look predictively at an actor's hand (Eisenberg et al., 2018), predictive looking was operationalized as gaze directed toward locations where the actor's hands would be in the next few seconds. We compared predictive looking errors with measures of prediction error and uncertainty from SEM, and with human segmentation.

Methods

Eye Tracking

Participants. For this part of the study, we recruited participants from the subject pool at Washington University. In exchange for their time, participants were offered compensation in the form of either course credit or $10 per hour. A target sample size of 100 was decided before data collection to ensure adequate to detect small to middle effect size reported in (Gold et al., 2017).

Based on the power analysis, 70 participants were necessary for power of .80. Because some participants might not have usable eye-tracking data, we decided to recruit up to 100 participants. Data from 19 participants were excluded because of experiment program failure (n=13), eye-tracking calibration failure (n=5) or failure to remain on the headrest throughout the study (n=1).

A total of 81 individuals successfully completed the study. Each participant self-reported their age (M = 19.00 years, SD = 1.88) and gender identity, as well as their race and ethnicity. Among them, 54 were female individuals and 27 were male individuals. The sample's racial breakdown included 23 individuals identifying as Asian, 43 as White, 3 as Black or African American, 10 as more than one race, and 2 who did not report their race. In terms of ethnicity, 7 individuals reported being Hispanic.

Materials. Four movies of actors performing everyday activities were selected from the META corpus (Bezdek, et al., 2022). This stimulus set includes different actors exercising (586 s), making breakfast (586 s), grooming (646 s), and tidying up the room (679 s). The movies were filmed from a fixed, head-height perspective, with no pan or zoom. The frame rate was 60 *f*rames per second, and the frame's dimension was 1280*720 pixels. This movie set can be found from this link (https://osf.io/vsbaq) in the "stimuli" Folder. Figure. 2 shows representative frame from all 4 activities.

Figure 2.

Representative frames from the four movies.

Task and procedure. After giving informed consent to this study, participants passively watched four movies of an actor performing everyday activities. Breaks were offered between movies. Participants were instructed to watch the movies for comprehension while their eyes were recorded by the eye tracker. The eye tracker was calibrated by the experimenter before the start of the experiment and during the break if needed. Researchers monitored participants' attention periodically throughout the study. Any observed instances of inattention were documented, and the corresponding data were excluded from analysis. Gaze locations were obtained from the right eye using an infrared pupil-corneal eye tracker camera (EyeLink 1000; SR Research Ltd., Mississauga, ON, Canada) that sampled at 1000 Hz. The camera was mounted on the SR Research Desktop Mount. A chin/forehead rest was used to minimize head motion during the tasks. The camera was positioned 52 cm from the top of the forehead rest. Stimuli were presented on a 19-in. (74 cm) monitor (1440 × 3900 resolution, viewing distance of 58 cm from the forehead rest, viewing angle of 38.68).

Calibration of the eye tracker was conducted before beginning the study task. Participants were instructed to look at each of five to nine dots presented serially across the participant's central and peripheral visual field. Following calibration, the measurements were validated by having the participants look at each of these nine dots again as they appeared on the screen. This validation of calibration was considered good when there was an average error of 0.50 degrees of visual angle or less and when the maximum error for any given dot was 1.00 degree or less.

Movie order was counterbalanced across participants. Participants were debriefed after completing the study.

Event Segmentation Norms

Participants. To obtain highly reliable segmentation norms for comparison with the eye-tracking data, we combined two datasets. In the first dataset, 184 individuals recruited via Amazon Mechanical Turk self-reported their age (M = 35 years, SD = 11.95), gender identity, race, and ethnicity. Among these individuals, 70 were female, 112 were male, and 2 identified as other. Regarding race, 17 individuals identified as Asian, 121 as White, 23 as Black or African American, 2 as American Indian/Alaskan Native, 8 as more than one race, and 4 did not report their race. In addition, 14 individuals reported Hispanic ethnicity. The recruitment procedure is detailed in Bezdek et al. (2022).

In the second dataset, individuals were recruited from the Volunteer for Health participant registry maintained by Washington University. Data from 47 individuals were included in the analyses (age range: 18–35 years, M = 23 years). Among these individuals, 33 were female and 14 were male. Participants self-reported their race: 14 individuals identified as Asian, 26 as White, 2 as Black or African American, and 5 as more than one race. In addition, 6 individuals reported Hispanic ethnicity.

We combined these two datasets to better estimate the segmentation probability distributions for individual movies. Specifically, having a substantial number of timepoints with zero probabilities is a problem for model estimation. By increasing the sample size, we reduced the number of timepoints with zero segmentation probability, resulting in a more robust and reliable distribution.

Materials. The same four movies from the passive viewing part of the study were used for this task.

Procedure. After giving informed consent, participants were randomly assigned either to identify coarse event boundaries or fine event boundaries in those four movies. In the coarse condition, participants were instructed to push the button whenever they believe that a large meaningful unit of activity has ended. In the fine condition, participants were instructed to push the button whenever they believe that a small meaningful unit of activity has ended.

All participants practiced the segmentation task before segmenting the four movies. During the practice session, participants segmented a video with a duration of 2 min 35 s in which a man constructs a toy boat using interlocking building blocks. Participants repeated the practice until they pressed the button within a pre-defined range: 5-8 times for fine segmentation and 2–4 times for coarse segmentation. Participants were never informed of these ranges but were asked to repeat the practice to identify more (or fewer) activities in the video until performance was within range.

Estimating Predictive Looking

To establish a continuous measure of predictive looking error, this study employed regression models to formalize predictive looking. The residuals of the predictive looking model served as the quantifiable prediction error. Predictive looking was defined as the extent to which participants' prior gaze locations predicted the current hand location of the actor. The hands' location was selected as the region of interest to approximate the most crucial area to predict in each frame, given the frequent hand movements observed across all movies. Attending to hand locations was deemed vital for processing and comprehending the four activities in this study. Moreover, previous research has indicated that individuals naturally focus on hand locations during naturalistic viewing, and distinctive hand positions are associated with event boundaries (Zacks, 2004).

