Seeing the Future: Anticipatory Eye Gaze as a Marker of Memory

Yamin D.[1,2] *, Schmidig J.F.[1,2] *, Sharon O.[3] *, Nadu, Y.[2], Nir, J.[2,4], Ranganath, C. [5], Nir Y[1,2,6,7] †

1. Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
2. Department of Physiology & Pharmacology, Faculty of Medicine, Tel Aviv University, Israel
3. Center for Human Sleep Science, Department of Psychology, Helen Wills Neuroscience Institute, University of California, Berkeley, United States
4. The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, Israel 91904
5. Center for Neuroscience, Department of Psychology, University of California Davis, Davis, CA, USA
6. Department of Biomedical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv 6997801, Israel
7. The Sieratzki-Sagol Center for Sleep Medicine, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

* Equal contribution
† Corresponding author

Abstract

Human memory is typically studied by direct questioning, and the recollection of events is investigated through verbal reports. Thus, current research confounds memory per-se with its report. Critically, the ability to investigate memory retrieval in populations with deficient verbal ability is limited. Here, using the MEGA (Memory Episode Gaze Anticipation) paradigm, we show that monitoring anticipatory gaze using eye tracking can quantify memory retrieval without verbal report. Upon repeated viewing of movie clips, eye gaze patterns anticipating salient events can quantify their memory traces seconds before these events appear on the screen. A series of experiments with a total of 126 participants using either tailor-made animations or naturalistic movies consistently reveal that accumulated gaze proximity to the event can index memory. Machine learning-based classification can identify whether a given viewing is associated with memory for the event based on single-trial data of gaze features. Detailed comparison to verbal reports establishes that anticipatory gaze marks recollection of associative memory about the event, whereas pupil dilation captures familiarity. Finally, anticipatory gaze reveals beneficial effects of sleep on memory retrieval without verbal report, illustrating its broad applicability across cognitive research and clinical domains.

Keywords. eye tracking, consolidation, sleep, machine learning, real-life, gaze, eye-movements, anticipation

Introduction

Traditionally, human memory assessment has predominantly relied on explicit retrieval tasks, where participants verbally categorize learned items as familiar or novel[1–3]. However, relying exclusively on verbal reports entails significant limitations. First, from a basic science standpoint, it confounds memory per se with the ability to access and report the memory. Accordingly, it remains unclear whether the underlying neural process, a deleterious effect of brain disease, or the benefit of sleep pertain to memory, or to the ability to access or report the memory engram. In addition, studying memory retrieval is limited (or completely impossible) in populations where explicit reports are unreliable or absent, such as in aphasia patients, newborns, or animals. A reliable and well-validated method to assess episodic-like memory of events independent of the report would represent a major advance.

A similar challenge exists in the study of consciousness, where specialized paradigms have been developed to separate the neural correlates of consciousness from those related to its report[4]. Many such paradigms employ eye tracking, suggesting that an eye-tracking-based "no-report" paradigm could also be effective in studying memory retrieval. Additional motivation arises from rodent studies, where differences in exploratory behavior are routinely used to index memory[5,6]. Humans and other primates are visual-centric and primarily rely on eye movements to explore their environment[7–9]. Along this line, Kano & Hirata (2015)[10] leveraged visual exploration to assess memory retrieval by tracking great apes' eyes during a single movie clip. Inspired by this approach, we hypothesized that profiles of human visual gaze exploration could allow the creation of a general method to capture distinct memory traces.

Indeed, over the past two decades, eye tracking has been increasingly used to study human memory. Several studies have shown that memory can affect the way people visually explore static images[11–18]. For example, recognition of photographs is associated with a decrease in distributed overt attention[16] and familiar faces elicit fewer fixations compared to unfamiliar ones[15,19]. Such effects can occur in conjunction with verbal reports or in their absence[12,13]. Few studies have used single movie clips to establish visual exploration in animals[10] or children[20] Memory in preverbal infants has been specifically established using eye-tracking tests[21,22]. However, it remains unclear to what extent eye tracking could index episodic-like memory in adults at the single-trial level.

We sought to go beyond the state-of-the-art by validating anticipatory gaze during dynamic video clips as a proxy for episodic-like memory retrieval in adult humans. We predicted that by using video clips, visual gaze could index context-dependent memory-guided prediction[23]

in the absence of the remembered visual context about to appear, thereby going beyond familiarity and image recognition. In addition, gaze exploration is not limited to conscious retrieval[14] and could, therefore, potentially provide a more sensitive measure of memory than explicit reports. Thus, we developed a method that uses anticipatory gaze patterns to index and quantify memory-guided prediction of events occurring at a specific place and time.

Here we introduce and validate MEGA (Memory Episode Gaze Anticipation), a no-report method based on eye tracking during repeated movie viewing to assess memory for events. Participants watched movies with predefined surprising events. When participants watch the movies for the second time and have already formed a memory of the event, their gaze is drawn to its location, anticipating its occurrence before it appears. In a series of experiments, we evaluate MEGA as an approach to quantify memory without report. We introduce a distance-based metric and employ statistical analysis and machine learning on eye-tracking data to capture gaze characteristics indicative of memory at the single-trial level. We then explore its correspondence with multiple explicit memory reports about the movie event and contrast it with other eye-tracking metrics, such as pupillometry that index context and familiarity. Finally, we demonstrate one application by studying the influence of sleep on memory consolidation via MEGA without explicit reports.

Results

We constructed the Memory Episode Gaze Anticipation (MEGA) paradigm with the aim of capturing memory retrieval in its raw form without the additional layer of verbal reports. Participants viewed short (3-26s) movie clips in two sessions conducted a few hours apart while gaze was monitored using an infrared video-based eye tracking system (see Methods). Each movie contained a surprising event that saliently occurred in an unexpected location and time (e.g., an animal suddenly appearing behind a rock, Figure 1A; see A1 for the full movie). The movie clips were custom-designed animations depicting simple ecological scenarios. Precise event timings and locations were equally distributed across the movie length and screen space (see methods). We hypothesized that in the first viewing, gaze patterns before the event occurs (‘pre-event) would be exploratory, whereas, in the second viewing, after the formation of memory, gaze patterns would anticipate the event and preferentially occur around the event location (Figure 1B). Accordingly, the memory for the surprising event may manifest as a difference between the 1st and 2nd viewings at pre-event intervals around the anticipated location of the event (red rectangle and heat maps in Figure 1B). To test this, the first experiment included 34 participants who viewed 65 movie clips in two viewing sessions

separated by a two-hour break (Figure 1C and see detailed methods). In the 2nd viewing session, verbal reports and confidence measures were collected after each movie, to allow comparison with standard explicit reports.

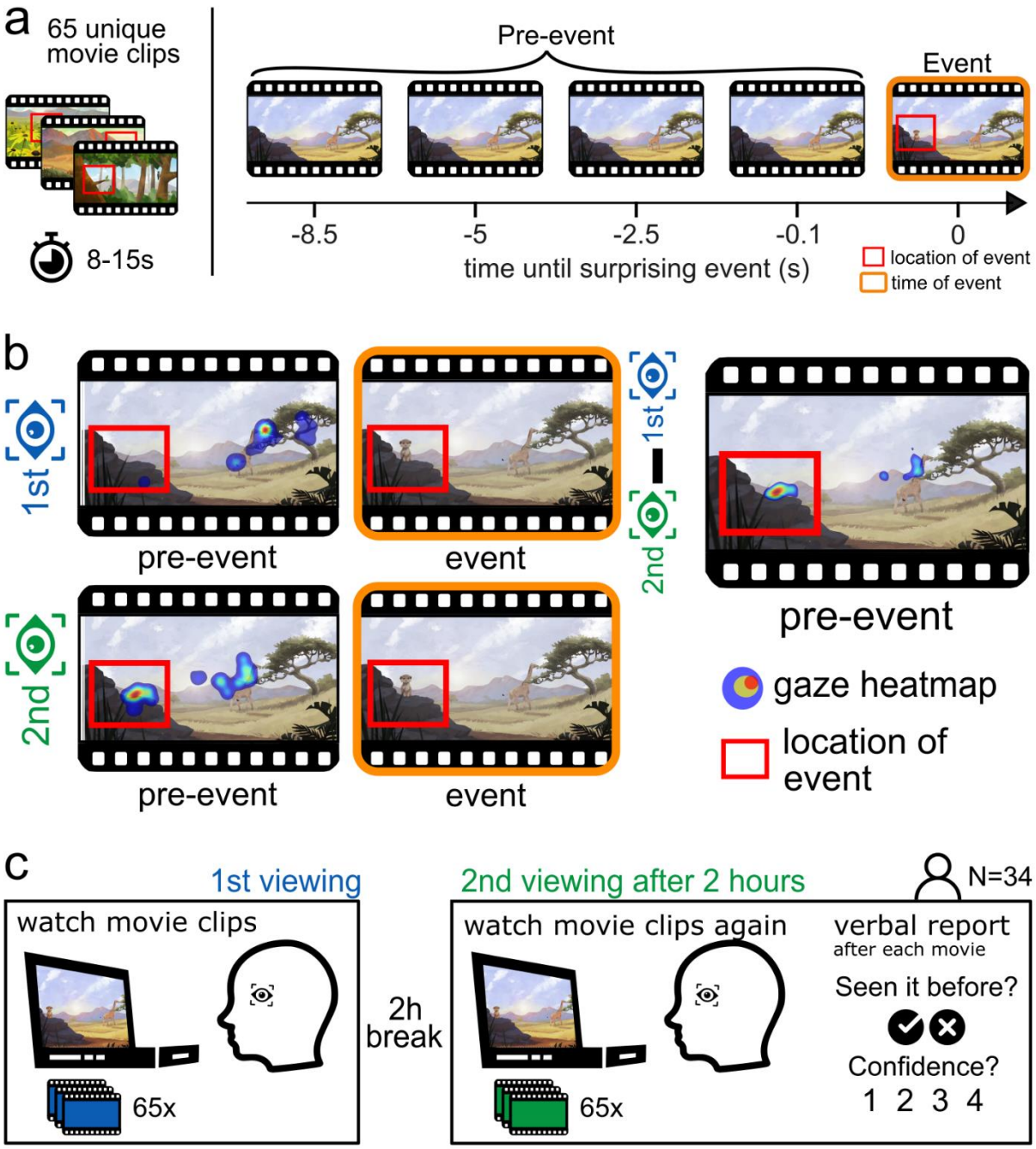

Figure 1. MEGA paradigm. (a) Visual Stimuli. 34 Participants watched 65 custom-made movie clips, each featuring a distinct surprising event. Each surprising event included a specific object at a specific location (red square) at a specific time (orange square). (b) Analysis rationale: gaze patterns were compared during the 1st and 2nd viewings to uncover memory retrieval. Top row: the average gaze location during 1st viewing (blue); Middle

row: the same average gaze during the 2nd viewing (green). Bottom row: the difference between the gaze during 1st and 2nd viewings before the event appears on the screen. Participants were expected to gaze more often toward the location of the upcoming event, indicating their memory of the event. Heat maps represent gaze locations of a single move, averaged over time. (c) Experimental design: 65 movie clips were presented twice in randomized order, with a consolidation interval of 2 hours between viewings. Eye-tracking data was collected during both viewings. During the 2nd viewing, after each movie, participants reported whether they recognized the movie clip from the first session and rated the confidence of their response.

