Forgone, but not forgotten: separate episodic memories underlie explicit reports and eye movements

Yul HR Kang* (yul.hr.kang@gmail.com)
Department of Biological and Behavioural Sciences, Queen Mary University of London
London, E1 4NS United Kingdom
Department of Engineering, University of Cambridge
Cambridge, CB2 1PZ United Kingdom

Johannes Mahr* (jmahr@fas.harvard.edu)
Department of Psychology, Harvard University
Cambridge, MA 02138 USA
Department of Philosophy, York University
Toronto, ON M3J 1P3 Canada

Márton Nagy
Department of Cognitive Psychology & MTA-ELTE Social Minds Research Group, Eötvös Loránd University, Egyetem tér 1-3
Budapest, 1053 Hungary

Krisztina Andrási
Department of Cognitive Psychology & MTA-ELTE Social Minds Research Group, Eötvös Loránd University, Egyetem tér 1-3
Budapest, 1053 Hungary

Gergely Csibra[†]
Department of Cognitive Science, Central European University, Quellenstraße 51
Wien 1040, Austria
Department of Psychological Sciences, Birkbeck, University of London,
London, WC1E 7HX United Kingdom

Máté Lengyel[†]
Department of Engineering, University of Cambridge
Cambridge, CB2 1PZ United Kingdom
Department of Cognitive Science, Central European University, Nádor u. 9
Budapest, 1051 Hungary

*[†] equal contribution

Abstract

Episodic memories are thought to integrate multiple aspects of a past experience into a unified engram. A key prediction is that such integrated memories should drive behavior consistently across different response modalities (e.g., explicit/implicit). This however, has remained largely untested, as previous approaches couldn't resolve the content of memories underlying implicit responses. Here we used ideal observer-based trial-by-trial/gaze-by-gaze analyses of explicit memory reports and spontaneous eye movements to reconstruct participants' memories underlying different response modalities. We used a false memory paradigm, where human participants studied a sequence of object-location pairings, followed by suggestions (50% false). After participants judged the correctness of all suggestions, they recalled each object's location while their gazes were recorded. Our analyses revealed that explicit recall reflected 'conditioned memory': the recalled location reflected only either the suggested or the original location, depending on whether the suggestion had been deemed correct. In contrast, eye movements did not show this effect, and reflected a combination of memories distinct from those underlying explicit recall. Challenging the notion of unified engrams, these results suggest the formation of multiple distinct memory traces of the same experience that are maintained independently, show different vulnerability to false suggestions, and control different response modalities.

Keywords: episodic memory; ideal observer; Bayesian modeling; causal inference; eye movements

Eye movements have been suggested to provide a unique window into memory processes that evade instructed, or explicit, reports (Ryan & Shen, 2020; Lancry-Dayan, Ben-Shakhar, & Pertzov, 2022). Here, building on research on "retrieval-dependent eye movements" (RDEs; Richardson & Spivey, 2000; Johansson, Holsanova, Dewhurst, & Holmqvist, 2012; Staudte & Altmann, 2017), we examined explicit memory reports and eye movements to study the representations and mechanisms underlying episodic memory retrieval.

We used a location memory task to compare how RDEs are affected by a memorized object location and a subsequent suggestion. In the "learning phase", participants were presented with a series of object-location pairs on a circular array and were asked to remember the location of each object for a later memory test ("*L*"earned location; Figure 1A). Next, in the "suggestion phase", participants were again presented with a series of (50% correct) object-location pairs ("*S*"uggested location; allegedly produced by another participant in the upcoming memory test) and were asked to make a binary judgment about the correctness of each pair ("*D*"ecision). Finally, in the "response phase", participants were sequentially presented with a subset of the objects from the learning phase at the center of the screen. For each of these objects, participants were asked to indicate its original location on the circular array ("*R*"esponed location).

Explicit reports

To understand the mechanisms determining the explicit reports given two sources of information (*L* & *S*, where *S* is the same as *L* or a random different location, depending on whether it is "*C*"orrect, i.e. $C = 1$ or 0), we fit and compared two models that perform causal inference about whether these two sources of information derive from the same cause (Shams & Beierholm, 2010). First, the optimal cue combination model (Figure 1B without dark red arrow), optimally combines noisy memories $\tilde{L}^{\mathrm{R}}$ and $\tilde{S}^{\mathrm{R}}$ of *L* and *S* to produce the reports *D* (reflecting inferences about *C*) and *R* (reflecting inferences about *L*) by performing Bayesian inference (Körding et al., 2007). Second, in the conditioned memory model (Figure 1B with dark red arrow), *D* is produced as in the optimal model but (in analogy to models of conditioned perception; Stocker & Simoncelli, 2008) *R* is determined considering only the possibility that the judgment *D* was veridical (i.e., conditioning on $C = D$; dark red arrow from *D* to *R*). We used statistically principled maximum-likelihood to fit both memory models to the trial-by-trial reports (R, D) of individual participants given the stimuli (L, S), and compared the two models with Bayesian Information Criterion (BIC). Using synthetic data, we extensively validated our model fitting and comparison approach, and confirmed that the models were identifiable (see e.g., Figure 1E, left).