Spatial location was analyzed by dividing the screen into an 8-by-6 grid and modeling the probability that an actor's hand would be in a given grid square based on the recent history of eye gaze. The 1280-by-720-pixel screen image was divided equally into 48 rectangles (160 pixels wide, 120 pixels high). The dependent variable in the predictive looking model, denoted as H in equations 1-6, was a binary variable coding for the presence of a hand in the relevant grid square. To determine the presence of actor's hands for each grid in each frame for each movie, we first identified the hand locations using the OpenPose algorithm, a computer vision model for pose detection (Cao et al., 2019). The hand location was determined by the OpenGaze algorithm as a single point on the screen. Because hands occupy more than one pixel on the screen, we convolved these point-locations with a two-dimensional Gaussian kernel (sd = 50 pixels, determined through visual inspection to ensure that the resulting density map accurately represented the majority of the hand area). Following this convolution, the sum of hand location probabilities for pixels within each grid was computed. An “elbow point” in the distribution of these probabilities across all grids for each frame—representing a significant change in the rate of loss—was then identified and used as a cut-off threshold. Grids surpassing this threshold were assigned a value of 1 (indicating the presence of hands), while grids falling below the threshold were coded as 0.

The independent variables in the predictive looking model were participants' gaze density for each grid in each frame, denoted as F in equations 1-6. To calculate gaze density for each grid in every frame, we convolved each participant's gaze location using a two-dimensional Gaussian kernel. The kernel's standard deviation was calibrated to the participants' viewing angle (sd = 37 pixels). We set the kernel size using a spatial sigma of 2 degrees of visual angle, chosen because this corresponds to the approximate size of the human foveal region.

This convolution process generated the gaze density value for each pixel corresponding to each participant. Subsequently, the density values of all participants for all pixels within a given grid were aggregated, resulting in the final gaze density measurement for each grid.

To enhance the sensitivity of the analysis to dynamic prediction errors, grids containing the actor's face were excluded from analysis. Faces, like hands, carry rich predictive information —in particularly, facial expression and eye gaze can be important sources of predictions. However, those features’ predictive value is not spatially arrayed: For example, if an actor looks at a coffee cup and frowns the viewer might predict that they will avoid reaching to that coffee cup, whereas if the person looks to the coffee cup and smiles the viewer might predict that they will reach to it. However, eye-tracking cannot register whether the viewer is encoding the smile or frown, nor which coffee cup the actor is looking at. Thus, the predictive information in the face is not predictive information that can be assayed with eye-tracking. The location of faces was determined using the OpenPose algorithm (Cao et al., 2019), and a two-dimensional Gaussian kernel (sd=50) was applied to each face location. We identified an 'elbow point' within the distribution of these facial location probabilities across all grids, representing a significant change in the rate of loss. This elbow point served as our cut-off threshold, and any grids with values exceeding this threshold were subsequently removed from the analysis. Importantly, we replicated all of our findings with the face grid included, confirming the robustness of our results (see Supplemental Section 2).

To improve robustness, we applied temporal smoothing to the gaze and hand data before analysis. A Gaussian kernel was used to smooth the probability signals for gaze and hand movements over the duration of each movie, reducing skew in the hand probabilities that could otherwise cause convergence issues in the predictive model. The size of the smoothing kernel (120 frames) was determined through visual inspection. We then sampled these smoothed values at intervals of one second (60 frames).

To form a measure of predictive looking error, we fit mixed effects models to predict the likelihood of a grid containing a hand at a given timepoint from the recent history of gaze locations and then used the residuals from the model as the error measure. Equation 1 describes the form of the fixed effects in the predictive looking model:

$$\text{Equation 1: } H_{tr}={}^{\beta}{}_{i0}F_{t-ir}+{}^{\beta}{}_{i-10}F_{t-(i+1)r}+\ldots+{}^{\beta}{}_{00}F_{tr}+\varepsilon$$

H_{tr} is the log odd of a hand being in region r (pixel) at time point t(s); the F variables represent the density of gaze locations in the region in the i seconds before the current time point.

Models were fit as multivariate mixed-effects logistic models using the lme4 package (Version 1.1.27.1; Bates et al., 2015) in R software (R Core Team, 2014). To account for potential variance stemming from specific movies and the diverse locations of the grid within frames, both grid index and movie were integrated as random intercepts within the models. To evaluate the degree of predictive looking, a stepwise model comparison was employed. Starting with a model that included only gaze on the current time point, fixed effects representing successive previous time points were added until the additional time point did not add predictive value. The stepwise model comparison approach can be conceptualized as follows:

Equation 2: $H_{tr}= {}^{\beta}{}_{00}F_{tr}+\varepsilon$

Equation 3: $H_{tr}= {}^{\beta}{}_{10}F_{t-1r}+{}^{\beta}{}_{00}F_{tr}+\varepsilon$

Equation 4: $H_{tr}={}^{\beta}{}_{20}F_{t-2r}+ {}^{\beta}{}_{10}F_{t-1r}+{}^{\beta}{}_{00}F_{tr}+\varepsilon$

Equation 5: $H_{tr}={}^{\beta}{}_{30}F_{t-3r}+{}^{\beta}{}_{20}F_{t-2r}+{}^{\beta}{}_{10}F_{t-1r}+ {}^{\beta}{}_{00}F_{tr}+\varepsilon$

…

Equation 6: $H_{tr}={}^{\beta}{}_{i0}F_{t-ir}+{}^{\beta}{}_{i-10}F_{t-(i+1)r}+\ldots{}^{\beta}{}_{30}F_{t-3r}+{}^{\beta}{}_{20}F_{t-2r}+{}^{\beta}{}_{10}F_{t-1r} + {}^{\beta}{}_{00}F_{tr}+\varepsilon$

The optimal model was determined based on the relative variation in both Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) values among these models.

Prediction errors were quantified as the magnitude of the deviance residuals for each grid in each frame based on the optimal predictive looking model. This measure differs from prediction uncertainty, as it does not reflect the stability of the optimal model’s prediction of the hand location but rather quantifies the discrepancy between the predicted and ground truth values. Deviance residuals in logistic regression are a measure of the difference between observed and expected outcomes under the model. The mathematical formula for the magnitude of the deviance residual is:

$$\text{Prediction Error} = \sqrt{-2\log\left(Prob_{event}\right)}$$

Predictive Signals of Computational Model of Event Comprehension

To compare predictive looking error with other measures of prediction quality during movie viewing, we extracted two measures from computational models: prediction error from PE-SEM and prediction uncertainty from uncertainty-SEM. Both PE-SEM and uncertainty-SEM are variants of a computational architecture that models human event comprehension. These models represent different event schemas as distinct Recurrent Neural Network (RNN) weight matrices, with the current working event model represented as hidden unit activation within one of the event schema RNNs.

