Supplementary Information

A smartphone intervention that enhances real-world memory and promotes differentiation of hippocampal activity in older adults

Chris B. Martin, Bryan Hong, Rachel N. Newsome, Katarina Savel, Melissa E. Meade, Andrew Xia, Christopher J. Honey, Morgan D. Barense

Table of Contents

1. Behavioral Methods

1.1 Autobiographical Memory Tests

1.1.1 Administration

We assessed memory performance in Experiment 1 and Experiment 2 using an adapted version of the autobiographical interview (AI) (1). Typical administration of the AI involves asking participants to verbally describe their memory for self-selected events from five different life periods. Following a free-recall response in the typical AI, participants respond to a general probe intended to encourage retrieval of additional information without providing specific guidance. They are then asked to respond to a series of specific probes designed to elicit further detail corresponding to events, time, time integration, place, sensory information, emotions, and thoughts. In our adapted version of the AI, participants were instructed to recall events for which they had previously recorded an autobiographical cue. Memory for each event was tested by first

having the participant view the event-specific cue and then provide a verbal description of what they remembered based on the following instructions:

"Please tell me as much as you can remember about the event that was just cued. Try to restrict your description to the specific event that was the subject of the recording. It is important that your description goes beyond the contents of the video and verbal description that you recorded. In other words, do not simply narrate the video, tell me about the entirety of the event."

The length of event-specific verbal descriptions elicited by this approach varied across events and participants. Early testing cued-recall responses ranged from 15 seconds to 20 minutes per event. Given that we were probing memory for up to 40 distinct events, rather than five as is typical in the AI, we did not consistently administer general or specific probes. As such, all data described in this report reflect details described in initial cued-recall responses.

1.1.2 Scoring

All raters successfully completed the AI training protocol (1). Cued-recall responses were quantified using an adapted version of the AI scoring protocol that specifies criteria for distinguishing between Internal and External details, as well as subtypes of Internal and External details (1, 2). Our adaptation further quantified In-Cue details.

Internal Detail Subtypes

- *Internal Event* details: recall of information that is episodic in nature and pertains to the cued event. This includes happenings, individuals present, weather conditions, physical/emotional actions, or reactions in others.
- *Internal Time* details: year, season, month, day of week, time of day.
- *Internal Place* details: localization of an event including the city, street, building, room, part of room.
- *Internal Perceptual* details: auditory, olfactory, tactile, taste, visual and visual details, body position, duration.

- *Internal Thought/Emotion* details: emotional state, thoughts, implications.

External Detail Subtypes

- *External Event* details: recall of information that is episodic in nature but pertains to non-target events.
- *General Semantic* details: culturally shared knowledge of facts, public events, people, and concepts.
- *Personal Semantic* details: semantic knowledge of one's personal past, which is further divided into three subtypes.
    - *Autobiographical Facts*: basic units of knowledge about one's personal past, such as the name of the street on which you lived as a child, or your dog's name.
    - *Self-Knowledge*: awareness of one's disposition and preferences, such as claiming that you are hot tempered or noting that you dislike spicy food.
    - *Repeated Events*: descriptions of features that are common to multiple instances of an episode, such as mentioning that you and a friend always order the same cobbler for dessert when you get together for dinner on Saturdays.
- *Repeated Details*: information provided multiple times while describing memory for a specific event.
- *Other*: utterances that could not otherwise be scored as Internal or any of the above External subtypes (e.g., "Give me a moment to think about this", "I don't remember this event very well").

In-Cue Details

- In-Cue details refer to verbal descriptions that capture information that was apparent in the cue from all trial-specific detail counts. These details were omitted from Internal and External detail counts. For example, if a cue included video footage of a dinner partner who was wearing a red dress and eating chicken, any

mention of the red dress or specific comments about the chicken dish were not counted toward Internal and External detail counts. This strategy ensured that our dependent measure reflected episodic reliving rather than mere recall of content from the multi-modal cue.