Gaze anticipation indexes memory for events

To quantify anticipatory gaze towards the remembered location in each movie viewing, we assessed, for each time point separately, the Euclidean distance from the gaze location to the event location (apriori defined by the animation studio that compiled the movies, Methods). Figure 2A illustrates this calculation, which utilized all data points (not only fixations) to transform the multivariate eye tracking data into a single time series that encapsulates anticipatory gaze behavior. During the 1st viewing, the distance from the event location was mostly large, but upon 2nd viewing, we observed that the gaze gravitated toward the expected location of the event before its onset (see representative example of a single movie in Figure 2B). We quantified this by the Gaze Distance (GAD) — the mean distance from each gaze point to the event location from the movie's beginning until the event onset. GAD is a simple indicator that can reveal a tendency to gaze closer to the event location before its onset. Computing the GAD across all movies for each participant (Figure 2C) revealed a significant convergence to the event location upon 2nd viewing, observed in 31/34 (91%) of participants (mean GAD 1st viewing=18.13±0.62° vs. mean GAD 2nd viewing=16.91±1.05°; $t(33)=583$, $p=4.07e-09$; Cohen's $d = 1.4$). Next, to observe the temporal dynamics of anticipatory gaze, GAD was averaged across all movies and participants without averaging across time (Figure 2D). This analysis revealed a closer gaze towards the event location in the 2nd viewing that was present throughout the seconds leading to its appearance (Figure 2D, green time-course), followed by a sudden drop in GAD after-event appearance (reflecting gaze towards the surprising event once it occurred).

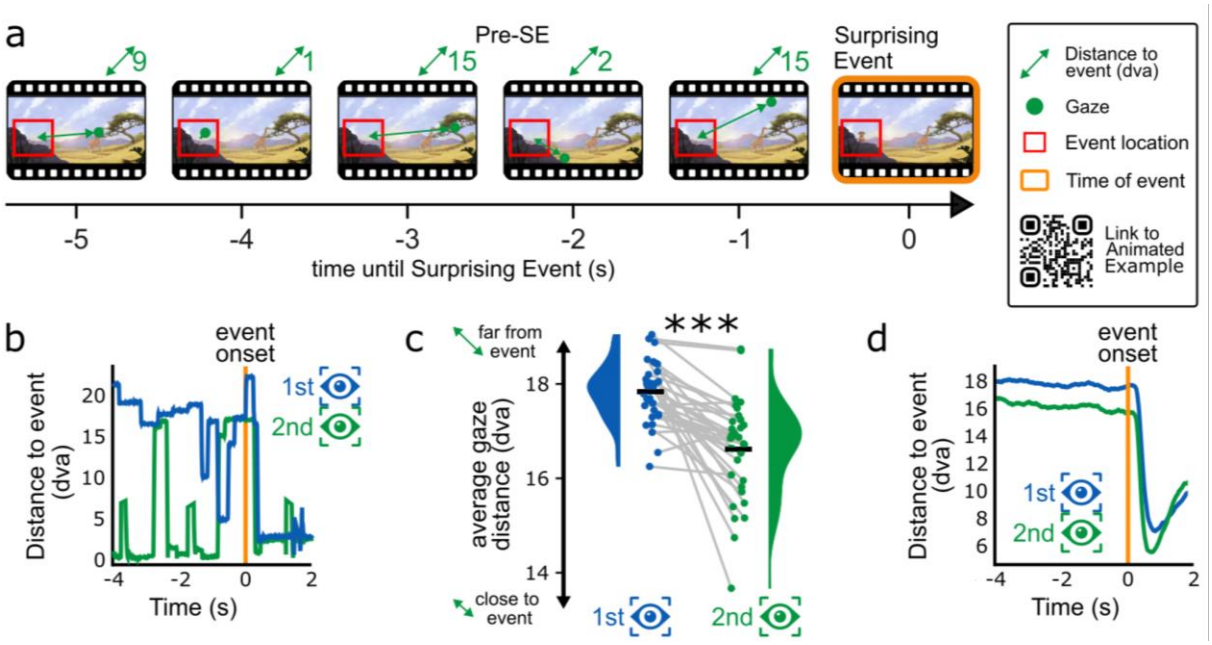

Figure 2. Anticipatory gaze effect. A) Gaze Proximity Measurement: participant's gaze trajectory while watching a movie clip. The Euclidean distance to the location of the surprising event center point is calculated at each time point in Degrees of Visual Angle (DVA). The QR code leads to an animated illustration of the methodology (for example, a movie with superimposed eye movements and gaze distance measures). B) Temporal Gaze Distance of a single participant during first (blue) and second (green) viewings of a clip. The SE's onset is marked by an orange line, serving as a temporal marker for gaze behavior. C) Gaze distance (GAD) comparison within participants across two viewings. The gaze average distance is calculated relative to the event center prior to its appearance on the screen and averaged across movies. GAD significantly decreases during the second viewing compared to the first viewing indicating memory for the event's location. Black horizontal line: group average, colored area: density estimate of GAD distribution, dots: individual participant average, line connects 1st and 2nd viewing of the same participant. D) Averaged gaze temporal dynamics: An aggregate view of gaze distance from the event across all participants and movies, time-locked to the event onset. The distances as a function of time for the first (blue) and second (green) viewings show the participants' anticipation of the event based on their prior viewing experience. Participants were surprised and looked at the surprising event immediately after its appearance, which can be seen in the steep slope after the event onset (orange line). *** = $p<0.0001$

Machine learning discriminates first and second viewings at a single trial resolution

We tested to what extent machine learning (ML)-based classification of multiple eye tracking features could extend the intuitive GAD metric and accurately identify memory traces at the single-trial level. First, we focused on reducing gaze distance as captured by the GAD metric (above). We employed the XGBoost classification algorithm[24] at the single-trial level, complemented by a leave-one-subject-out cross-validation method (Figure 3A). Models achieved an average correct classification of 69±10% (chance: 50%), demonstrating the model's ability to distinguish, based only on GAD, whether a single trial's data represented the first movie viewing or whether it had been viewed before- 1st or 2nd viewing (Figure 3B). The Receiver Operating Characteristic (ROC) curves for each subject-specific model and the

average ROC curve indicated consistent model performance across subjects (Figure 3C). Accordingly, the area under the ROC curve (AUC-ROC) was 0.75±0.26, quantifying the model's overall effectiveness ($t(33) = 10.76$, $p = 2.47e-12$ via 1-sample t-test). In 32/34 of the participants (94%), model accuracy was greater than chance levels, with an average accuracy of 0.69±0.1, attesting to its reliable prediction of identifying memory traces (Figure 3D).

Next, we employed an exploratory 'bottom-up' ML strategy by broadening our analytical scope to encompass a wider array of eye-tracking features, irrespective of our initial GAD metric. We manually engineered multiple features from eye-tracking data centered around the event location, thereby minimizing potential session-specific confounds such as differences in time-of-day and cognitive load due to tasks and reports in 2nd viewing (Methods). Such features (87 in total) included aspects such as fixation count ratios relative to the event location, the velocity and visual angle of saccades directed towards the event, and the change in pupil radius during fixations within the event location compared to prior fixations, thereby capturing event-related eye tracking spatial and temporal features. With these features, models achieved an average classification accuracy of 71±11.73% and 70±12.34% for 1st and 2nd viewings, respectively. Associated AUC-ROC metrics exhibited an average of 0.75±0.26 ($t(33) = 10.85$, $p = 1.98e-12$ via 1-sample t-test). In 32/34 of the participants (94%), model accuracy was greater than chance levels, with an average accuracy of 0.7±0.11. Remarkably, our initial analyses based solely on the GAD metric already captured nearly all variance within the dataset and showed comparable performance to the extensive 87-feature model, attesting to GAD's potent predictive value and showing that distance metrics alone are sufficient for precise identification of episodic-like memory traces in the MEGA paradigm.

To further understand which eye-tracking features are most effective in capturing event memory, we incorporated SHAP (SHapley Additive exPlanations) feature importance analysis, a game theory-based method to distill and quantify each feature's influence on the model's output[25]. SHAP provides an average impact of each feature on model prediction by examining its performance with and without the presence of each feature across all possible feature combinations. This analysis identified distance-based features as particularly informative, with GAD ranking as the top feature (Figure 3E). Along the same lines, a waterfall plot of a representative trial illustrates how individual features (especially GAD) cumulatively influence the model’s decision-making process for a single trial, highlighting the predictive power of GAD in indexing memory in the MEGA paradigm (Figure 3F).