Following theoretical principles of movie processing, at each timestep, the currently active RNN receives a scene vector and predicts the subsequent scene vector. PE-SEM tracks prediction error by calculating the Euclidean distance between the predicted vector and the actual future scene vector. When prediction error exceeds a threshold, the current working event model is replaced with a new one through an approximate Bayesian process. Similarly, uncertainty-SEM tracks prediction uncertainty by calculating the variability of the predicted scene through small perturbations of the weight matrix. When prediction uncertainty exceeds a threshold, a new event model replaces the current one via the same process.

The two variants are identical in all other aspects of their design except for their updating signals. Both models were trained on a single pass through 18 hours of naturalistic human activities, with an additional 3.5 hours used to test each variant's agreement with human segmentation and categorization. Both models were able to learn to predict human activity, and it developed segmentation and categorization approaching human-like performance. For detailed comparisons of PE-SEM and uncertainty-SEM performance against human behavior, see Nguyen et al., 2024. Table 1 provides an overview of each key concept's psychological meaning, along with how it was operationalized in this study.

Table 1

Key Concepts, Psychological Definitions, and Their Operationalizations in This Study

Key Concept	**Psychological Meaning**	**Operationalization**
Predictive Looking Error	The difference between the location where people look in anticipation of the target (predicted location) and the actual location of the target.	The deviance residuals of the optimal predictive gaze model within each grid for each frame.
Model Prediction Error	The difference between the predicted outcome generated by an individual's internal model and the actual observed outcome.	The Euclidean distance between an observed scene vector and the current working event model prediction generated by the SEM's RNN.
Model Prediction Uncertainty	The extent to which predictions based on an internal model are distributed across potential outcomes.	The variability in the SEM's working event model RNN predictions resulting from perturbations to the RNN's weights.
Segmentation Probability	The probability that a given	The density function of a

	timepoint is perceived as the moment when individuals distinguish between two events, defined as segments of time at a particular location with a perceived beginning and end.	group of individuals' segmentation points, scaled by the number of participants.

Transparency and Openness

The anonymized data and analysis code to reproduce our results are publicly accessible in our Open Science Framework folder at https://osf.io/vsbaq. This study was not preregistered. The experimental stimuli of this study is also included in our Open Science Framework folder at https://osf.io/vsbaq. Throughout this work, we report how we determined our sample size and list all data exclusions (if any), all manipulations, and all measures in each study. All data were analyzed using R, Version 4.2.2. Our experiments received Institutional Review Board approval and complied with all ethical regulations.

Results

The data were analyzed to address three main questions 1) What is the extent of participants' predictive looking? To address this, we constructed statistical models that use participants' past gaze locations to predict the actor's current hand location. We then employed a stepwise model comparison to identify the best-fitting predictive looking model, allowing us to determine the extent of participants' predictive looking. 2) What is the relationship between predictive looking error and computational models of prediction error? To explore this question, we examined the correlation between the magnitude of the residual in the most optimal predictive looking model and the prediction error derived from the computational model (pe-SEM). Additionally, we also examined the relationship between the same magnitude of the residual with prediction uncertainty generated by another computational model (uncertainty-SEM). And finally, 3) Is predictive looking error associated with event segmentation? To address this question, we analyzed the relationship between the magnitude of residual in the most optimal predictive looking error with segmentation probability.

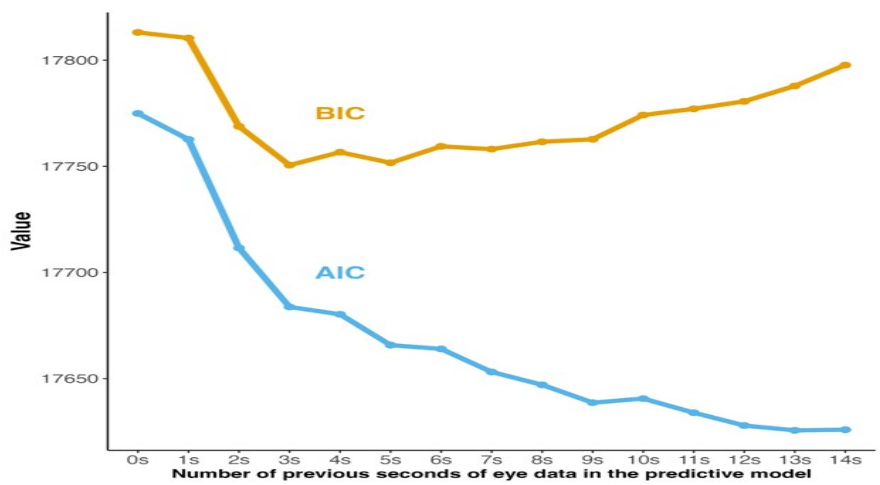
Participants exhibited predictive looking during passive movie viewing

To quantify predictive looking, a forward selection stepwise model comparison approach was implemented. The analysis began with the simplest model (Equation 2), where the current gaze density predicts the probability of the hand being in a particular grid square. Then, gaze density during the previous second was added, followed by gaze density two seconds before, etc. All models included random intercepts of grid locations and movie. To assess model fit, both AIC and BIC index were calculated for each model. The final model selected included gaze density up to 9 seconds ago (equation 6). The comparative AIC and BIC values across models are shown in Figure 2. To select the optimal predictive looking models, we evaluated both BIC and AIC parameters. While both measure model fitness, BIC penalizes model complexity more heavily. For our analysis, which examines how past gaze predicts current hand locations while controlling for previous gaze location, models incorporating past gaze patterns necessarily include more variables and would be overly penalized by BIC. Therefore, we consulted both BIC and AIC for optimal model selection, as indicated in Figure 3. The AIC values continued to decrease until the model included gaze positions up to 9 seconds ago, while BIC values remained stable, suggesting a genuine improvement in model fit rather than overfitting due to additional parameters. Beyond 9 seconds, BIC began to increase, and the rate of AIC decrease slowed, indicating that including gaze patterns beyond 9 seconds likely led to overfitting.

One possibility is that the predictive value of the previous eye position simply reflects that people look to where the hands are now and that is correlated with where the hands will be in the future. To address this possibility, we compared two models: one that included only previous hand positions, and another that incorporated both previous hand and gaze positions. Across time points, the model including gaze positions showed significantly better fit compared to the model with hand positions alone. This demonstrates that previous gaze positions contribute unique predictive information beyond the location of the previous hand position. For detailed model specifications and results, refer to Section S1 of the Supplementary Material.