Both the optimal and the conditioned memory models gave comparable fits to the probability of judging the suggestion correct (Figure 1C, Row 1). However, their predictions about the pattern of responded locations differed markedly (Figure 1C, Row 2). The optimal model (Figure 1C, Column 2) predicts that when the suggestion is deemed correct (orange line), the average report falls between the suggested (diagonal line) and the learned locations (horizontal line), and the learned location is followed more closely when the suggested location is farther away from the learned location (Figure 1C, Row 3, Column 2, orange line tends downward towards the extremes). This is because the optimal model weights the hypothesis that the suggestion comes from the same source as the learned location by its probability, which is lower when $|S - L|$ is large. In contrast, the conditioned memory model predicts that the report follows the suggestion when it is deemed correct, and even more so when the suggested location is farther away from the learned location (Figure 1C, Row 3, Column 3: orange line tends upward towards the extremes), which matches the data better (orange markers, reproduced from Column 1). Bayesian model comparison provided overwhelming support for the conditioned memory model (Figure 1E, Row 1, Column 2; $\Delta\mathrm{BIC} > 10$). Thus, these results suggest that participants first performed causal inference based on memories $\tilde{L}^{\mathrm{R}}$ & $\tilde{S}^{\mathrm{R}}$, and made the explicit report *R* ignoring the possibility that their decision *D* might be wrong. Hence, from explicit reports alone, it appears as though the memory of *S*, once forgone, is forgotten.