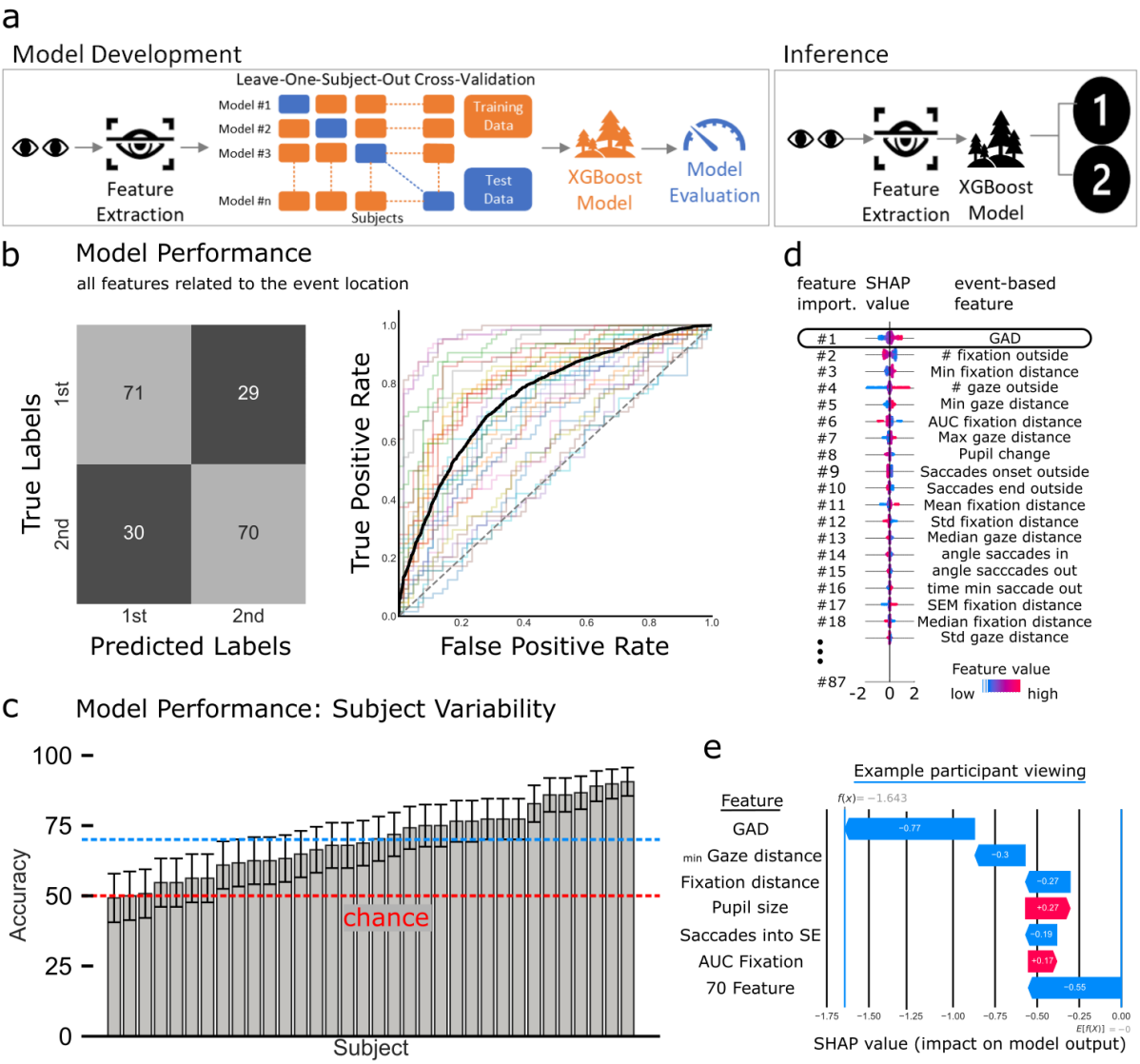

Figure 3. Single-trial Predictive Modeling Using Machine Learning of Eye Tracking Data Features (a) Model Development and inference: We utilized an XGBoost classifier, leveraging 87 features derived from the event location and timing to perform single-trial level classification. The model's robustness was ensured through a leave-one-subject-out cross-validation technique, where the algorithm was iteratively trained on the dataset excluding one subject, which was then used for testing. (b) Confusion Matrix: The average confusion matrix across all subjects, summarizing the classification performance of XGBoost models. Values along the diagonal (top left and bottom right) entries (71%,70%) represent the percentage of correct classifications, indicating true positives and true negatives, respectively. Off-diagonal values (bottom left to top right) entries (30%, 29%) denote misclassification probability. The Receiver Operating Characteristic (ROC, on the right) curves for XGBoost models, each tested on a distinct subject within our study cohort. These curves plot the true positive rate (TPR) against the false positive rate (FPR) at various decision threshold levels. Each trace represents the ROC curve for a subject-specific model (different color for each participant), delineating the trade-off between sensitivity and specificity. The black line denotes the mean ROC curve across all models. (c) Performance across individual participants (bottom left): bar chart displaying the classification accuracy for each subject, with the average accuracy across all subjects (blue, 70%) and chance-level (red, 50%). Error bars signify confidence intervals reflecting the precision of the model for each subject. (d) SHAP feature analysis: ranking of the influence of various gaze metrics on the model's predictions. Out of 87 SE-based features, the distance-based features, particularly

GAD, emerged as the most important feature. (f) A waterfall plot provides an example of how individual features contribute to a single trial's classification.

Anticipatory gaze marks event recollection, whereas pupil size indexes context recognition

What aspects of memory does anticipatory gaze capture? To what degree does it reflect episodic-like memory for the event versus familiarity with the general context? To address these questions, we first tested how MEGA relates to simple verbal reports (‘Have you seen this movie before?’). We compared GAD scores in movies reported as ‘seen before’ vs. ‘not seen before’, focusing only on trials with reports associated with high confidence ratings (Methods, Figure 1B). We found a significant reduction in GAD upon 2nd viewing for movies that were reported to be seen before ($GAD_{1st\ v.}$= 17.97 ± 0.72, $GAD_{2nd\ v.}$= 16.7 ± 1.13, t(33)=572, p=3.72e^{-08}, Cohen's D = 1.34, Figure 4A) but also a significant effect for movies that (incorrectly) reported as ‘not seen before’ ($GAD_{1st\ v.}$= 18.35 ± 1.93, $GAD_{2nd\ v.}$= 17.44 ± 1.78, t(33)=556, p=4.4e^{-07}, Cohen's D = 0.49, Figure 4A). To directly compare the difference between the first and second viewings of the explicitly recognized and not-recognized movies, we computed the normalized decrease of GAD from 1st to 2nd viewing ($GAD^{1st\ viewing}$ - $GAD^{2nd\ viewing}$ / max($GAD^{1st\ viewing}$, $GAD^{2nd\ viewing}$), Figure 4B). Higher values reflect an anticipatory effect during the second viewing and thus reflect memory-guided behavior. This score exhibited a trend towards higher values for movies that were subsequently recognized than for not-recognized movies ($M_{MEGA\ recognized}$=0.07 ± 0.07, $M_{MEGA\ not\ recognized}$=0.05 ± 0.05, paired t-test: t(66)=1.7, p=0.093, Cohen's D = 0.41, Figure 4B).

To better understand what aspects of memory are captured by anticipatory gaze beyond ‘Have you seen this movie before?’, a second experiment was conducted to evaluate multiple dimensions of memory reports in detail, aiming to distinguish between event recollection and context recognition (Figure 4C). Participants watched 48 animated movies depicting similar surprising events. After a four-hour break, they watched these movies again together with 12 novel movies in a randomized order. However, in this experiment, the 2nd session included the exact same movies, except the surprising event was omitted. Because the surprising event was not presented during the 2nd viewing, we could follow up with an array of retrieval tasks immediately after each movie. These tasks aimed to provide additional sensitivity to better investigate the relationship between explicit memory and anticipatory gaze. We collected the following five verbal reports: recognition, free recall, object recognition, event location recall, and temporal recall (for details, see Methods).

Results reliably replicated the anticipatory gaze effect with a new group of participants. 30/30 of participants (100%) demonstrated significantly greater gaze proximity to the event location during the second viewing ($M_{MEGA\text{-}score}$=0.086 ± 0.037, t(29)=12.8, p=2e^{-13}, Figure 4D). Next, we analyzed the GAD of each movie based on its explicit report in three categories (Methods): event recollection (correct movie recognition, as well as object recognition and event location recall), context recognition (scenery was recognized, but the object recognition or event location recall was incorrect), and unrecognized movies (incorrect recognition task, independent of the answer in event location recall or object recognition). Temporal recall (when precisely the event occurred) was not further analyzed because retrieval performance was at chance (p=0.81). Behaviorally, retrieval performance for movie recognition, object recognition, and location recall were all significantly above chance (all p<2e^{-09}, Figure S1).

Analyses revealed that MEGA scores were significantly higher for movies where participants recollected the event in full, highlighting the sensitivity of MEGA to episodic-like recall (Figure 4E & 4F). Accordingly, ANOVA with the dependent variable MEGA score and the factor memory content (event, context, and no memory) revealed a significant main effect (F(2,87)=14.11, p=4.9e^{-06}). Post-hoc pairwise comparison revealed a significantly higher MEGA-score for full recollection of the event compared to MEGA-scores of movies where only context was recognized (M_{event}=0.12 ± 0.06, $M_{context}$=0.08 ± 0.04, p_{Tukey}=0.002) or compared to unrecognized movies (M_{event}=0.12 ± 0.06, $M_{no\text{-}recognition}$=0.05 ± 0.05, p_{Tukey}=0.000004), but the latter two conditions did not differ significantly (context vs. no-recognition: p_{Tukey}=0.2). Moreover, the MEGA score of each movie correlated with participants' precision in reporting the event location such that stronger anticipatory gaze effects are associated with higher proximity to the event location (pearson r = 0.22, p<2.2e^{-16}, 1438 movies). Accordingly, in each trial, the higher the precision in explicitly reporting the event location, the closer the anticipatory gaze was to that location before it occurred on the screen. Accordingly, the MEGA score of participants correlated with the number of trials categorized as event recollection (pearson r=0.37, p=0.043, N=30) but did not correlate with the number of trials that were not recognized (r=-0.046, p=0.81). Correlation with the number of trials where only the context was recognized exhibited a marginally significant negative correlation (r=-0.36, p=0.05).

Finally, we investigated anticipatory gaze within the context of participants' free recall of surprising events, rather than location recall and forced-choice object recognition. As expected, the MEGA score was bigger if the object of a surprising event was explicitly recalled, compared to movies for which participants could not remember the event's object (free recall: MEGA score $_{recalled}$=0.14 ± 0.08, MEGA score $_{not\ recalled}$=0.07 ± 0.03, paired t-test: t(28)=5.46, p=7.8e^{-06}, Cohen's D =1.22). Furthermore, this difference in anticipatory gaze linearly

increased for the participants' ratio of objects recalled and forgotten (pearson r=0.56, p=0.0016, N=29).