Figure 3

Model comparisons results



Notes: Model comparisons demonstrate that gaze positions can predict an actor's hand locations up to nine seconds into the future. Model fit, as indicated by AIC, significantly improves when including gaze data up to nine seconds prior. BIC suggests that improvements in fit stabilize with gaze data five seconds prior and begin to decline when including data from ten seconds before.

Predictive looking error correlates with computational model-based prediction error

We operationalized prediction error as the magnitude of the deviance residuals, as it represents the discrepancies between prediction and the reality. We refer to this variable as *predictive looking error* because it was derived from participants' gaze patterns.

To explore the relationship between predictive looking errors and prediction error generated by the computational model pe-SEM, we conducted a mixed-effect linear regression. In this regression, predictive looking error served as the dependent variable, while pe-SEM's prediction error acted as the independent variable. This model included random intercepts for the type of movies and the grid number. The results revealed a significant positive relationship between prediction error and predictive looking errors (prediction error model: $\beta = 0.036$, $t(105,359) = 13.62$, $p < .05$). Analysis of the marginal R^2 value revealed that SEM's prediction error accounted for 0.1% of the variance in predictive looking errors, while controlling for the hierarchical random effects of movie and grid structure.

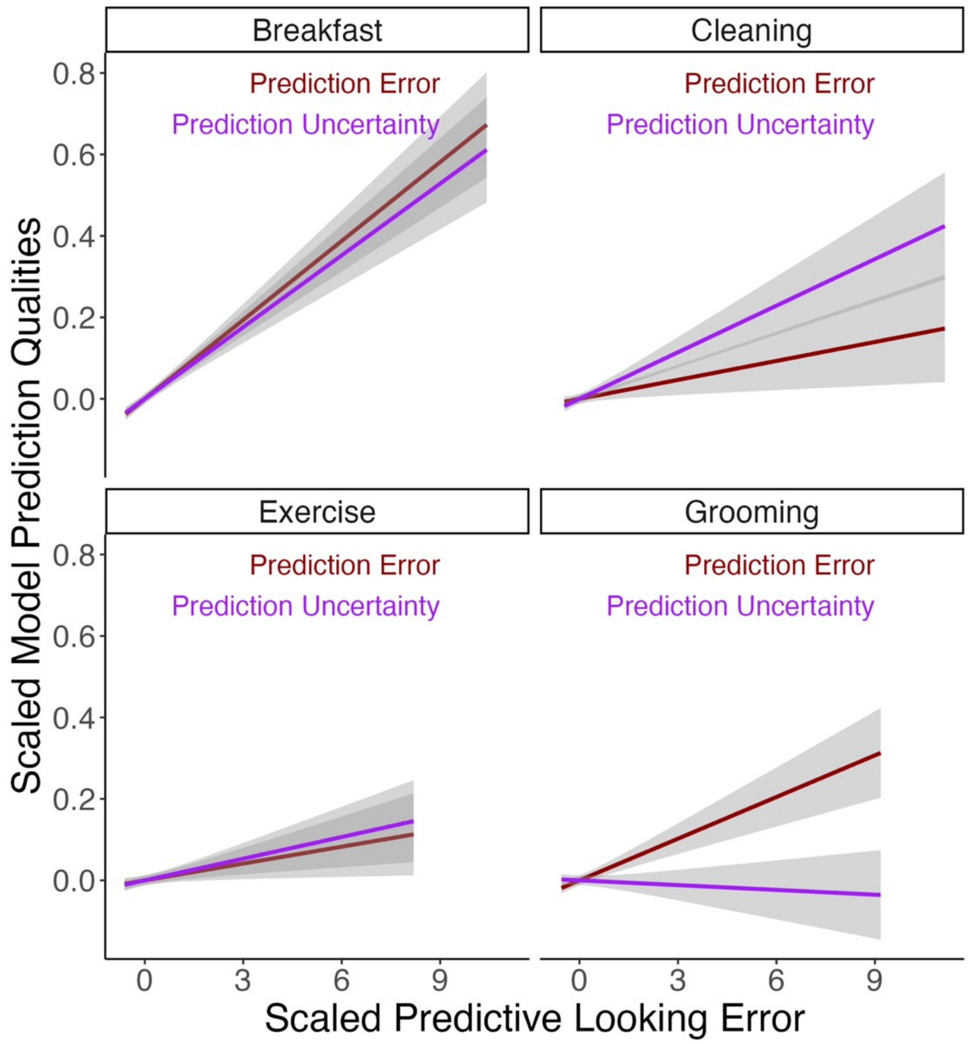
In addition, we explored the relationship between predictive looking error and prediction uncertainty to assess the specificity of this measure of prediction error. A similar mixed-effects linear regression was constructed with predictive looking error as the dependent variable and the prediction uncertainty of uncertainty-SEM as the independent variable, controlling for the random effects of movie types and grid numbers. The results indicated a significant positive correlation between prediction uncertainty and predictive looking errors (prediction uncertainty model: $\beta = 0.028$, $t(105,359) = 10.53$, $p < .05$). An analysis of the marginal R^2 value revealed that SEM's prediction uncertainty accounted for 0.08% of the variance in predictive looking errors while controlling for the hierarchical random effects of movie and grid structure.

By comparing two mixed-effects linear regression models—one in which PE-SEM's prediction error predicts predictive looking errors and another in which Uncertainty-SEM's prediction uncertainty predicts predictive looking errors—the AIC and BIC values suggest that the prediction error model provides a better fit for predictive looking errors than the uncertainty model (prediction error model: AIC = 267,363.9, BIC = 267,438.4; prediction uncertainty model: AIC = 267,438.4, BIC = 267,486.2). According to Akaike weights, the model incorporating PE-SEM's prediction error is significantly superior to the one involving Uncertainty-SEM's prediction uncertainty (prediction error model Akaike weight = 1.00; prediction uncertainty model Akaike weight = 0.00).

Figure 4 illustrates the relationship between predictive looking errors and two key measures of prediction quality from two variants of SEM computational models: prediction error (PE-SEM) and prediction uncertainty (Uncertainty-SEM). Through simple linear regression analysis, the data reveals that both prediction error and uncertainty show positive correlations with predictive looking errors. When analyzing the relationships across different movie types, an interesting pattern emerges for "Breakfast" and "Grooming" movies, the relationship between predictive looking error and model prediction error shows a steeper slope compared to the relationship with prediction uncertainty. However, this pattern reverses for "Exercise" and "Grooming" movie types. Overall, when controlling for movie-specific random effects, the analysis indicated that SEM's prediction error serves as a more reliable predictor of human predictive looking errors compared to prediction uncertainty. This finding suggests that the computational model's prediction error metric more closely aligns with actual human predictive gaze behavior.