1.1.3 Stimulus Selection

Anecdotally, the quality of cues varied within and across participants and experiments. Some cues corresponded to interesting and complex events (e.g., attending an outdoor concert at a summer festival), whereas others captured more mundane moments from daily life (e.g., preparing lunch). On average, participants in Experiment 1 recorded a total of 66.9 cues (33.8 were randomly assigned to the Replayed condition and 33.1 to the Baseline condition) over their 14-day use period. Participants in Experiment 2 recorded a total of 66.75 cues (33.9 in the Replayed condition and 32.9 in the Baseline condition) over their 70-day use period. Because of the extensive nature of our autobiographical memory tests, we selected up to 40 of these cues for the purpose of constructing our behavioral and fMRI assessments (up to 20 Replayed and up to 20 Baseline). In some cases, fewer than 40 cues were selected based on three criteria. First, across conditions we aimed to match the events for which we were testing memory at the level of event frequency, event significance, and memory age. Second, we selected cues that sampled broadly from the range of recorded behaviors and events to ensure that we were probing memory for experiences that were representative of our participants' lives. Cues were excluded from testing if they received low significance ratings (i.e., a rating of 1 or 2 out of 5). Third, in Experiment 1, multiple cues that obviously corresponded to the same event (e.g., a birthday party with one cue of the cake and another of presents being opened), but were randomly assigned to different conditions (i.e., one cue was Replayed whereas the other was Baseline), were not tested in either the behavioral or fMRI experiments.

Using this approach, we successfully matched Experiment 1 events across the Replayed and Baseline conditions at the level of event frequency ($M_{Replayed}$ = 2.52 on a 5-point scale where 1 corresponds to unique events and 5 corresponds to daily events, $M_{Baseline}$ = 2.60; paired $t(21) = 0.72$, P = .48) and event significance ($M_{Replayed}$ = 3.13 on a

5-point scale where 1 corresponds to the least significant events and 5 corresponds to the most significant events, $M_{Baseline}$ = 3.13; paired $t(21) = 0.01$, P = .99). There was a statistically significant difference between Replayed and Baseline events at the level of memory age, such that Replayed memories were older (Early M_{age}: Replayed = 4.3 days ± .16 SEM, Baseline = 4.0 days ± .20 SEM, paired $t(20) = 2.49$, P = .02, $d = 0.54$; Delayed M_{age}: Replayed = 123.8 days ± 4.92 SEM, Baseline = 123.4 days ± 4.90 SEM, paired $t(18) = 2.17$, P = .04, $d = 0.50$). We note, however, that this difference was rather small in practical terms and that its direction (Replayed older than Baseline) works against our hypotheses regarding enhanced memory in the Replayed condition.

In Experiment 2, Replayed and Baseline memories were matched at the levels of event significance ($M_{Replayed}$ = 3.32 on a 5-point scale, $M_{Baseline}$ = 3.24; paired $t(11) = 1.08$, P = .30), event frequency ($M_{Replayed}$ = 2.52 on a 5-point scale, $M_{Baseline}$ = 2.28; paired $t(11) = 2.01$, P = .06), and memory age (Early M_{age}: Replayed = 43.35 days ± 2.31 SEM, Baseline = 44.98 days ± 1.93 SEM, paired $t(11) = 0.54$, P = 0.59; Delayed M_{age}: Replayed = 144.55 days ± 4.51 SEM, Baseline = 146.16 days ± 4.11 SEM, paired $t(11) = 0.26$, P = 0.79).

1.2 Linear Mixed Modeling with Autobiographical Memory Test Data

Detail counts from the autobiographical memory tests were analyzed using multilevel modeling in R 4.1.0 (R Core Team, 2021) with the *lme4* package (3). The *performance* package was used to obtain intraclass correlations (ICC) and both conditional and marginal coefficients of determination (R^2_C and R^2_M, respectively) to determine model fits (4, 5). To test hypotheses related to autobiographical memory test detail counts, we used 2-level multilevel generalized Poisson models with individual trials nested within participants. Poisson models were used to best account for the count-based nature of the number of details recollected (6). For both Experiment 1 (2-Week Intervention) and Experiment 2 (10-Week Intervention), individual models were specified for Internal details, External details, and In-Cue details to obtain separate estimates for each detail type. Multilevel modeling was appropriate given the degree of variance explained by individual participants in Experiment 1 and Experiment 2 (Table S1). For all models, we first fit the model with the maximal random effects structure (7). To investigate

the effect of condition and test time point, this entailed estimating fixed effects for condition (Replayed vs. Baseline), test time point (Early vs. Delayed), and their interaction, as well as a random intercept estimated for each participant and a random slope estimated for each fixed effect. condition and test session were effect coded (condition: Baseline = -1, Replayed = 1; test time point: Early = -1, Delayed = 1). In the situation that the maximal model failed to converge due to overparameterization, we employed a backward-selection heuristic (8). To probe for any significant interactions, simple effects tests were conducted with the Tukey adjustment for pairwise comparisons using the *emmeans* package (9)—adjusted *P*-values are reported for simple effects. To compare the total number of details recalled across both experiments, the maximal model predicting the total number of details recalled on a trial with a fixed effect for Experiment (Experiment 1 vs Experiment 2), a random slope for Experiment, and a random intercept for each participant was first specified—Experiment was effect coded (Experiment 1 = -1, Experiment 2 = 1). Model selection was performed in the same fashion as described above. All models were estimated with an unstructured covariance matrix.