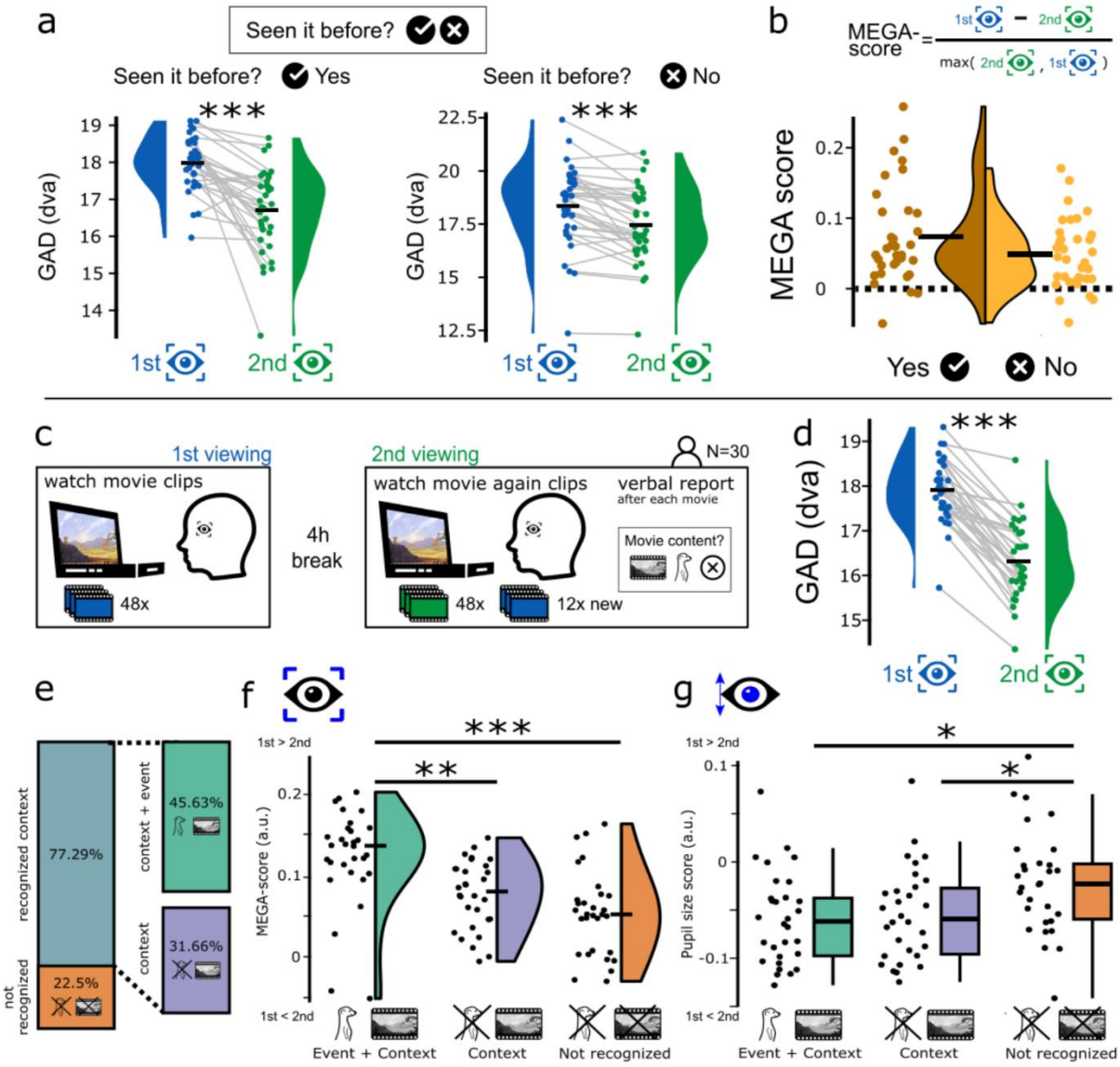

Figure 4: The relationship between anticipatory gaze and explicit memory reports A) First experiment: gaze in the second viewing of movies explicitly reported as "seen" (green) showed greater proximity to the event location compared with the first viewing (blue) for both recognized and unrecognized movies, signifying memory retrieval. B) The MEGA score quantifies the change in GAD between the first and second viewings, calculated using the formula above. This MEGA score is higher for movies that were explicitly remembered versus not remembered. C) Procedure of the second experiment: Participants completed a similar task except surprising events were omitted from the second viewing and followed by an extensive explicit memory report. Memory was assessed to determine what exactly participants remember. D) Replication of the anticipatory gaze effect: consistent with experiment 1, anticipatory gaze appears in the second viewing (green) and not in the first viewing (blue) across all naïve 30 subjects. E) Categorization of explicit report. Participants did (i) not recognize the movie clip at all (orange), (ii) recognize the context alone (violet), or (iii) recognize the context and recollected the event (green). F) The MEGA score is higher for movies in which participants remembered the event in full, compared to those with only scenery memory or completely forgotten movies. There is no significant difference between movies with scenery recognition

and movies that were not recognized. G) Pupil size variation and memory: Examines the decrease in pre-event pupil size from first to second viewing, in relation to scene recognition and event recollection, highlighting smaller changes (1st vs 2nd viewing) in unrecognized cases. Box plots represent the median, 95% confidence intervals, density plots, and dot plots. * = p<0.05, ** = p<0,01, *** = p<0,001

To further distinguish MEGA from context recognition, we focused on pupil dilation as an index of familiarity upon repeated stimulus presentation[26–29]. Specifically, previous findings suggest that pupil dilation is increased for recognized words compared to unrecognized words[30,31]. In line with this literature, we also found that pupil size was larger for the second viewing compared to the first viewing (pupil size $_{1st\ v}$=4436. 25 ± 378.12, pupil size $_{2nd\ v}$=4647.57 ± 378.12, t = 59, p-value = 1.974e^{-05}). Next, we compared changes in pupil size across repeated movie viewings in relation to the verbal report. To this end, we computed a normalized pupil size score using the same formula used for the MEGA score (1st viewing - 2nd viewing / max(1st viewing, 2nd viewing), negative values reflect a larger increase in pupil size, Methods). We found a larger increase in pupil size for movies that were recognized compared to movies that were not recognized (higher pupil size score for 'not recognized' in Figure 4G). ANOVA with pupil dilation score as the dependent variable and the explicit memory questions as a factor revealed a significant main effect ($F(2,87)=5.05$, $p=0.008$). Post-hoc pairwise comparison revealed a significantly higher pupil score for movies where the event was recollected in full (M_{event}=-0.059 ± 0.048, $M_{no\text{-}recognition}$=-0.021 ± 0.054, $p_{Tukey}=0.013$) and for movies where the context was recognized ($M_{context}$=-0.054 ± 0.049, $M_{no\text{-}recognition}$=-0.021 ± 0.054, $p_{Tukey}=0.033$) compared to unrecognized movies. Crucially, pupil dilation score did not differ between the movies with event recollection and context recognition, suggesting no specific relationship with event memory (event vs context: $p_{Tukey}=0.93$). These findings suggest that while the pupil is indicative of recognition of the context alone, the anticipatory gaze is guided by a richer memory that includes the recollection of event details, such as its location within the context.

Anticipatory gaze is Replicated in Naturalistic Movies

To what extent can anticipatory gaze be revealed using other movies, not necessarily animations compiled specifically for scientific research? We set out to test the degree to which anticipatory gaze captures episodic-like memory recall in settings that closely mimic real-world experiences, with the aim of bridging laboratory research and everyday memory. To this end, we performed a third experiment, where 32 naïve participants viewed 100 YouTube videos (Figure 5A). First, it was necessary to define the surprising event location and timing since these were not defined as a-priori as in the tailor-made animations. A group of 55 independent participants marked the spatial and temporal coordinates of the surprising event. Each movie's

event location was defined as the median (X, Y, t) of their choice of coordinates. 48 movies that exhibited a maximal level of consensus (within one standard deviation of the median time or location; see Methods) were used for subsequent analysis. Once again, analysis of GAD preceding the surprising event replicated the anticipatory gaze effect, with a completely different set of stimuli and a different group of subjects. A significant increase in gaze proximity to the event location was observed upon 2nd viewing in all participants but one, robustly indexing memory without report ($t(31)=495$, $p=9.3e^{-10}$; W-SRT, Cohen's d = 1.52, Figure 5C). Increased proximity of anticipatory gaze to the event location was evident throughout the seconds leading to the event (Figure 5B). GAD declined upon the event onset in both 1st and 2nd presentations, reflecting gazing towards the event once it occurred, but this drop was less steep and interestingly began already *prior* to event onset, arguably since the gaze of participants was already “drawn” towards the event location by narrative cues in naturalistic movies compared to the highly unexpected event appearance in tailor-made animations. Analyzing GAD and MEGA computed separately for recognized and unrecognized movies (according to verbal report, Figure 5A) showed that the MEGA score significantly exceeded chance-level for both remembered and forgotten movies, replicating the observation in the previous experiments ($M_{1st\ v.}=10.96$ +-0.46, $M_{2nd\ v.}=10.04$+-0.72, $t(31)=560$, $p=2.5e-7$, for explicitly remembered trials; $t(31)=539$, $p=3.6e-6$ for explicitly forgotten events, Figure 5D). Together, the results show that the anticipatory gaze effect robustly replicates with naturalistic movies, attesting to the utility of this approach in diverse contexts, including real-life situations.

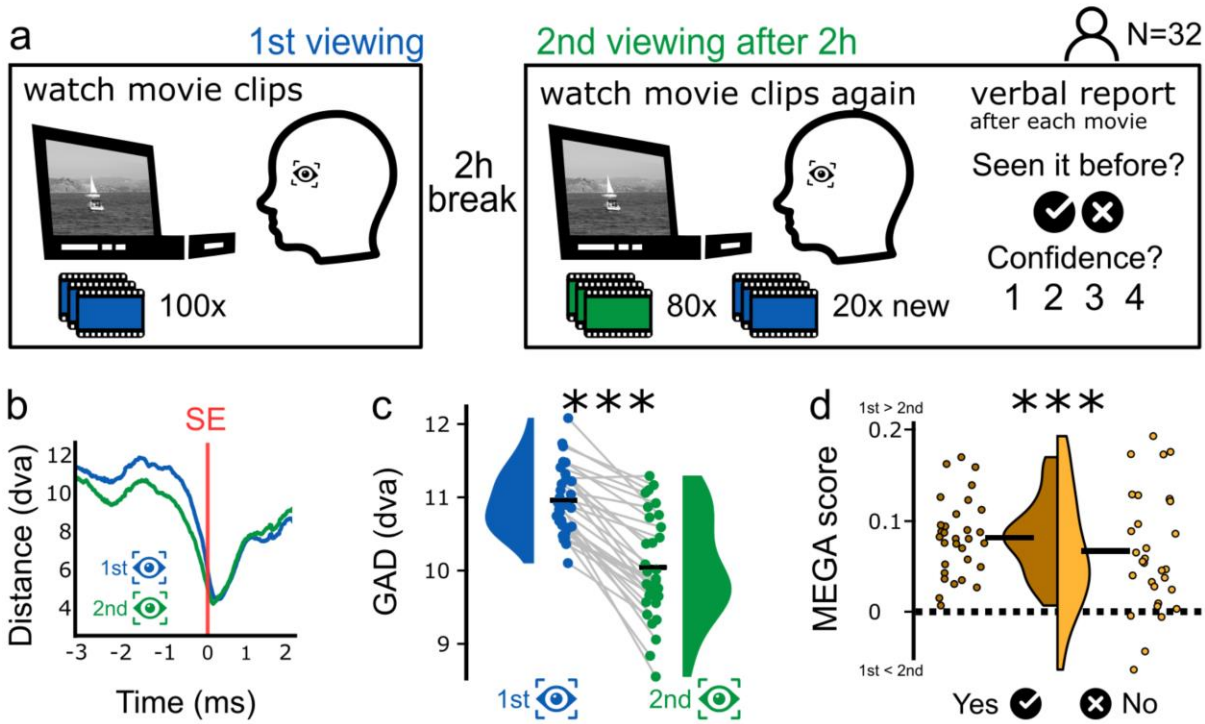

Figure 5: The anticipatory gaze effect is replicated in naturalistic (YouTube) videos. (a) Experimental procedure: 32 participants viewed 100 YouTube movies in two sessions spaced 2 hours apart. In the 2nd session, 20 novel ('lure') movies were included along with 80 of the original movies, and verbal reports were collected after each movie. (b) Average gaze distance to the surprising event for 1st (blue) and 2nd (green) viewing. A drop in gaze distance that already begins before event timing shows some anticipation of the surprising event in naturalistic movies, probably due to narrative and camera movements. (c) GAD decreases in 2nd viewing (green) compared to 1st viewing (blue): p = 9.3e-10, Wilcoxon Signed-Rank Test). (d) MEGA scores as a function of explicit memory reports, showing a higher MEGA score for movies that participants remember (p<0.0001).