Figure 4

Relationship Between Model Prediction Qualities and Predictive Looking Errors



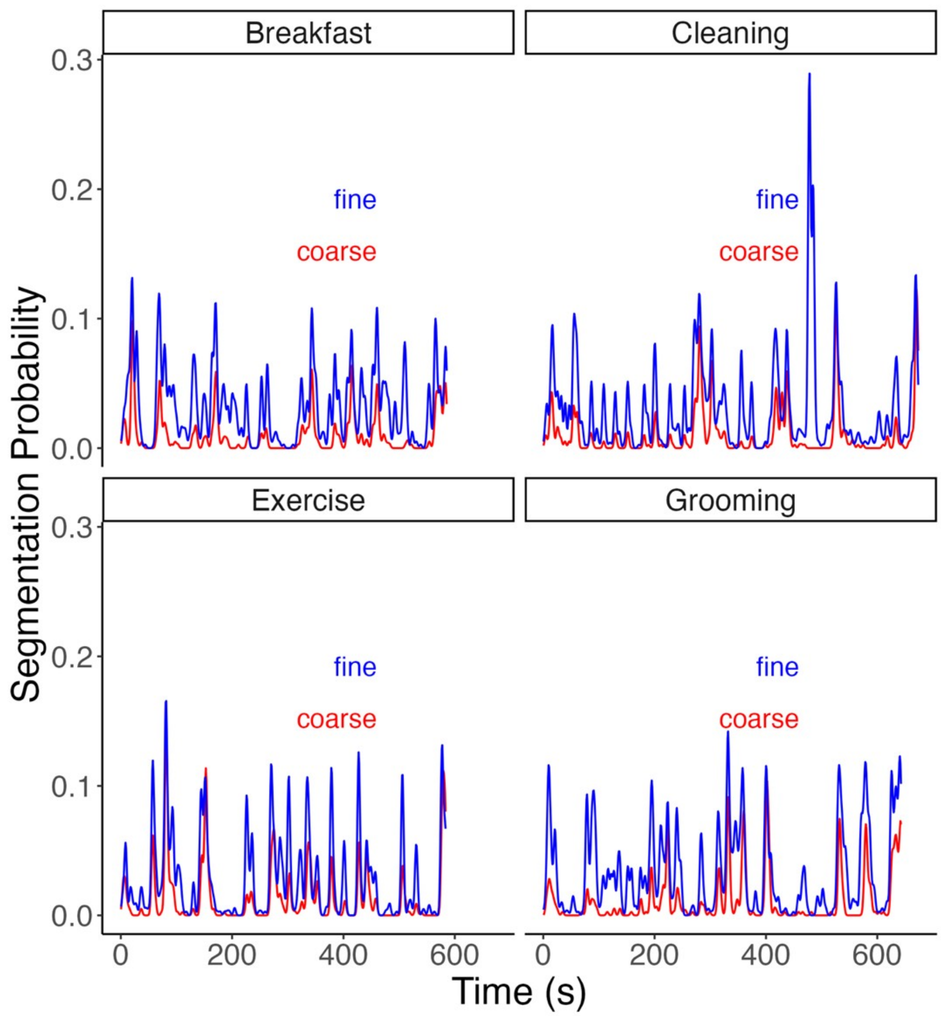
Notes. Linear regression lines showing the relationship between two predictive quality measurements from computational models of event comprehension—prediction error and prediction uncertainty—and predictive looking error across four movies. Shaded regions indicate the 95% confidence intervals of the estimates.

Predictive looking error is associated with segmentation probability

To characterize participants’ segmentation behavior, we estimated the probability density of event segmentation over time for each movie, for both fine and coarse segmentation (Figure 5). Segmentation density was estimated by convolving the time course of individual button presses with a Gaussian kernel using a 2-second bandwidth. This bandwidth—equivalent to 120 frames at a 60 frames per second rate—was selected to match the smooth transitions observed in the hand and gaze probabilities. Segmentation probability is derived by multiplying the segmentation density by the average number of button presses. As can be seen in Figure 5, some timepoints within the movies were consistently identified by many participants as event boundaries, while other time points were never selected as event boundaries. This high level of agreement replicates previous results (e.g., Sasmita & Swallow, 2022; Zacks & Tversky, 2001).