2. Behavioral Results

Overall Internal, External, and In-Cue detail counts are plotted in Fig. S1 to allow for direct comparison. Model fits for overall details of all types are reported in Table S1. Statistical analyses pertaining to overall Internal detail counts are reported in the main text of our manuscript (Fig. 2).

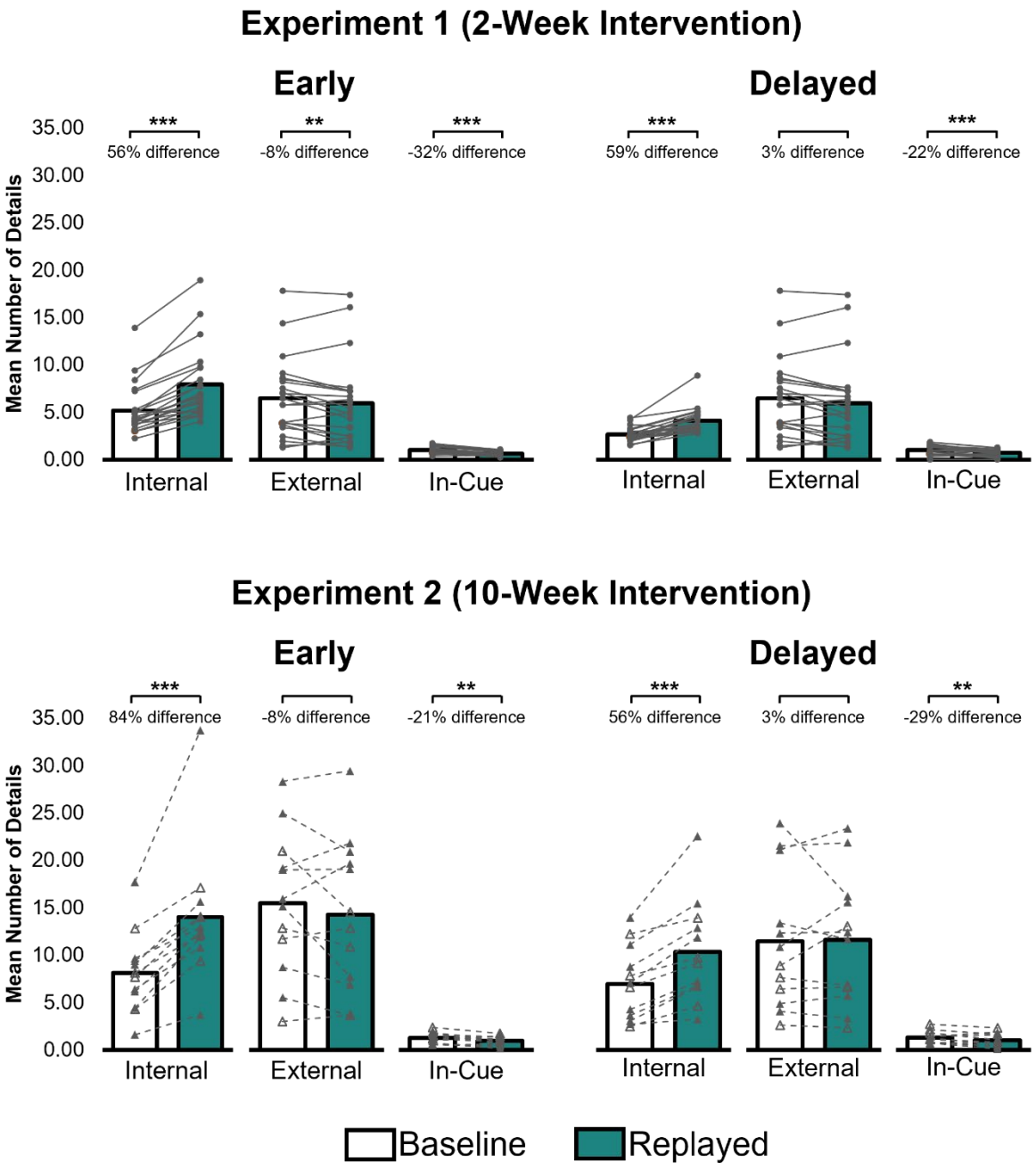

Fig. S1. Behavioral Results – Total Detail Counts. Mean number of Internal, External, and In-Cue details for Experiment 1 and Experiment 2. Early testing corresponds to behavioral performance measured during (Experiment 1) and shortly after (Experiment 2) HippoCamera use. Delayed testing corresponds to behavioral performance after a 3-month delay, during which time participants did not have access to their memory cues. Open triangles denote Experiment 2 participants who failed the Montreal Cognitive Assessment. * < .05, ** < .01, *** < .001.