Anticipatory gaze reveals sleep's benefit for memory consolidation without report

We demonstrate one potential application of the MEGA paradigm in terms of sleep benefits for memory consolidation. Naïve participants were recruited for a fourth experiment (Figure 5A), viewing naturalistic videos (identical to experiment 3) in two sessions separated by a 2h break that included either a nap opportunity for some individuals (n=19) or an equally long interval of wakefulness for other individuals (n=32, same data as in "naturalistic" experiment 3 shown in Figure 5). First, regardless of sleep or wake, we replicated the anticipatory gaze effect with a new group of participants. We observed significantly lower GAD reflecting anticipatory gaze (Figure 5B) in 16/19 of participants (84%), substantiated by statistical analysis ($M^{1st\ v.}$=11.3±0.62, $M^{2nd\ v.}$=10.12±0.87, t(18)=184, p=2.67e-5; W-SRT; Cohen's d = 1.54). This constitutes a third successful replication of the anticipatory gaze effect, reinforcing its reliability in capturing memory. Next, comparing sleep and wake, we found that the MEGA score in the nap condition was 23.6% higher than in the awake condition (M_{wake}=0.069 ± 0.043, M_{nap}=0.09 ± 0.048, W=1.7, p=0.0439, Cohen's d=0.48). This indicates that the nap had a

positive impact on memory consolidation as reflected in anticipatory gaze, demonstrating the potential of MEGA as a no-report paradigm for studying the relation between sleep and episodic-like memory consolidation.

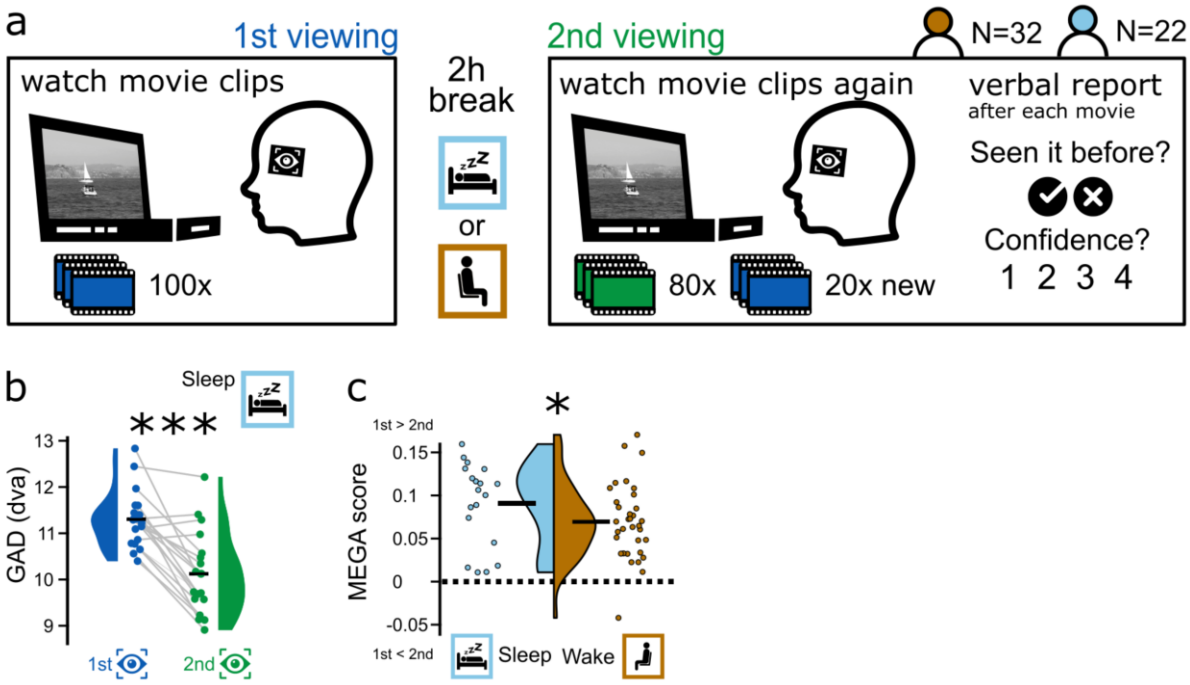

Figure 6. MEGA scores improve after sleep compared to wakeful rest. (a) Experimental design: participants (N = 19 for nap condition; N=32 for wakeful rest condition) watched naturalistic videos as in a "naturalistic" experiment with YouTube videos. The 2-hour interval between the 1st and 2nd viewing included either a nap opportunity or wakeful rest. (b) GAD in 1st (blue) and 2nd (green) viewing for the nap condition participants reveals a significant decrease in GAD during 2nd viewing (p = 9.3e-10, Wilcoxon Signed-Rank Test) (c) A comparison of MEGA scores (normalized difference in GAD between 1st and 2nd viewings) in the wake (brown) and sleep (cyan) conditions reveals greater MEGA after a nap (Mann-Whitney U test, p = 0.0439)

Discussion

This study shows that tracking gaze during repeated viewings of movies with surprising events constitutes an effective method for investigating memory without verbal reports. The results establish that during the second movie viewing, the gaze gravitates towards the event location, exhibiting memory-guided prediction that anticipates its occurrence. Gaze distance (GAD) can be used as an intuitive metric to capture the degree of this predictive anticipation, showing significantly higher proximity to the event location during the second viewing. The anticipatory gaze effect was captured before changes were visible in the movie and even when the surprising event was entirely absent in the second movie viewing. This establishes that it corresponds to memory-guided prediction irrespective of the visual cues that mark it. Anticipatory gaze is a highly robust effect that is consistently observed and replicated several times across multiple stimulus types and in different naïve groups of participants (N=126) - attesting to its versatility and utility across settings. Machine learning classifications of features extracted from gaze data identify memory traces at a single-trial level in new participants. In a separate experiment where we collected verbal reports about the movie events, we found that anticipatory gaze effects are largest when the surprising event was fully recalled, showing dissociation from pupil size measures not associated with recalling the surprising event. Finally, we illustrate how applying MEGA without verbal reports can effectively replicate the classical beneficial effect of sleep on declarative memory[32–34]. MEGA can be successfully employed using either naturalistic movies or custom animations specifically designed for this purpose, a resource made available for any future follow-up study (https://yuvalnirlab.com/).

The current study extends the growing literature on studying memory with eye-tracking[12,13]. Most previous studies in this domain probed the recognition of still images influences viewing patterns[11,13–16,29]. Accordingly, such studies have examined the “looking at nothing” effect, where empty regions in which events were previously encoded are preferably sampled[35–37]. By contrast, inspired by research on great apes[10], our anticipatory gaze approach captures memory-guided predictions in the absence of the encoded item (when it is not presented), a phenomenon distinct from cue-induced recall. Moreover, our results with dynamic videos resemble eye-tracking results in visual search[38,39] and contextual cueing[40] in picture scenes. Importantly, MEGA captures many single-shot memories of events in a way that requires knowledge about their properties (e.g., location) and correlates with explicit reports about event content. Strikingly, the anticipatory gaze effect is very sensitive, as it even detects memory when it cannot be consciously recalled. Thus, MEGA represents important and new features of memory retrieval that go above and beyond eye-tracking memory research to date.

Two complementary approaches were used to quantify the anticipatory gaze effect. The first and intuitive metric, GAD, captures the average Euclidean distance to the location of the surprising event from the beginning of the movie presentation until the event occurs and enables identification of memory-guided gaze in ~90% of participants. A second machine learning approach using the XGBoost classification algorithm is aimed at the predictive power of single-trial gaze. Surprisingly, machine learning applied on merely seven-second intervals of gaze data during pre-event movie viewing - without any averaging across movies or participants - was sufficient to significantly identify whether this viewing was associated with memory for the event (first or second viewing). Strikingly, classification performance was maintained even when we only used gaze distance instead of a comprehensive set of 87 gaze distance-related eye-tracking features. A post-hoc comparison of all features' importance confirmed that GAD was most informative.

What can estimate memory with MEGA, independent of verbal reports, provide? This approach entails several distinct advantages. First, MEGA can provide memory measures in clinical settings where verbal communication is challenging. For instance, individuals with cognitive impairments, such as aphasia or developmental disorders, often struggle with complex instructions or language comprehension and production. Along the same lines, a stroke patient may be unable to speak, yet we would like to estimate their memory capacity. Second, MEGA provides higher sensitivity than the reports typically collected, as is evident by the presence of significant effects even when participants reported not recognizing movies they had seen before. Such heightened sensitivity may hold significant promise for detecting early signs of preclinical Alzheimer's disease, where subtle memory deficits may go undetected by conventional questionnaire-based assessments. Third, MEGA can enhance consistency in memory research by increasing generalizability across participants who speak different languages. Finally, from a basic science perspective, current research employing reports confounds memory itself with the ability to articulate it. In this context, MEGA can help distill the brain activities and diseases affecting memory per se beyond its access and report.

Another important advantage of MEGA is that it goes beyond a binary ‘recognized’ vs. ‘non-recognized’ report to represent memory as a continuous quantitative variable. The MEGA-score metric is sensitive to various anticipatory behaviors—whether through a few prolonged fixations at the event location, multiple fixations around it, or even subtle proximity to the event location. This granularity allows us to reveal a linear relation between the anticipatory gaze scores and the individual precision in reporting the event's location for single movies. Accordingly, the degree of event recollection per participant correlates with the participant's MEGA score.

The findings broaden the discourse on unconscious memory as captured by eye-tracking[12] by extending the results from still images to dynamic videos. While lower than when consciously reported, we nevertheless find a significant anticipatory gaze score even in cases where participants confidently report they haven't seen the movie before. This suggests that, even without conscious indication, memory processes are at play to a sufficient extent to modulate eye movement behavior. While previous studies[13,41–43] have demonstrated unconscious relational memory in eye movements, MEGA leverages dynamic videos to detect unconscious memory effects without the object about to appear visible on the screen. Because this effect can reveal whether a stimulus has been encoded – despite explicit indication of not remembering it, MEGA opens new avenues for research investigating why some memories transcend into explicit recall while others remain implicit. Future studies can compare this memory-guided behavior beyond explicit reports with established measures of unconscious memory. This can disentangle to what extent the above-chance performance in "not recognized" scenarios stems from nuanced memory activation compared to mere exploratory behavior or attentional diversion. The fact that MEGA emerges independently of conscious recognition highlights its potential to uncover unconscious memory independent of traditional explicit memory reports.

Future research can apply and further develop MEGA across several domains. Follow-up questions concern the neuronal underpinnings associated with anticipatory gaze: to what extent does MEGA depend on activity in the hippocampus and the medial temporal lobe (MTL)? Is it comparable to other implicit eye-movement-based memory effects?[26,47,48] Future studies should assess how anticipatory gaze unfolds in entirely passive viewing conditions without any instructions. It should be acknowledged that the current results involve a task where participants were required to report if the movies were previously seen or not, thereby introducing an additional layer of cognitive processing that could influence the natural recall of events. Another application lies in research on memory consolidation. The present results already bear important implications in suggesting that sleep's benefits for memory consolidation extend beyond the mere ability to report memories, and such research can be developed further.

In clinical settings, the current findings in healthy adults hold promise for improving and refining existing non-verbal tasks[18,49–51] for early diagnosis of memory disorders in mild cognitive impairment (MCI) and for monitoring Alzheimer's disease progression. Standard cognitive and neuropsychological assessments like the MMSE[52] or the MoCA[53] are limited in detecting preclinical memory deficits. Beyond degeneration, MEGA can be used to assess memory upon damage to the medial temporal lobe (MTL)[54], when it is often unclear to what extent

damage affects mnemonic systems or their interface with other brain systems that enable conscious report. Another clinical application pertains to patients suffering from motor or language disorders that limit verbal report, such as those with aphasia[55,56], offering a non-verbal means of assessing memory integrity.