Figure 5

Probability of fine and coarse segmentation for each of the four movies

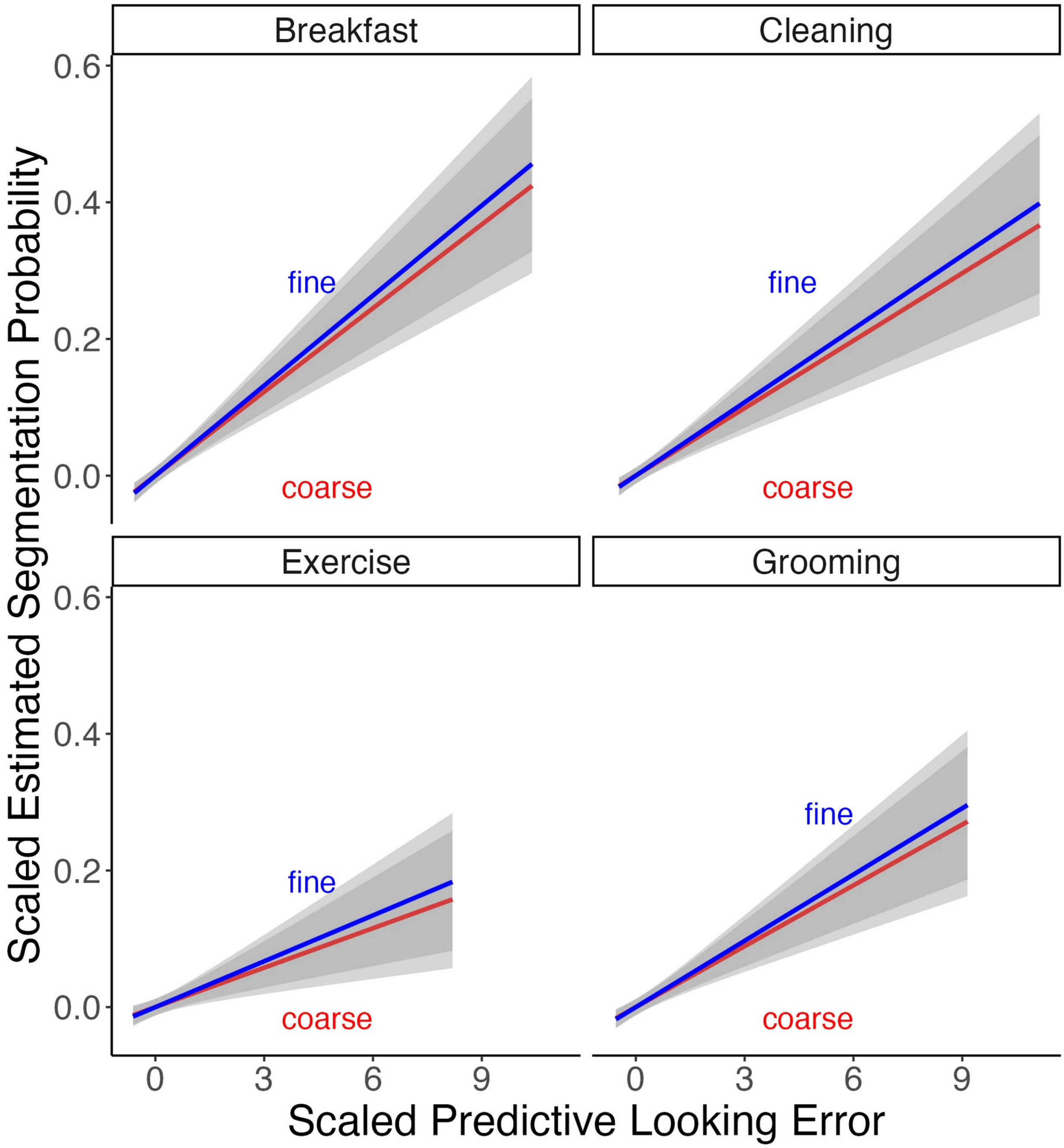


We used a mixed effects multivariate regression model to assess the relationship between predictive looking error and segmentation probability. The independent variable of Grain of segmentation was entered as a covariate, and the movies and grid number were entered as the random effects. There was a significant positive relationship between segmentation probabilities and predictive looking errors ($\beta = 0.001$, $t(6{,}635) = 6.63$, $p < .05$). An analysis of the marginal R^2 value revealed that predictive looking errors accounted for 11% of the variance in segmentation probabilities while controlling for the covariate of Grain in the hierarchical random effects of movie and grid structure.

As expected, segmentation probabilities were significantly higher in the fine segmentation condition. There was no significant interaction between segmentation grain and predictive looking error. Figure 6 indicates a consistent positive relationship between predictive looking errors and event segmentation probabilities across all four movies. When viewers make larger errors in predicting upcoming actions, they are more likely to perceive event boundaries. This supports the proposal that increased prediction errors signal transitions between discrete events.

Figure 6

Relationship between Predictive Looking Error and Segmentation Probability



Notes: Linear regression lines are fitted between estimated looking error and segmentation probability for all four movies where color indicates the segmentation condition. Shaded region represents the 95% confident interval of the estimates.

Discussion

The present study uncovered three critical phenomena related to event perception: First, individuals exhibit predictive looking up to 9 seconds into the future while passively observing everyday activities. Thus, gaze position can be used to track prediction error. Second, there was a significant and positive correlation between predictive looking errors and computationally generated prediction errors, indicating that both measures encapsulate the dynamics of prediction error in the brain. Finally, predictive looking errors increase around event boundaries.

People looked predictively up to 9 seconds in the future during the unfolding of everyday activities

Model comparison approach revealed that viewers' gaze predicted the locations of an actor's hands up to nine seconds in the future. For example, in the breakfast movie, participants started directing their gaze towards the toaster once the actor picked up the plates from the counter, anticipating the subsequent movement of the actor's hand in the upcoming seconds. This finding remained robust even when facial regions were included in the analysis grid and across varying smoothing kernel parameters (see Supplemental Material section 2). To distinguish between predictive gaze and simple hand-following behavior, we compared two models: one using only previous hand locations to predict current hand position, and another incorporating both previous hand and gaze locations. The model including gaze showed significant improvement, demonstrating genuine predictive looking rather than mere autocorrelation between hand movements and gaze following the hands. Detailed results of these additional analyses are available in the supplemental materials section 1.

Prediction manifests in various forms of eye movements. For example, anticipatory pursuit can begin as early as 200 milliseconds before the onset of object motion when the motion direction and onset are predictable (e.g., Badler & Heinen, 2006; Barnes & Collins, 2008). Saccades, too, synchronize with target steps rather than exhibit a time lag if the target is predictable (Smit & Van Gisbergen, 1989). Beyond tracking a single object or target in highly controlled stimuli, individuals demonstrate the ability to anticipate future events in naturalistic activity. In sports, professional cricket players fixate on the ball as it leaves the bowler's hand and then rapidly make a saccade to where the ball is expected to bounce, awaiting its image to return to the fovea (Land & McLeod, 2000). In daily activities outside of sports, Eisenberg et al., (2018) showed participants increased their gaze time on a target object as the actor approached, suggesting an anticipation of the actor making contact with the object in the near future. Building on previous research, the current data-driven approach unveils that predictive looking extends beyond critical objects or specific time intervals; rather, it is a continuous and integral aspect of natural perception.

Predictive looking errors aligned more with model prediction error compared to model uncertainty

Although prediction errors and prediction uncertainty are often correlated in naturalistic activities, they represent distinct signals of model updating. Prediction error measures the discrepancy between a model's predicted output and the observed ground truth, quantifying how well the model's predictions align with the actual data. Prediction uncertainty, on the other hand, reflects a lack of confidence or knowledge about a prediction, indicating the model's level of uncertainty regarding its own prediction. This study presented a way to quantify prediction error during passive viewing of everyday activities. The validity of this measure was confirmed by comparing it with model prediction error and model prediction uncertainty. Statistical models incorporating either the updating signal of SEM-uncertain (uncertainty) or SEM-PE (prediction error) to predict predictive looking errors demonstrated that computationally generated prediction error provides a significantly better fit for predictive looking compared to a model incorporating computationally generated prediction uncertainty. This suggests that the current measure of prediction errors accurately captures the true prediction errors occurring in the brain.