2.1 Autobiographical Memory Test: Total External Details

We performed exploratory analyses focused on total External details (Fig. S1, Table S1-2). For Experiment 1, we found a significant main effect of condition (Replayed vs. Baseline: b = -0.0335, SE = 0.0171, z = -1.964, *P* = .049) but not test time point (Early vs. Delayed: b = -0.0811, SE = 0.0427, z = -1.899, *P* = .057). There was also a significant interaction between condition and test time point (b = 0.0342, SE = 0.0172, z = 1.986, *P* = .047). Comparison of simple slopes revealed that participants provided significantly more External details for Baseline events at Early testing test sessions (*b* = -0.135, SE = 0.0520, *z* = -2.603, *p* = .0092), but not Delayed testing test sessions (*b* = 0.00139, SE = 0.0447, *z* = 0.031, *p* = .975). In contrast to Experiment 1, the main effect of condition was not significant in Experiment 2 (*b* = -0.0259, SE = 0.0189, *z* = -1.373, *p* = .170), but it was for test time point (*b* = -0.146, SE = 0.0355, *z* = -4.104, *p* < .001). Participants provided more External details at Early testing than they did at Delayed testing. The interaction term between condition and test time point was not significant (*b* = 0.0322, SE = 0.0301, *z* = 1.069, *p* = .285). Overall, these data suggest that HippoCamera replay specifically enhances recall of Internal details from everyday experiences.

2.2 Autobiographical Memory Test: In-Cue Details

In-Cue details were common but represented a relatively small proportion of total details recalled (Fig. S1, Table S1-2). Analysis of In-Cue details revealed converging results across Experiment 1 and Experiment 2. In both cases, we found a main effect of condition (Replayed vs. Baseline; Experiment 1: *b* = -0.207, SE = 0.0295, *z* = -7.024, *p* < .001; Experiment 2: *b* = -0.194, SE = 0.0624, *z* = -3.111, P = .00186) that reflected fewer details in the Replayed condition than in the Baseline condition. The main effect of test time point was not significant (Early vs. Delayed; Experiment 1: *b* = -0.00115, SE = 0.0381, *z* = -0.030, *p* = .976; Experiment 2: *b* = 0.0269, SE = 0.0307, *z* = 0.876, *p* = .381). There were also no significant interactions between condition and test time point (Experiment 1: *b* = 0.0133, SE = 0.0295, *z* = 0.450, *p* = .652; *b* = 0.00563, SE = 0.0307, *z* = 0.183, *p* = .855). These results suggest that participants relied less on information reflected in their cues and more on elaborative retrieval processes following repeated replay with HippoCamera.

Table S1. Fit statistics and fixed-effects parameters for best fitting models for Internal, External, and In-Cue detail counts.

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 1	Internal	~ condition × test time point + (condition + test time point | participant)	.493	.264	intercept	1.477	0.0649	22.757	< 2e-16 ***
					condition	0.212	0.0183	11.593	< 2e-16 ***
					test time point	-0.292	0.0336	-8.676	< 2e-16 ***
					condition × test time point	-0.00533	0.0134	-0.397	.691
Experiment 1	External	~ condition × test time point + (condition × test time point | participant)	.650	.015	intercept	1.525	0.125	12.201	< 2e-16 ***
					condition	-0.0335	0.0171	-1.964	.0575
					test time point	-0.0811	0.0427	-1.899	.0495 *
					condition × test time point	0.0342	0.0172	1.986	.0471 *
Experiment 1	In-Cue	~ condition × test time point + (test time point | participant)	.158	.045	intercept	-0.249	0.0754	-3.297	.000976 ***
					condition	-0.207	0.0295	-7.024	2.15e-12 ***
					test time point	-0.00115	0.0381	-0.030	.976
					condition × test time point	0.0133	0.0295	0.450	.653
Experiment 2	Internal	~ condition × test time point + (condition × test time point | participant)	.828	.162	intercept	2.183	0.144	15.114	< 2e-16 ***
					condition	0.222	0.0264	8.407	< 2e-16 ***
					test time point	-0.120	0.0312	-3.840	0.000123 ***
					condition × test time point	-0.0369	0.0142	-2.607	0.009143 **