In conclusion, the MEGA paradigm introduces a significant advance in memory research. Utilizing eye tracking to study memory without verbal reports offers wide implications for both basic research in cognitive neuroscience and clinical fields.

Methods

Experimental Procedures

Four different experiments were carried out, each included a first viewing session (encoding), a break (consolidation period with different durations), and a second viewing session (retrieval) of the same movies, and in some cases additional new movies.

Experiment 1: tailor-made animation movies. Experiment 1 (Figures 1, 2, 3) was conducted during a daytime session with a two-hour break between the first and second movie viewing sessions. Following the completion of consent forms, and eye tracking setup (see below), the first viewing session started around noon (11:48 AM on average) and lasted an average 45 minutes. Participants watched 65 animated movies while pupils and gaze were monitored (see below). The movies were separated by a 2-second fixation cross presented on a blank grey screen. Then, the participants had a ~ two h break in which they were free to leave the lab unsupervised. During the second viewing session, each movie was followed by participants feedback indicating their explicit memory recall ("*Have you seen this movie before?*" with options Yes (1) or No (2)) and a second screen in which they rated their confidence ("*How confident are you in your answer?*" on a scale from "*Not at all (1)*" to "*Very confident (4)*").

Experiment 2: animation movies with extensive explicit memory reports. This experiment (Figure 4) examined in more detail the relation between anticipatory gaze and verbal reports, probing different aspects of memory. The experiment employed the exact same animation used in Experiment 1 for the first viewing. However, in the second viewing, the participants viewed an edited version of these movies where the surprising event E was omitted from each movie. In the first viewing, participants watched 48 movies. After a 2h break, participants were instructed about the tasks associated with the second viewing ahead and a single training example verified they could successfully follow the tasks. During the second session, participants watched the edited, event-lacking version of the 48 movies and 12 new movies in a random order. After each movie, the following five retrieval tasks were presented:

i) **movie recognition:** "Do you remember watching this movie before?" ("Yes" or "No"),
ii) **free recall:** "Please describe what was missing in the video and where did it happen?" Participants were instructed to say their answer aloud (for example: "snake on the right" or "frog in the center").
iii) **object recognition:** two objects were presented, and the participant was asked to indicate, "What was missing in the movie?". The lure object was created such that it was not unlikely to exist in the presented environment. For example, if the scene was underwater, possible lures would be "fish" or "crab" but not "elephant". Hence the lure list was fixed in such a way that each object appeared once as correct and once as incorrect answer.
iv) **event location recall:** a frame from the movie was presented, and participants were instructed to click on the screen where they thought the event should have happened. An answer was considered correct if they clicked in the correct quarter of the screen.
v) **temporal recall:** participants were asked about the timing of the event within the Movie: "When did the Surprising Event happen?". Answer options were a) "In the middle of the movie" or b) "In the end of the movie". Event timing was considered in

the middle or in the end according to the distribution of all events. If an event appeared before the median time of all SE's it was considered in the middle, and vice versa.

The free recall was self-paced, whereas the other four questions were limited to 5 seconds. The movie order in the first and second viewing was randomized, but which movies were presented once or twice was fixed and identical for all participants.

Experiment 3: Naturalistic movies. Experiment 3 (Figure 5) investigated anticipatory gaze at our sleep lab using naturalistic videos from YouTube. After setting up the eye tracking and EEG (below), the first viewing session started around 2 PM and included watching 100 naturalistic movies. Then, the participants had a 2-hour break and were instructed to remain awake. The second viewing session also included 100 movies (80 seen movies and 20 new ones) as well as a simple recognition task ("*Have you seen this movie before?* ("Yes" or "No"),") and confidence feedback ("*How confident are you in your answer?"* on a scale from "*Not at all (1)*" to "*Very confident (4)*"))") as in Experiment 1. Since the movies were publicly available on YouTube, there was a possibility that the participants had already seen some of them before the experiment. To address this, a recognition task was also carried out during the first viewing, and any trials with positive responses were excluded.

Experiment 4: Naturalistic movies with nap or wake. In Experiment 4, the setup was identical to Experiment 3, but introduced new participants and a modification: participants were given a nap opportunity during the 2h break while we monitored EEG, electrooculogram (EOG) and electromyogram (EMG) to monitor sleep. Data collection and sleep scoring were previously described[34]. Data from 8/29 participants was excluded due to short sleep duration (<50% sleep efficiency or < 15 minutes in N2/N3 sleep, Figure 6).

Participants

We tested a total of 126 participants across the four different experiments. Written informed consent was obtained from each participant prior to their involvement in the study as approved by the Institutional Review Board at Tel Aviv University (Experiments 1,2) or by the Medical Institutional Review Board at the Tel Aviv Sourasky Medical Center (Experiments 3,4). All participants were required to have normal or corrected-to-normal vision, reported overall good health, and confirmed the absence of any history of neurological or psychiatric disorders. Experiment 1 (animation movies) included a total of 34 participants (age range: 19-61, M ± SD = 26.2 ± 9.01 years; 23 (67 %) female). Experiment 2 (animation movies - with elaborate memory assessments) included 32 participants (age range: 18-40, M ± SD = 26.48 ± 4.3 years; 23 (72 %) female). 2 participants in this experiment were excluded from the analysis due to technical issues. Experiment 3 (naturalistic movies) included 32 participants (age range: 19-44, M ± SD = 27.03 ± 5.09 years; 9 (27 %) female). One participant was excluded from this experiment due to the loss of more than 50% of the eye-tracking data. Experiment 4 (naturalistic movies - with sleep consolidation) initially recruited 25 participants. Three participants were excluded due to incomplete data, lacking either EEG or eye tracking data across two sessions, and three other participants were excluded due to insufficient sleep (<50% sleep efficiency: the ratio of time spent asleep to the total time from sleep onset to final awakening). Therefore, 19 participants were subsequently analyzed (age range: 20-34, M ± SD = 27.13 ± 3.44 years; 10 (34%) female).

Movie stimuli and visual presentation

Overall, we used 185 different movie clips. We presented movies in full-screen mode to maximize engagement. Between the movie clips, a gray background with a fixation cross was displayed for 2 seconds to standardize the visual field and prepare the participants for the upcoming stimulus. In all experiments the order of the movie clips was pseudo-randomized for each participant to avoid order effects that could influence memory encoding and retrieval processes. Previously unseen movies were incorporated in the second session to increase task difficulty and enhance performance variability but were not subsequently analyzed. The experiments were coded in Python, using the PsychoPy[58] package and the Pylink package, which facilitates interfacing with EyeLink eye-trackers (see below).

Animation Movies (Experiments 1 & 2). For Experiment 1, we used 65 silent animated colored movie clips (see examples in Supp. Movies A1 and A2), each rendered at a resolution of 2304x1296 pixels. These clips varied in duration from 8 to 22 seconds, with an average length of 12.68 seconds. Movies were prepared with the aim of being universally comprehensible, regardless of age, cultural background, or cognitive ability. A key feature of these animations was the inclusion of a surprising event designed to be unmistakably noticeable upon first viewing while adhering to the following criteria: i) The event occurred after a minimum of 5 seconds from the start of the clip, ensuring viewers are engaged and have enough time to potentially recognize the context. ii) Once introduced, the event remained visible for at least 2 seconds. iii) It appeared peripherally, not centrally, varying in location and timing across clips to prevent predictability. iv) There were little or no hints to spoil the surprise and indicate the event's appearance. Additionally, the movies were crafted to include background scenes that are visually similar, yet not identical, across different movie clips (e.g., more than one movie had the same underground water scenery) to increase difficulty and avoid ceiling performance. Animations were crafted by Studio Plonter® (https://www.plonteranimation.com/). For the second viewing in Experiment 2, the animation studio produced an alternative version of 60 animations, identical in every aspect but without the surprising event. Visual stimuli were presented using a 15.6-inch (35x20 cm) monitor with a resolution of 3840x2160 pixels.

Naturalistic movies (Experiments 3 & 4). Movies (examples in Supp. Movies R1 and R2) included 100 silent black and white movie clips with a duration of 3.5 to 26 seconds downloaded from YouTube. This collection encompassed carefully selected clips from classic films (e.g., Charlie Chaplin) or contemporary single-shot natural scenes characterized by minimal camera movement, containing at least one SE. Scenes covered a broad spectrum of topics ranging from famous soccer goals through animal behaviors to human social interactions. Visual stimuli were presented using a 24-inch monitor (51x29 cm) with a resolution of 1920x1080 pixels.

Determining the surprising event. Determining the surprising event is central for quantification to the anticipatory gaze, as event onset needs to be defined for data analysis. For animated movies (Experiments 1, 2), the animation studio defined the location and timing of the surprising events. The onset and location of the surprising event were defined by the first pixel of the event, which becomes visible on the screen. For naturalistic movies (Experiments 3,4) the surprising event was defined experimentally by an online human focus group. We followed protocols used in prior research[59–61] and conducted an additional

dedicated online experiment with 55 participants who were asked to mark the time and location of the surprising event E in a random subset of the 80 movies used for analysis (excluding the 20 unseen and those that were replaced by them, M = 51.05±1.16 rated movies per participant). Participants watched the movies and then, during an additional viewing, pressed their mouse at the moment (t) and location (x, y coordinate) where the surprising event occurred. Each movie's event onset, duration, and location were then defined as the median of marked events. We discarded from the study movies that did not reach a minimal consensus on the timing and location of the surprising event since such movies typically contained multiple events and could not yield consistent eye-tracking patterns. Accordingly, movies where less than 70% of participants agreed on the location were excluded (consent meant localizing the location within 1/9 of the screen size). Overall, 48 of the 80 YouTube movies (60%) were deemed suitable for eye-tracking experiments.

Eye tracking

Eye tracking employed EyeLink 1000 Plus (SR Research) as in Sharon et al.[62] with a sampling rate of 500Hz. We first determined the dominant eye of each participant, utilizing a modified version of the Porta test[63,64]Participants were then instructed to position their heads on a chin rest 50-70cm from the screen to maximize eye tracking quality. Next, a 9-point calibration and validation process was performed until the error was below 0.5° of visual angle. EyeLink software recorded gaze patterns, including fixations, saccades, and blinks, using its standard parser algorithms. The participants' gaze and pupil size data were recorded from the dominant eye, ensuring the capture of relevant eye movement metrics.

Data Analysis

Eye Tracking Analysis. For each session, participant, and movie separately, a PKL file included time-stamped X-Y gaze coordinates and pupil radius measurements. A custom validation tool was employed to ensure data integrity, running several checks such as for the correct number of files and recording rates (see code below). Movie trials with more than 30% tracking loss were excluded. Furthermore, all movies that were not shown in the first encoding session were excluded as well (12 animations & 20 naturalistic videos). The EyeLink 1000 Plus's classifications (Fixation, Saccades, Blinks) were used to structure the analysis. Data points were organized into multi-index pandas DataFrames for in-depth processing with custom-made Python scripts. Next, we computed the Euclidean gaze distance between each sample and the event location point in degrees of visual angle (DVA). DVA calculation was carried out according to the geometry of the setup according to the formula: DVA = 2 * arctan((object size/2) / viewing distance). Data quality was further ensured by observing the raw data by superimposing the instantaneous gaze and the average distance metric on the actual movies, as in Supp. Movies A1, A1, R1, R2.