When controlling for the random effects of movement types and grid positions, model prediction error explained 11% of the variance in predictive looking error. Although this effect size may appear small, two important factors should be considered: First, both measures contain inherent noise in their measurement. Second, the model accounts for multiple sources of variance, with visual salience being a particularly strong driver of eye movements. The fact that we observed a significant positive relationship between these measures despite these factors suggests both measures capture genuine variation in prediction error within the human brain.

Predictive looking errors increase around event boundaries

Previous studies have shown that event boundaries are related to various characteristics of eye movements. Eisenberg et al., (2018) documented that predictive looking is impaired at event boundaries. Moreover, viewers initiate an exploratory processing phase around event boundaries, before transitioning to focal viewing as the event progresses. This has been interpreted as a response to the unpredictability of activity around event boundaries (Eisenberg & Zacks, 2016).

One limitation of previous research on prediction errors in event comprehension was that prediction error was measured sparsely (Eisenberg et al., 2018) or in ways that interrupted ongoing perceptual processing (Zacks et al., 2009, 2011). To overcome these limitations, here we introduced a direct and continuous method for estimating predictive looking errors during movie viewing. Having derived and validated this measure of prediction error, we then showed that there was a significant and positive relationship between predictive looking errors and segmentation probabilities. This finding provides evidence for theories of event cognition that suggest prediction error drives event segmentation (We note, however, that in a direct comparison of SEM's prediction errors and prediction uncertainty with segmentation, it was uncertainty that was more strongly related to human segmentation; Nguyen et al., 2024)

Implications for theories of event perception

Currently, theories of event segmentation focus on three types of mechanisms for event segmentation: detecting prediction errors (Zacks et al., 2007), detecting feature changes (Baker & Levin, 2015; Hymel et al., 2016), and detecting statistical structure (Baldwin et al., 2008; Schapiro et al., 2013). According to Event Segmentation Theory (EST; Zacks et al., 2007), comprehension systems make predictions based on a working event model. When prediction errors increase, the model resets and updates to reflect the new situation, resulting in the perception of an event boundary. Consistent with EST, here individuals engaged in predictive looking, and errors in predictive looking were correlated with segmentation probabilities.

Other accounts of event segmentation emphasize learning the sequential structure to chunk continuous experience into units. Some accounts note if there are recurring sequences in a domain, viewers can learn to segment at the relatively unpredictable joints between these predictable sequences (e.g., Baldwin et al., 2008). These models entail that there should be a time lag between event boundaries and the peak of prediction error, because an observer would anticipate the increase in prediction error in the imminent future and then segment before the peak of prediction error. We did not observe that prediction errors peaked after event boundaries were detected (see Supplemental Materials section 3). However, it is possible that in the segmentation task button-pressing lags relevant internal processes, because participants take some time to decide to press the button and to execute the response. Future studies could measure this lag directly by recording neural activity or manipulate it with instruction. Related theories focus on community structure—statistical sequences in which clusters of states tend to occur nearby in time (Schapiro et al., 2013). Theories proposing that viewers segment at transitions from one community to another entail that there should be circumstances in which prediction error can be equated, but event boundaries occur at points of transition from one collection of associated states to another. This could be directly tested using predictive looking but would likely require activities to be constructed with somewhat unnatural statistical structures.

Finally, some theories propose that event boundaries are determined through retroactive inferences (Baker & Levin, 2015; Hymel et al., 2016; Papenmeier et al., 2019). Baker and Levin (2015) suggest the spatial configuration of a recently encountered scene is maintained in memory and compared to current perceptual input. Changes in spatial configurations lead to segmentation. Papenmeier et al., (2019) showed that participants segment fewer times when causal continuation information can be retroactively inferred. The present findings indicate that prediction error precedes and thus could cause segmentation; however, this does not rule out that retroactive change detection also can lead to event model updating.

Limitations and future directions

One limitation of the approach taken here is that it depends on predictive looking to actors' hands. This is appropriate for activities involving goal-directed interactions with objects, but not for activities such as engaging in conversation or navigating an environment. Within the current stimulus set, this can be seen in the fact that predictive looking was less evident in the exercise movie, for which the hands' location was less informative.

A potential generalization of the approach would be to leverage computational models to derive saliency maps for each frame and treat them as the target of prediction. However, a challenge lies in selecting an appropriate computational model for movie viewing—particularly one that is appropriate for movie, rather than still frame.

Conclusion

Life is not a spectator sport. Event comprehension is an active process that, most of the time, serves to regulate action control. Even when someone observes events whose outcomes they cannot control, as in watching a movie, comprehension actively guides gaze to acquire information that will be informative. This means that the eyes offer a window on the brain's predictive processing, showing how the brain constructs models of the environment and updates them when prediction fails.

Author Note

Preliminary Dissemination

Preliminary versions of the ideas and data presented in this manuscript were disseminated at Psychonomic 2022 and VSS 2024. We gratefully acknowledge the Dynamic Cognition Lab for their insightful discussions and constructive feedback, which significantly enhanced our work. Disseminating these findings at these forums allowed us to engage with the broader research community and refine our analyses prior to final submission.

Author Contributions Statement

Sophie (Xing) Su: Conceptualization, Software, Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft, Writing – Review & Editing

Matthew A. Bezdek: Conceptualization, Software, Data Curation, Project Administration, Writing – Review & Editing

Tan T. Nguyen: Data Curation, Writing – Review & Editing

Christopher S. Hall: Data Curation, Project Administration, Writing – Review & Editing

Jeffrey M. Zacks: Conceptualization, Funding Acquisition, Resources, Supervision, Writing – Review & Editing

Constraints on Generality

This article, along with the supplement, utilizes data from two participant samples. Event segmentation data were obtained from participants recruited on Amazon Mechanical Turk using CloudResearch. For the eye-tracking study, we recruited participants from the Washington University in St. Louis subject pool. The CloudResearch sample is more diverse in age, education, and culture. Both segmentation patterns and viewing patterns could be influenced by participants' knowledge and culture; thus, it would be valuable to replicate both the behavioral and eye-tracking data collection across cultures.

The chosen films depicted activities that are typical aspects of daily life. However, it is plausible that the cognitive mechanisms observed may be specific to these highly familiar activity types. Further exploration is needed to investigate whether this relationship holds true for unfamiliar activities or during the learning phase, such as when children are acquiring new skills.

It is also possible that the cognitive mechanisms producing these effects may differ across development, aging, or neurodiversity.