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 2	External	~ condition × test time point + (condition × test time point | participant)	.893	.041	intercept	2.434	0.185	13.133	< 2e-16 ***
					condition	-0.0278	0.0181	-1.536	.124
					test time point	-0.138	0.0323	-4.288	1.8e-05 ***
					condition × test time point	0.0314	0.0289	1.087	.277
Experiment 2	In-Cue	~ condition × test time point + (condition | participant)	.366	.043	intercept	-0.0173	0.151	-0.114	.909
					condition	-0.194	0.0624	-3.111	.00186 **
					test time point	0.0269	0.0307	0.876	.381
					condition × test time point	0.00563	0.0307	0.183	.855
N/A	Overall	~ intervention length + (1 | participant)	.815	.336	intercept	2.667	0.0792	33.681	< 2e-16 ***
					condition	0.376	0.0792	4.751	2.02e-06 ***

Legend: R^2_C = conditional coefficient of determination, R^2_M: marginal coefficient of determination, SE: standard error, ** *P* < .01, *** *P* < .001. Note: models specified here are multilevel generalized Poisson models to best account for the count-based nature of the number of details recollected.

2.3 Autobiographical Memory Test: Internal Details by Subtype

Our primary behavioral analyses revealed that HippoCamera replay selectively enhanced recall of Internal details. We next examined whether this effect reflected an increase in all subtypes of Internal details. Consistent with previous evidence from older adults (1), most Internal details were of the Event detail subtype (57% of all Internal details were Event details in both experiments). The remaining details were roughly evenly distributed among Place, Perceptual, and Thought/Emotion details. Time details were least frequently described. The most consistent result across detail types was a main effect of condition (Replayed > Baseline; Fig. S2; Table S2-3).

Internal Event: We found a significant main effect of condition (Replayed vs. Baseline) in Experiment 1 (b = 0.224, SE = 0.0257, z = 8.726, p < .001) and Experiment 2 (b = 0.276, SE = 0.0420, z = 6.577, p < .001). There was also a significant main effect of test time point (Early vs. Delayed) in Experiment 1 (b = -0.294, SE = 0.0405, z = -7.267, p < .001) and Experiment 2 (b = -0.120, SE = 0.0391, z = -3.079, p = .00208). There were no significant interactions between condition and test time point in either experiment (Experiment 1: b = -0.0177, SE = 0.0181, z = -0.976, p = .329; Experiment 2: b = -0.0309, SE = 0.0200, z = -1.549, p = .121). In sum, participants recalled more Internal Event details for Replayed events than Baseline events, and at Early than Delayed testing.

Internal Time: We found a significant main effect of condition (Replayed vs. Baseline) in Experiment 1 (b = 0.137, SE = 0.0448, z = 3.059, p = .00222) and Experiment 2 (b = 0.236, SE = 0.0631, z = 3.434, p < .001). There was also a significant main effect of test time point (Early vs. Delayed) in Experiment 1 (b = -0.379, SE = 0.0674, z = -5.634, p < .001) and Experiment 2 (b = -0.233, SE = 0.0538, z = -4.331, p < .001). There were no significant interactions between condition and test time point in either experiment (Experiment 1: b = 0.00328, SE = 0.0448, z = 0.073, p = .942; Experiment 2: b = 0.0336, SE = 0.0538, z = 0.625, p = .532). In sum, participants recalled more Internal Time details for Replayed events than Baseline events, and at Early than Delayed testing.

Internal Thought/Emotion: We found a significant main effect of condition (Replayed vs. Baseline) in Experiment 1 (*b* = 0.139, SE = 0.0501, *z* = 2.764, *p* = .00571) and Experiment 2 (*b* = 0.249, SE = 0.0423, *z* = 5.882, *p* < .001). There was also a significant

main effect of test time point (Early vs. Delayed) in Experiment 1 (b = -0.326, SE = 0.0434, z = -7.520, p < .001) and Experiment 2 (b = -0.128, SE = 0.0484, z = -2.634, p = .00843). There were no significant interactions between condition and test time point in either experiment (Experiment 1: b = -0.0492, SE = 0.0331, z = -1.483, p = .138; Experiment 2: b = -0.0421, SE = 0.0285, z = -1.481, p = .139). In sum, participants recalled more Internal Thought/Emotion details for Replayed events than Baseline events, and at Early than Delayed testing.