Degrees of Visual Angle (DVA). To compute the Degrees of Visual Angle (DVA), we utilized a series of steps designed to precisely measure the angular distance between the participant's gaze point and the event location on the screen. First, we calculated the difference between the gaze point and the event location separately for the X and Y coordinates by subtracting the ROI's middle point coordinates from the gaze coordinates for each data point. Next, we computed the Euclidean distance in pixels between the gaze point and the event location using the Pythagorean theorem, which involved squaring the differences in X and Y coordinates,

summing these squares, and taking the square root of the result. To convert the pixel distances to degrees of visual angle, we used the following formula:

$$DVA = \frac{\text{distance_in_pixels} \times \left(\frac{\text{atan2}(0.5 \times \text{screen_size_in_cm}, \text{distance_between_screen_and_eyes}) \times 180}{\pi}\right)}{0.5 \times \text{resolution_in_pixels}}$$

For this conversion, we considered the specific dimensions of the screen (both width and height) and the distance between the screen and the participant's eyes, using parameters such as screen width and height in centimeters and the screen resolution in pixels. Finally, the DVA measures for the X and Y coordinates were combined using the Euclidean distance formula, ensuring that the angular distance accounted for both horizontal and vertical deviations. The resulting DVA values were then indexed and organized in the DataFrame to align with other metrics, such as subjects, sessions, movies, and timestamps, allowing seamless integration with other data processing steps and statistical analyses.

Gaze Average Distance (GAD). GAD quantifies the mean Euclidean distance between each gaze data point and the SE center point, calculated separately for each time point prior to the appearance of the SE and averaged across all computed distances (not just fixations). Thus, the GAD captures a cumulative estimate of how closely participants' gazes approximate the surprising event by reflecting the number of entries near the SE center point and the duration of each stay. GAD was averaged for each subject across all movies within each session.

MEGA Score: a metric for gaze anticipation. The MEGA Score is a normalized metric reflecting how anticipatory gaze behaviors change across repeated viewings of the same movie: MEGA Score = ($GAD^{1st\ viewing}$ - $GAD^{2nd\ viewing}$)\ max($GAD^{1st\ viewing}$, $GAD^{2nd\ viewing}$). A higher MEGA Score indicates that participants were looking closer to upcoming SEs during the second viewing, suggesting an enhanced anticipatory gaze indicative of effect once memory is present.

Pupillometry. Pupil size was extracted during all fixations prior to the surprising event using the EyeLink segmentation tool. As above, trials with >30% missing data were excluded. Pupil size dynamics comprised a time course prior to the events used to compute a normalized pupil size score (as for the MEGA score): pupil size score = (pupil size $^{1st\ viewing}$ - pupil size $^{2nd\ viewing}$) / max(pupil size $^{1st\ viewing}$, pupil size $^{2nd\ viewing}$). The pupil size score provides a nuanced measure of change in pupil size across repeated movie viewings, where a higher pupil size score reflects a larger pupil size in the 2nd viewing.

Behavioral analysis

Explicit report

To disentangle the relationship between anticipatory gaze and explicit report, we collected verbal reports in all four experiments. For Experiment 1, 3 and 4 we asked after each movie "*Have you seen this movie before?* (“Yes” or “No”)," and collected their confidence rating. In Experiment 2, five retrieval tasks were presented: i - movie recognition, ii - free recall, iii - object recognition, iv - location recall, v - temporal recall.

Anticipatory gaze and explicit report

For the anticipatory gaze analysis in relation to the explicit report, we computed the MEGA score for each movie. To estimate the difference between the first and the second viewing, we averaged the MEGA score over all movies for each participant ('new' movies were excluded) and then tested the MEGA score against chance. To estimate the subsequent memory effect, we computed the MEGA score according to the explicit report. That means, for example, we aggregated the MEGA score once for all movies that were recognized and all movies that were not recognized, according to the answer to the question "Have you seen this movie before?". Next, we compared the average recognized and unrecognized MEGA score (via the Wilcoxon signed-rank test and a binomial test, respectively). This was done for naturalistic and animated movies. Moreover, we used the same procedure to compare the sleep and the wake groups.

Anticipatory gaze and event-specific recollection. In Experiment 2, after extracting raw responses to each separate question (available in Supp. Figure 1), responses were combined into an integrated measure as follows:
Based on the four forced-choice tasks (i - recognition, iii - what, iv - where, v - when), and the defined three labels of remembered memory content: (A) Context and event recollection: movies where participants accurately recall the context and the surprising event. Namely, a correct answer for the movie recognition, object recognition, and the event location recall. (B) Context recognition: recognition of the movie's setting without specific recollection of the event. Namely, a correct answer for the movie recognition but an incorrect answer for the object recognition and spatial recall, and (C) not recognized: when participants did not recognize the movie at all. Namely, indicating that they do not recognize the movie, independent of the retrieval performance in object recognition and event location recall. Temporal recall was not further analyzed since retrieval performance was at chance (p=0.81, chance=50%, see Discussion). The free recall (ii) was analyzed separately. A movie was labeled as successfully recalled if a participant was able to recall the object of the surprising event before any visual cue of the object itself was presented.

Single-Trial Decoding with Machine Learning

Feature engineering for machine learning (ML) analysis. Our ML analysis relied on extracting and transforming raw eye-tracking data into a set of features. We calculated a total of 243 eye-tracking features. These features were broadly categorized into four groups: Fixation, Saccades, Blinks, and Pupil. Each category encapsulated different aspects of the participants' eye movements and responses. The features were primarily computed using each movie clip's time interval leading up to the event onset. Of these 243 features, 87 were related to the event: where the event was about to appear ("Event-Based Features"). These features leveraged the location and timing of the event, offering a more nuanced view of how participants' gaze interacted with specific elements of the visual stimuli. A comprehensive list and description of the Event-Based Features can be found in Supp. Table T4.

XGBoost Classifiers. ML employed an ensemble of XGBoost (Extreme Gradient Boosting) classifiers, operating through the construction of multiple decision trees in a gradient boosting framework, where each tree is sequentially built to correct the errors of its predecessors. The optimization of node splits and tree structure is driven by gradient statistics of a loss function, aiming to minimize prediction errors. Our implementation adhered to the standard specifications of XGBoost, as delineated by Chen and Guestrin[24]. The algorithm was

implemented by its application to each fold in our leave-one-subject-out (LOSO) cross-validation (CV) framework, ensuring robust and generalizable findings. LOSO iteratively trains the model on the dataset excluding data from one subject, which is then used as the test set. This approach mitigates potential biases and accounts for inter-individual variability, enhancing the model's applicability to new, unseen data. To optimize parameters, we used a grid search process to pinpoint optimal hyper-parameters within predefined ranges: 'learning_rate' (0.001-0.5), 'max_depth' (2-4), and 'n_estimators'(100-150). Final average hyperparameters used were 'learning_rate': 0.058, 'max_depth': 3.26, and 'n_estimators': 118.52, were instrumental in fine-tuning our models to achieve optimal performance.

Performance evaluation. We quantified model performance via several metrics: A) Classification Accuracy: the proportion of correctly predicted viewings, capturing the model's ability to accurately differentiate between 1st and 2nd viewings. B): Average Confusion Matrix: a more nuanced breakdown of model predictions separating true positives, true negatives, false positives, and false negatives. The average confusion matrix extends this analysis across all the models we created, providing precision and recall capabilities. C) Receiver Operating Characteristic (ROC) Curve: The ROC curve shows the trade-off between the true positive rate and false positive rate across different thresholds, with the Area Under the Curve (AUC-ROC) providing a summary measure of model performance. Finally, presentation of individual and average ROC curves for each model facilitates a deeper exploration of the model's sensitivity and specificity.

SHAP Analysis. Understanding the intricate decision-making processes of our ensemble, XGBoost classifiers necessitated the integration of SHAP (SHapley Additive exPlanations) analysis grounded in cooperative game theory[25], to interpret the contribution of each feature to individual predictions. SHAP analysis was implemented in Python 3.9 Anaconda, with the shap package version 0.42.1.

Statistical analysis

The error probability of 5% was chosen for all statistical tests. For hypothesis testing, we used mainly nonparametric statistics. Most data distributions approximate normality, suggesting that parametric tests could be adequate; however, to ensure methodological consistency across all comparisons, we opted for the Wilcoxon signed-rank test for testing whether the GAD values were different between first and second viewings (Figures 2C, 4A, 4B, 4D, 4F, 4G, 5C, 5D, and). In our analysis of gaze anticipatory scores (Figures 4B, 5D, and 6C), we tested the statistical significance of the observed changes between each subject's first and second viewings using a binomial test. To compare wake and sleep data, the MEGA score was compared between two groups (Figure 6B) using the Wilcoxon rank-sum statistic for two independent samples of unequal sizes. The statistical significance of ML classification was quantified by applying a one-sample t-test to the ROC AUC score compared to the chance level (50%). Lastly, we employed Cohen's d (Cohen, 1988) to assess the magnitude of effects observed in the study. Behavioral data were analyzed in R (https://www.r-project.org/, version 2022.07.0) and Python 3.9 Anaconda, with Scipy version 1.7.1. The sleep EEG data and eye tracking analyses were conducted using Python 3.9 Anaconda.

Acknowledgments

We extend our deepest gratitude to Dr. Noa Regev for her administrative assistance and unwavering support throughout this project. We are indebted to Yuval Shapira for their initial analyses and for developing a user interface that enabled participants to accurately select event coordinates in natural movies. Special thanks go to Odessa Goldberg, Alexandra Klein, Yael Gat, Rotem Falach, Or Ra'anan, and the additional research assistants for their invaluable help with data collection. We also want to express our heartfelt thanks to the participants of this study, whose involvement was crucial to the success of our research. Additionally, we are grateful to Studio Plonter ® for their exceptional work in creating the tailor-made animations that played a pivotal role in our experiments. We are also thankful to Dr. Roy Amit for his advice on eye-tracking data collection and preprocessing in the initial stages of the project.

This research was supported by ERC-2019-CoG 864353 (Y.N.). Additionally, Flavio J. Schmidig was supported by a Tel Aviv University Sagol School of Neuroscience Postdoctoral Fellowship.

Declaration of interests

The authors declare no competing interests.

Author contributions

- Conception and Design of Research: D. Yamin, O. Sharon, F.J. Schmidig, and Y. Nir.
- Funding Acquisition: Y. Nir and O. Sharon.
- Compilation of Movie Stimuli: O. Sharon and D. Yamin.
- Data Collection: D. Yamin, F.J. Schmidig, O. Sharon, and Y. Nadu.
- Data Analysis: D. Yamin, F.J. Schmidig, O. Sharon, and J. Nir.
- Manuscript Writing: D. Yamin, F.J. Schmidig, and Y. Nir.
- Advice on Experimental Design and Data Analysis: C. Ranganath.
- Critical Review of Results and Manuscript Comments: All authors.