References

Badler, J. B., & Heinen, S. J. (2006). Anticipatory Movement Timing Using Prediction and External Cues. *The Journal of Neuroscience*, *26*(17), 4519–4525. https://doi.org/10.1523/JNEUROSCI.3739-05.2006

Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition*, *136*, 14–29. https://doi.org/10.1016/j.cognition.2014.11.030

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407. https://doi.org/10.1016/j.cognition.2007.07.005

Baldwin, D., & Kosie, J. E. (2021). How Does the Mind Render Streaming Experience as Events? *Topics in Cognitive Science*, *13*(1), 79–105. https://doi.org/10.1111/tops.12502

Barnes, G. R., & Collins, C. J. S. (2008). Evidence for a Link Between the Extra-Retinal Component of Random-Onset Pursuit and the Anticipatory Pursuit of Predictable Object Motion. *Journal of Neurophysiology*, *100*(2), 1135–1146. https://doi.org/10.1152/jn.00060.2008

Bezdek, M., Nguyen, T. T., Hall, C. S., Braver, T. S., Bobick, A. F., & Zacks, J. M. (2022). The multi-angle extended three-dimensional activities (META) stimulus set: A tool for studying event cognition. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01980-8

Bubic. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2010.00025

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., & Sheikh, Y. A. (2019). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Corlett, P. R., Mollick, J. A., & Kober, H. (2022). Meta-analysis of human prediction error for incentives, perception, cognition, and action. *Neuropsychopharmacology*, *47*(7), 1339–1349. https://doi.org/10.1038/s41386-021-01264-3

Diaz, G., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013). Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of Vision*, *13*(1), 20–20. https://doi.org/10.1167/13.1.20

Eisenberg, M. L., & Zacks, J. M. (2016). Ambient and focal visual processing of naturalistic activity. *Journal of Vision*, *16*(2), 5. https://doi.org/10.1167/16.2.5

Eisenberg, M. L., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications*, *3*(1), 53. https://doi.org/10.1186/s41235-018-0146-z

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. (2020). Structured Event Memory: A neuro-symbolic model of event cognition. *Psychological Review*, *127*(3), 327–361. https://doi.org/10.1037/rev0000177

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, *100*(1–3), 70–87. https://doi.org/10.1016/j.jphysparis.2006.10.001

Geerligs, L., Van Gerven, M., & Güçlü, U. (2021). Detecting neural state transitions underlying event segmentation. *NeuroImage*, *236*, 118085. https://doi.org/10.1016/j.neuroimage.2021.118085

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, *108*(supplement_3), 15647–15654. https://doi.org/10.1073/pnas.1014269108

Gold, D. A., Zacks, J. M., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive Research: Principles and Implications*, *2*(1), 1. https://doi.org/10.1186/s41235-016-0043-2

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research*, *217*(1), 125–136. https://doi.org/10.1007/s00221-011-2979-2

Henderson, J. M. (2017). Gaze Control as Prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

Hymel, A., Levin, D. T., & Baker, L. J. (2016). Default processing of event sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 235–246. https://doi.org/10.1037/xhp0000082

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Kuperberg, G. R. (2021). Tea With Milk? A Hierarchical Generative Framework of Sequential Event Comprehension. *Topics in Cognitive Science*, *13*(1), 256–298. https://doi.org/10.1111/tops.12518

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience*, *3*(12), 1340–1345. https://doi.org/10.1038/81887

Maia, T. V. (2009). Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 343–364. https://doi.org/10.3758/CABN.9.4.343

Malcolm, G. L. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision*, *10*(2), 1–11. https://doi.org/10.1167/10.2.4

Maldonato, M., & Dell'Orco, S. (2012). The Predictive Brain. *World Futures*, *68*(6), 381–389. https://doi.org/10.1080/02604027.2012.693846

Mann, D. L., Spratford, W., & Abernethy, B. (2013). The Head Tracks and Gaze Predicts: How the World's Best Batters Hit a Ball. *PLoS ONE*, *8*(3), e58289. https://doi.org/10.1371/journal.pone.0058289

Mushtaq, F., Bland, A. R., & Schaefer, A. (2011). Uncertainty and Cognitive Control. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00249

Nguyen, T. T., Bezdek, M. A., Gershman, S. J., Bobick, A. F., Braver, T. S., & Zacks, J. M. (2024). Modeling human activity comprehension at human scale: Prediction, segmentation, and categorization. *PNAS Nexus*, *3*(10), pgae459. https://doi.org/10.1093/pnasnexus/pgae459

Papenmeier, F., Brockhoff, A., & Huff, M. (2019). Filling the gap despite full attention: The role of fast backward inferences for event completion. *Cognitive Research: Principles and Implications*, *4*(1), 3. https://doi.org/10.1186/s41235-018-0151-2

Richmond, L. L., & Zacks, J. M. (2017). Constructing Experience: Event Models from Perception to Action. *Trends in Cognitive Sciences*, *21*(12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005

Romberg, A. R., & Saffran, J. R. (2013). Expectancy Learning from Probabilistic Input by Infants. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00610

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2016). Task and context determine where you look. *Journal of Vision*, *7*(14), 16. https://doi.org/10.1167/7.14.16

Sasmita, K., & Swallow, K. M. (2022). Measuring event segmentation: An investigation into the stability of event boundary agreement across groups. *Behavior Research Methods*, *55*(1), 428–447. https://doi.org/10.3758/s13428-022-01832-5

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Shin, Y. S., & DuBrow, S. (2021). Structuring Memory Through Inference-Based Event Segmentation. *Topics in Cognitive Science*, *13*(1), 106–127. https://doi.org/10.1111/tops.12505

Smit, A. C., & Van Gisbergen, J. A. M. (1989). A short-latency transition in saccade dynamics during square-wave tracking and its significance for the differentiation of visually-guided and predictive saccades. *Experimental Brain Research*, *76*(1), 64–74. https://doi.org/10.1007/BF00253624

Su, X., & Zacks, J. M. (2025, January 13). Predictive Looking and Predictive Looking in Everyday Activities. Retrieved from osf.io/vsbaq

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979–1008. https://doi.org/10.1207/s15516709cog2806_5

Zacks, J. M. (2020). Event Perception and Memory. *Annual Review of Psychology*, *71*(1), 165–191. https://doi.org/10.1146/annurev-psych-010419-051101

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction Error Associated with the Perceptual Segmentation of Naturalistic Events. *Journal of Cognitive Neuroscience*, *23*(12), 4057–4066. https://doi.org/10.1162/jocn_a_00078

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307–327. https://doi.org/10.1037/a0015305

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29