Internal Place: We found a significant main effect of condition (Replayed vs. Baseline) in Experiment 1 (b = 0.363, SE = 0.0409, z = 8.885, p < .001) and Experiment 2 (b = 0.227, SE = 0.0363, z = 6.258, p < .001). There were no significant main effects of test time point (Early vs. Delayed) in either experiment (Experiment 1: b = -0.0755, SE = 0.0532, z = -1.420, p = .156; Experiment 2: b = -0.0417, SE = 0.0489, z = -0.854, p = .393). There were no significant interactions between condition and test time point in either experiment (Experiment 1: b = -0.0154, SE = 0.0409, z = -0.376, p = .707; Experiment 2: b = -0.0378, SE = 0.0363, z = -1.040, p = .298). In sum, participants recalled more Internal Place details for Replayed events than Baseline events.

Internal Perceptual: We found a significant main effect of condition (Replayed vs. Baseline) in Experiment 1 (b = 0.179, SE = 0.0569, z = 3.137, p = .00170) and Experiment 2 (b = 0.212, SE = 0.0269, z = 7.865, p < .001). There was also a significant main effect of test time point (Early vs. Delayed) in Experiment 1 (b = -0.331, SE = 0.0612, z = -5.404, p < .001) and Experiment 2 (b = -0.148, SE = 0.0479, z = -3.077, p = .00209). The interaction between condition and test time point was not significant in Experiment 1 (b = 0.0816, SE = 0.0553, z = 1.474, p = .141) but was in Experiment 2 (b = -0.0745, SE = 0.0269, z = -2.766, p = .00567). The significant interaction in Experiment 2 was driven by the fact that the increase in Internal Perceptual details for Replayed events relative to Baseline events was greater at Early testing (b = 0.573, SE = 0.0720, z = 7.956, p < .001) than it was at Delayed testing (b = 0.275, SE = 0.0802, z = 3.426, p < .001).