Data availability

Source data underlying results presented in figures will be made publicly available before publication.

Code and stimuli availability

The code used to preprocess and analyze the gaze data in this study is available at: https://github.com/dyamin/MEGA

Specific code for detailed analysis is available from the authors upon reasonable request.

The experiment code is available at: https://github.com/dyamin/MEGA-Experiment

The movie clips used in the experiments will be made publicly available before publication at: https://yuvalnirlab.com/

References

1. Gardiner, J. M. Functional aspects of recollective experience. *Mem. Cognit.* **16**, 309–313 (1988).
2. Migo, E. M., Mayes, A. R. & Montaldi, D. Measuring recollection and familiarity: Improving the remember/know procedure. *Conscious. Cogn. Int. J.* **21**, 1435–1455 (2012).
3. Tulving, E. Memory and consciousness. *Can. Psychol. Psychol. Can.* **26**, 1–12 (1985).
4. Tsuchiya, N., Wilke, M., Frässle, S. & Lamme, V. A. F. No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends Cogn. Sci.* **19**, 757–770 (2015).
5. Morris, R. Developments of a water-maze procedure for studying spatial learning in the rat. *J. Neurosci. Methods* **11**, 47–60 (1984).
6. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
7. Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H. & Lakatos, P. Dynamics of Active Sensing and perceptual selection. *Curr. Opin. Neurobiol.* **20**, 172–176 (2010).
8. Henderson, J. M. & Hollingworth, A. Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Percept. Psychophys.* **65**, 58–71 (2003).

9. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. *Trends Cogn. Sci.* **9**, 188–194 (2005).
10. Kano, F. & Hirata, S. Great Apes Make Anticipatory Looks Based on Long-Term Memory of Single Events. *Curr. Biol.* **25**, 2513–2517 (2015).
11. Johansson, R., Nyström, M., Dewhurst, R. & Johansson, M. Eye-movement replay supports episodic remembering. *Proc. Biol. Sci.* **289**, 20220964 (2022).
12. Hannula, D. E. *et al.* Worth a Glance: Using Eye Movements to Investigate the Cognitive Neuroscience of Memory. *Front. Hum. Neurosci.* **4**, (2010).
13. Ryan, J. D. & Shen, K. The eyes are a window into memory. *Curr. Opin. Behav. Sci.* **32**, 1–6 (2020).
14. Hannula, D. E. & Ranganath, C. The eyes have it: hippocampal activity predicts expression of memory in eye movements. *Neuron* **63**, 592–599 (2009).
15. Olsen, R. K. *et al.* The relationship between eye movements and subsequent recognition: Evidence from individual differences and amnesia. *Cortex J. Devoted Study Nerv. Syst. Behav.* **85**, 182–193 (2016).
16. Sharot, T., Davidson, M. L., Carson, M. M. & Phelps, E. A. Eye Movements Predict Recollective Experience. *PLOS ONE* **3**, e2884 (2008).
17. Urgolites, Z. J., Smith, C. N. & Squire, L. R. Eye movements support the link between conscious memory and medial temporal lobe function. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7599–7604 (2018).
18. Haque, R. U. *et al.* VisMET: a passive, efficient, and sensitive assessment of visuospatial memory in healthy aging, mild cognitive impairment, and Alzheimer's disease. *Learn. Mem. Cold Spring Harb. N* **26**, 93–100 (2019).
19. Heisz, J. J. & Ryan, J. D. The effects of prior exposure on face processing in younger and older adults. *Front. Aging Neurosci.* **3**, 15 (2011).
20. Cannon, E. N. & Woodward, A. L. Infants generate goal-based action predictions. *Dev. Sci.* **15**, 292–298 (2012).

21. Wong-Kee-You, A. M. B. & Adler, S. A. Anticipatory eye movements and long-term memory in early infancy. *Dev. Psychobiol.* **58**, 841–851 (2016).
22. Rovee-Collier, C. The Development of Infant Memory. *Curr. Dir. Psychol. Sci.* **8**, 80–85 (1999).
23. Aly, M. & Turk-Browne, N. B. Attention promotes episodic encoding by stabilizing hippocampal representations. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E420-429 (2016).
24. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in 785–794 (2016). doi:10.1145/2939672.2939785.
25. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).
26. Kragel, J. E. & Voss, J. L. Looking for the neural basis of memory. *Trends Cogn. Sci.* **26**, 53–65 (2022).
27. Goldinger, S. D. & Papesh, M. H. Pupil dilation reflects the creation and retrieval of memories. *Curr. Dir. Psychol. Sci.* **21**, 90–95 (2012).
28. Kucewicz, M. T. *et al.* Pupil size reflects successful encoding and recall of memory in humans. *Sci. Rep.* **8**, 4949 (2018).
29. Kafkas, A. & Montaldi, D. Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched. *Neuropsychologia* **50**, 3080–3093 (2012).
30. Papesh, M. H., Goldinger, S. D. & Hout, M. C. Memory strength and specificity revealed by pupillometry. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **83**, 56–64 (2012).
31. Võ, M. L.-H. *et al.* The coupling of emotion and cognition in the eye: introducing the pupil old/new effect. *Psychophysiology* **45**, 130–140 (2008).
32. Geva-Sagiv, M. *et al.* Augmenting hippocampal–prefrontal neuronal synchrony during sleep enhances memory consolidation in humans. *Nat. Neurosci.* **26**, 1100–1110 (2023).
33. Rasch, B. & Born, J. About sleep’s role in memory. *Physiol. Rev.* **93**, 681–766 (2013).

34. Schmidig, F. J. *et al.* A visual paired associate learning (vPAL) paradigm to study memory consolidation during sleep. *J. Sleep Res.* e14151 (2024) doi:10.1111/jsr.14151.
35. Ryan, J. D., Althoff, R. R., Whitlow, S. & Cohen, N. J. Amnesia is a deficit in relational memory. *Psychol. Sci.* **11**, 454–461 (2000).
36. Ferreira, F., Apel, J. & Henderson, J. M. Taking a new look at looking at nothing. *Trends Cogn. Sci.* **12**, 405–410 (2008).
37. Wynn, J. S., Shen, K. & Ryan, J. D. Eye Movements Actively Reinstate Spatiotemporal Mnemonic Content. *Vis. Basel Switz.* **3**, 21 (2019).
38. Brockmole, J. R. & Henderson, J. M. Using real-world scenes as contextual cues for search. *Vis. Cogn.* **13**, 99–108 (2006).
39. Chau, V. L., Murphy, E. F., Rosenbaum, R. S., Ryan, J. D. & Hoffman, K. L. A flicker change detection task reveals object-in-scene memory across species. *Front. Behav. Neurosci.* **5**, (2011).
40. Le-Hoa Võ, M. & Wolfe, J. M. The role of memory for visual search in scenes. *Ann. N. Y. Acad. Sci.* **1339**, 72–81 (2015).
41. Schwedes, C. & Wentura, D. The revealing glance: Eye gaze behavior to concealed information. *Mem. Cognit.* **40**, 642–651 (2012).
42. Hannula, D. E., Baym, C. L., Warren, D. E. & Cohen, N. J. The Eyes Know: Eye Movements as a Veridical Index of Memory. *Psychol. Sci.* **23**, 278–287 (2012).
43. Wuethrich, S., Hannula, D. E., Mast, F. W. & Henke, K. Subliminal encoding and flexible retrieval of objects in scenes. *Hippocampus* **28**, 633–643 (2018).
44. Henke, K. A model for memory systems based on processing modes rather than consciousness. *Nat. Rev. Neurosci.* **11**, 523–532 (2010).
45. Schwedes, C. & Wentura, D. Through the eyes to memory: Fixation durations as an early indirect index of concealed knowledge. *Mem. Cognit.* **44**, 1244–1258 (2016).
46. Mahoney, E. J., Kapur, N., Osmon, D. C. & Hannula, D. E. Eye Tracking as a Tool for the Detection of Simulated Memory Impairment. *J. Appl. Res. Mem. Cogn.* **7**, 441–453 (2018).

47. Ryals, A. J., Wang, J. X., Polnaszek, K. L. & Voss, J. L. Hippocampal contribution to implicit configuration memory expressed via eye movements during scene exploration. *Hippocampus* **25**, 1028–1041 (2015).

48. Bone, M. B. *et al.* Eye Movement Reinstatement and Neural Reactivation During Mental Imagery. *Cereb. Cortex N. Y. N 1991* **29**, 1075–1089 (2019).

49. Antón-Méndez, I., Talk, A. & Johnston, S. Gaze direction reveals implicit item and source memory in older adults. *PloS One* **14**, e0226018 (2019).

50. Wynn, J. S., Buchsbaum, B. R. & Ryan, J. D. Encoding and retrieval eye movements mediate age differences in pattern completion. *Cognition* **214**, 104746 (2021).

51. Ryan, J. D., Wynn, J. S., Shen, K. & Liu, Z.-X. Aging changes the interactions between the oculomotor and memory systems. *Aging Neuropsychol. Cogn.* **29**, 418–442 (2022).

52. Folstein, M. F., Folstein, S. E. & McHugh, P. R. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).

53. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).

54. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).

55. Robinson, G., Blair, J. & Cipolotti, L. Dynamic aphasia: An inability to select between competing verbal responses? *Brain J. Neurol.* **121**, 77–89 (1998).

56. van der Meulen, I., van de Sandt-Koenderman, W. M. E., Duivenvoorden, H. J. & Ribbers, G. M. Measuring verbal and non-verbal communication in aphasia: reliability, validity, and sensitivity to change of the Scenario Test. *Int. J. Lang. Commun. Disord.* **45**, 424–435 (2010).

57. Squire, L. R. & Zola, S. M. Structure and function of declarative and nondeclarative memory systems. *Proc. Natl. Acad. Sci.* **93**, 13515–13522 (1996).

58. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).

59. Smith, C. N., Hopkins, R. O. & Squire, L. R. Experience-Dependent Eye Movements, Awareness, and Hippocampus-Dependent Memory. *J. Neurosci.* **26**, 11304–11312 (2006).
60. Smith, C. N. & Squire, L. R. Experience-Dependent Eye Movements Reflect Hippocampus-Dependent (Aware) Memory. *J. Neurosci.* **28**, 12825–12833 (2008).
61. Smith, C. N. & Squire, L. R. When eye movements express memory for old and new scenes in the absence of awareness and independent of hippocampus. *Learn. Mem.* **24**, 95–103 (2017).
62. Sharon, O., Fahoum, F. & Nir, Y. Transcutaneous Vagus Nerve Stimulation in Humans Induces Pupil Dilation and Attenuates Alpha Oscillations. *J. Neurosci.* **41**, 320–330 (2021).
63. DELLA PORTA, Giovan Battista (c.1538-1615). De refractione optices parte: libri novem. Naples: Horatius Salvianus for Joannes Jacobus Carlinus and Antonio Pace, 1593. | Christie's. https://www.christies.com/en/lot/lot-6069542.
64. Gronwall, D. M. & Sampson, H. Ocular dominance: A test of two hypotheses. *Br. J. Psychol.* **62**, 175–185 (1971).