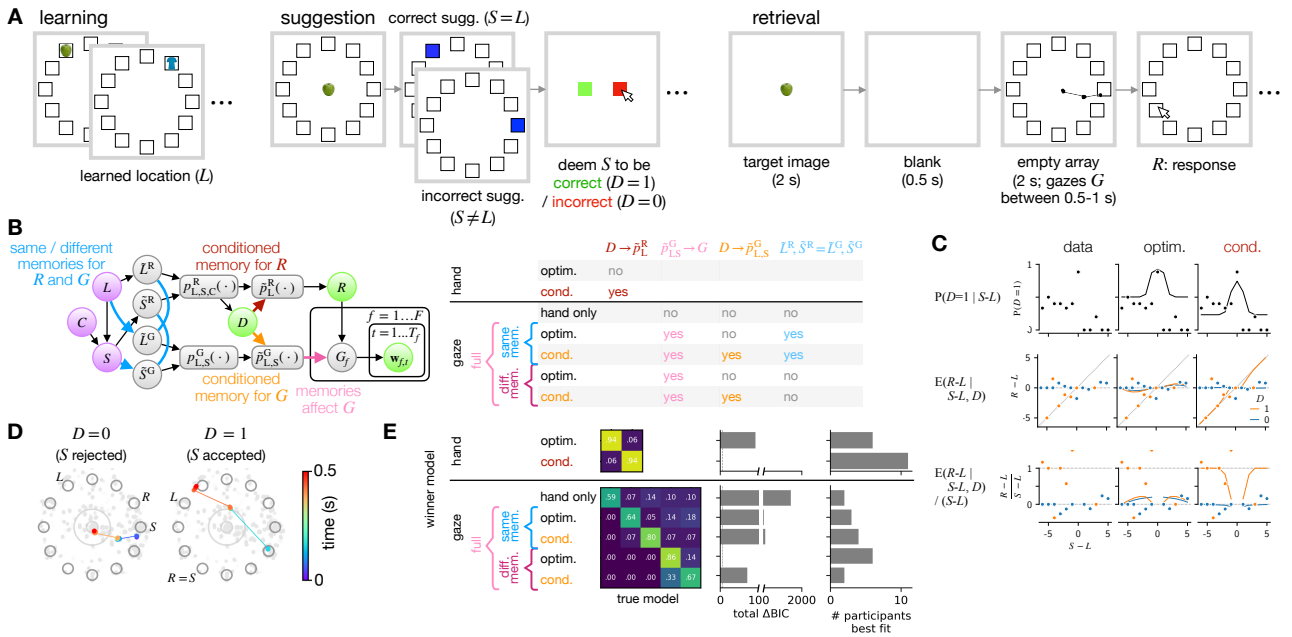

Figure 1. **A**. Behavioral task. **B**. Generative model of hand (R) and gaze responses (G_f). The suggestion (S) is either correct ($C=1$) such that it matches the learned location ($S=L$) or wrong ($C=0 \rightarrow S \neq L$). The memories of L and S are corrupted by noise, and are shared or distinct across the pathways controlling R and G ($\tilde{L}/\tilde{S}^{R/G}$; blue lines/arrows). Based on these memories, the observer infers the posterior distributions of L, S, and C ($p^{R/G}$). After the decision about the suggestion (D), these posteriors may be conditioned on the assumption that $C=D$ ($\tilde{p}^{R/G}$; red/orange arrows). R is generated based on $\tilde{p}^{R}$, and each fixation G_f is generated based on R, and possibly also $\tilde{p}^{G}$ (pink arrow). Gaze locations $\mathbf{w}_{f,t}$ corresponding to the same fixation f are sampled from a G_f-specific 2D Gaussian (one of 12 peripheral and 2 central gray open ellipses in D). Purple: known to the experimenter; gray: known to the participant; green: known to both. Arrows/lines with color: present only in a subset of models (see table). **C**. Explicit reports of an example participant and model fits. Data (Column 1 & markers) and the fit of the optimal (Column 2, curves) and the conditioned memory models (Column 3, curves) to the data. Models are fit to the full distribution of reports; mean is shown for visualization only. *Top:* probability of deeming the suggestion correct. given $S-L$. *Middle:* circular mean of $R-L$ given $S-L$, when S was deemed correct vs. wrong ($D=0/1$, blue/orange). Dotted lines indicate $R=S$ (diagonal) and $R=L$ (horizontal). *Bottom:* same as middle row, but normalized by $S-L$ (undefined when $S=L$). Dotted lines indicate $R=S$ (top) and $R=L$ (horizontal). **D.** Example trials of the same participant when $D=0/1$ (left/right). Color indicates within-trial time, gray dots are gazes from other trials, ellipses are identified clusters. **E.** Model recovery & comparison. Top and bottom rows show hand and eye models, respectively. *Left:* model recovery P(true model | winner model). *Middle:* Total BIC of each model relative to the winner model. *Right:* number of participants for whom each model was the winner.

Implicit responses

We then analyzed eye movements during the response phase to determine whether they reveal memories about L and S beyond what is already conveyed by explicit hand reports. We first constructed two classes of models to analyze eye movements gaze-by-gaze (Figure 1B). Either class assumes that gaze locations form 14 clusters (Figure 1D, open ellipses), and that each gaze location ($\mathbf{w}$) is sampled from a 2D Gaussian distribution associated with one of the clusters (G). The two models differ only in what determines G. The "hand only" model (Figure 1B without pink arrow) assumes that the explicit report R determines which cluster (G) the gaze is assigned to. In contrast, the "full" model (Figure 1B with pink arrow) assumes that both the explicit reports and the memory of L and S affect the gazes. Consistent with the full model, we found trials where the gaze was clearly directed to S even when it was rejected, and to L instead of S even when S was accepted (Figure 1D). Fitting the models revealed that gaze locations (their cluster assignment, G), after subtracting the influence of R, were biased toward the direction of L and S, consistent with the full model, but not with the hand-only model (not shown). Bayesian model comparison decisively supported the full gaze model (Figure 1E, 1st versus all other models). We further discriminated between 2×2 variants of the full model, depending on whether they performed optimal cue combination or conditioned memory (Figure 1B, orange arrow), and whether the effects of L & S on eye-movements manifested via the same or different memories as that underlying explicit recall (Figure 1B, blue lines vs. arrows). We found overwhelming evidence for separate memories underlying eye movements with optimal cue combination (Figure 1E, penultimate model). Therefore, our results demonstrate that there is a separate memory of S that is not forgotten even when it is explicitly forgone.

References

Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2012). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1289. doi: 10.1037/a0026585

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), e943. doi: 10.1371/journal.pone.0000943

Lancry-Dayan, O. C., Ben-Shakhar, G., & Pertzov, Y. (2022). The promise of eye-tracking in the detection of concealed memories. *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2022.08.019

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: looking at things that aren't there anymore. *Cognition*, *76*, 269–295. doi: 10.1016/s0010-0277(00)00084-6

Ryan, J. D., & Shen, K. (2020). The eyes are a window into memory. *Current Opinion in Behavioral Sciences*, *32*, 1–6. doi: 10.1016/j.cobeha.2019.12.014

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432. doi: 10.1016/j.tics.2010.07.001

Staudte, M., & Altmann, G. T. (2017). Recalling what was where when seeing nothing there. *Psychonomic Bulletin & Review*, *24*, 400–407. doi: 10.3758/s13423-016-1104-8

Stocker, A. A., & Simoncelli, P. (2008). A Bayesian Model of Conditioned Perception. In J. C. Platt, D. Koller, Y. Singer, & S. T. Roweis (Eds.), (pp. 1409–1416). Howard Hughes Medical Institute.