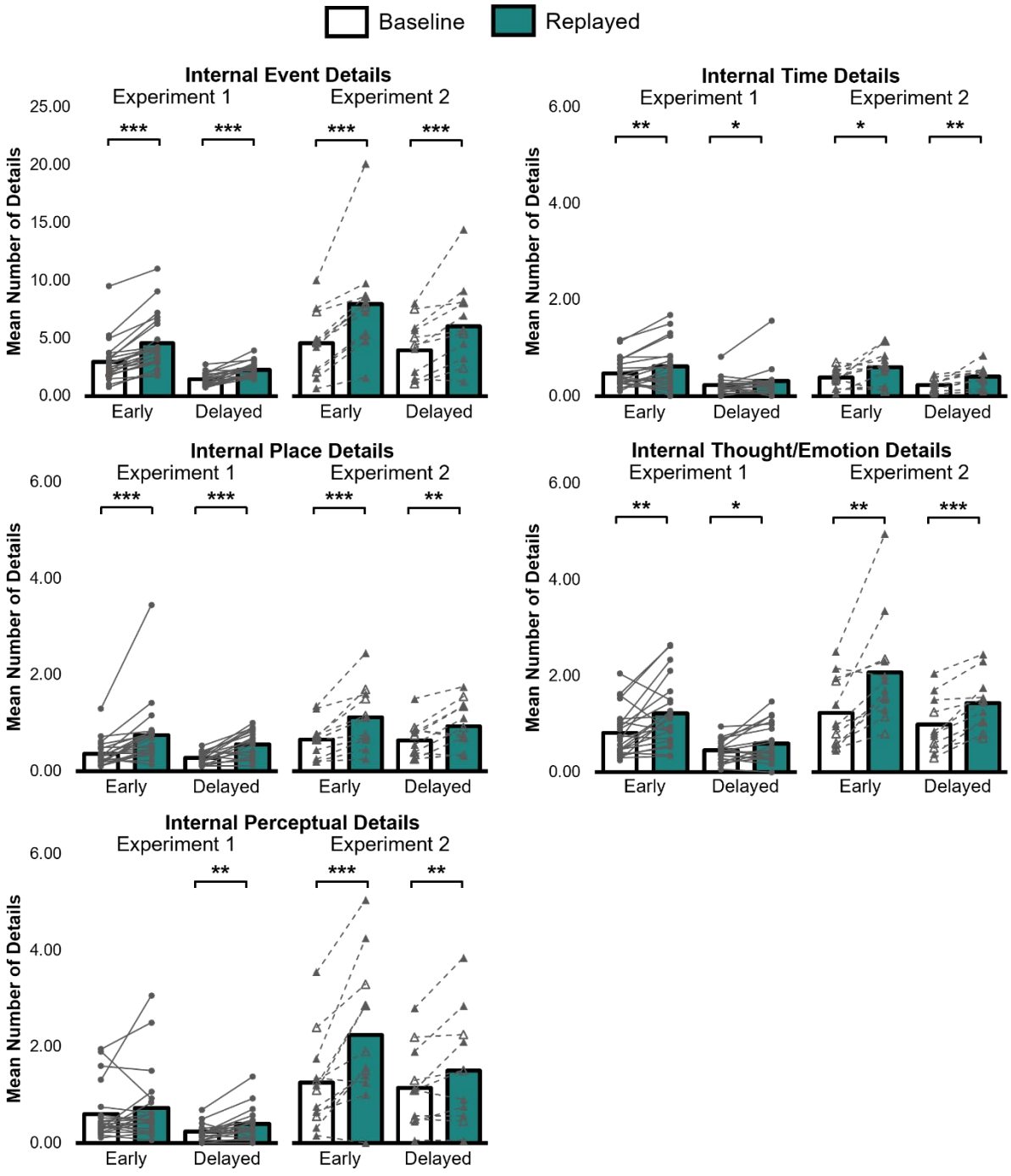

Fig. S2. Behavioral Results – Internal Details by Subtype. Mean number of Internal details by subtype for Experiment 1 and Experiment 2. Early testing corresponds to behavioral performance measured during (Experiment 1) and shortly after (Experiment 2) HippoCamera use. Delayed testing corresponds to behavioral performance after a 3-month delay, during which time participants did not have access to their memory cues. Open triangles denote Experiment 2 participants who failed the Montreal Cognitive Assessment. * < .05, ** < .01, *** < .001.

Table S2. Fit statistics and fixed-effects parameters for best fitting models for Internal detail counts by subtype.

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 1	Event	~ condition × test time point + (condition + test time point | participant)	.423	.199	intercept	0.886	0.0749	11.835	< 2e-16 ***
					condition	0.224	0.0257	8.726	< 2e-16 ***
					test time point	-0.294	0.0405	-7.267	3.69e-13 ***
					condition × test time point	-0.0177	0.0181	-0.976	.329
Experiment 1	Time	~ condition × test time point + (test time point | participant)	.257	.085	intercept	-1.157	0.129	-8.975	< 2e-16 ***
					condition	0.137	0.0448	3.059	.00222 **
					test time point	-0.379	0.0674	-5.634	1.76e-08 ***
					condition × test time point	0.00328	0.0448	0.073	.942
Experiment 1	Place	~ condition × test time point + (test time point | participant)	.228	.087	intercept	-0.897	0.107	-8.390	< 2e-16 ***
					condition	0.363	0.0409	8.885	< 2e-16 ***
					test time point	-0.0755	0.0532	-1.420	.156
					condition × test time point	-0.0154	0.0409	-0.376	.707
Experiment 1	Perceptual	~ condition × test time point + (condition × test time point | participant)	.313	.067	intercept	-1.040	0.158	-6.597	4.20e-11 ***
					condition	0.179	0.0569	3.137	.00170 **
					test time point	-0.331	0.0612	-5.404	6.52e-08 ***
					condition × test time point	0.0816	0.0553	1.474	.141

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 1	Thought/Emotion	~ condition × test time point + (condition + test time point | participant)	.274	.092	intercept	-0.445	0.106	-4.177	2.95e-05 ***
					condition	0.139	0.0501	2.764	.00571 **
					test time point	-0.326	0.0434	-7.520	5.47e-14 ***
					condition × test time point	-0.0492	0.0331	-1.483	.138
Experiment 2	Event	~ condition × test time point + (condition × test time point | participant)	.847	.151	intercept	1.505	0.181	8.314	< 2e-16 ***
					condition	0.276	0.0420	6.577	4.79e-11 ***
					test time point	-0.120	0.0391	-3.079	.00208 **
					condition × test time point	-0.0309	0.0200	-1.549	.121
Experiment 2	Time	~ condition × test time point + (condition | participant)	.188	.060	intercept	-1.059	0.152	-6.973	3.10e-12 ***
					condition	0.236	0.0631	3.734	.000188 ***
					test time point	-0.233	0.0538	-4.331	1.49e-05 ***
					condition × test time point	0.0336	0.0538	0.625	.532
Experiment 2	Place	~ condition × test time point + (test time point | participant)	.277	.045	intercept	-0.337	0.157	-2.147	.0318 *
					condition	0.227	0.0363	6.258	3.89e-10 ***
					test time point	-0.0417	0.0489	-0.854	.393
					condition × test time point	-0.0378	0.0363	-1.040	.298

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 2	Perceptual	~ condition × test time point + (test time point | participant)	.555	.043	intercept	0.0959	0.270	0.355	.722
					condition	0.212	0.0269	7.865	3.70e-15 ***
					test time point	-0.148	0.0479	-3.077	.00209 **
					condition × test time point	-0.0745	0.0269	-2.766	.00567 **
Experiment 2	Thought/Emotion	~ condition × test time point + (condition + test time point | participant)	.397	.108	intercept	0.216	0.129	1.677	.0936
					condition	0.249	0.0423	5.882	4.06e-09 ***
					test time point	-0.128	0.0484	-2.634	.00843 **
					condition × test time point	-0.0421	0.0285	-1.481	.139

Legend: R^2_C = conditional coefficient of determination, R^2_M: marginal coefficient of determination, SE: standard error, ** $P < .01$, *** $P < .001$. Note: models specified here are multilevel generalized Poisson models to best account for the count-based nature of the number of details recollected.

Table S3. Intraclass correlations for intercept-only models of Internal and External detail counts.

Analysis	Internal Details	ICC	External Details	ICC
Experiment 1	Internal Total	.422	External Total	.660
	Internal Event	.374	External Event	.365
	Time	.186	General Semantic	.465
	Place	.185	Autobiographical Fact	.499
	Perceptual	.307	Self-Knowledge	.332
	Thought/Emotion	.207	Repeated Event	.242
			Repeated Detail	.360
			Other	.395
Experiment 2	Internal Total	.697	External Total	.832
	Event	.655	External Event	.645
	Time	.155	General Semantic	.578
	Place	.243	Autobiographical Fact	.715
	Perceptual	.572	Self-Knowledge	.456
	Thought/Emotion	.240	Repeated Event	.624
			Repeated Detail	.362
			Other	.298

2.4 Autobiographical Memory Test External Details by Subtype

To provide a comprehensive picture of our data, we next examined whether HippoCamera replay differentially affected seven subtypes of External details (Fig. S3, Table S3-4).

External Event: There was no main effect of condition (Replayed vs. Baseline) in Experiment 1 (b = -0.0758, SE = 0.0418, z = -1.816, p = .0694) or Experiment 2 (b = -0.103, SE = 0.0643, z = -1.600, p = .110). There was a significant main effect of test time point (Early vs. Delayed) in Experiment 1 (b = -0.264, SE = 0.0753, z = -3.514, p < .001) and Experiment 2 (b = -0.246, SE = 0.0698, z = -3.520, p < .001). In both experiments, participants provided more External Event details at Early than Delayed testing. The interaction between condition and test time point was not significant in Experiment 1 (b = -0.0252, SE = 0.0417, z = -0.604, p = .546) but was in Experiment 2 (b = 0.200, SE = 0.0951, z = 2.106, p = .0352). The interaction in Experiment 2 was driven by more External Event details at Early than Delayed for the Baseline condition (b = -0.892, SE = 0.254, z = -3.516, p < .001) but no difference across test time periods in the Replayed condition (b = -0.0906, SE = 0.217, z = -0.418, p = .676).

General Semantics: The main effect of condition (Replayed vs. Baseline) was not significant in either Experiment 1 (b = 0.0293, SE = 0.0330, z = 0.888, p = .374) or Experiment 2 (b = 0.0133, SE = 0.0536, z = 0.248, p = .804). The main effect of test time point (Early vs. Delayed) was also not significant in Experiment 1 (b = -0.0603, SE = 0.0866, z = -0.696, p = .487) or Experiment 2 (b = -0.0898, SE = 0.0496, z = -1.812, p = .070). There was a significant interaction between condition and test time point in Experiment 1 (b = 0.0868, SE = 0.0330, z = 2.630, p = .00854) but not in Experiment 2 (b = 0.0339, SE = 0.0238, z = 1.422, p = .155). For Experiment 1, General Semantic details did not differ across Baseline and Replayed trials at Early testing (b = -0.115, SE = 0.0858, z = -1.340, p = .180), whereas more details were recalled for Replayed than Baseline at Delayed testing (b = 0.232, SE = 0.100, z = 2.314, p = .0207).

Autobiographical Facts: The main effect of condition (Replayed vs. Baseline) was not significant in either Experiment 1 (b = -0.0287, SE = 0.0364, z = -0.786, p = .432) or Experiment 2 (b = 0.0203, SE = 0.0453, z = 0.449, p = .654). There was a significant main

effect of test time point (Early vs. Delayed) in both experiments (Experiment 1: b = -0.135, SE = 0.0412, z = -3.265, p = .00109; Experiment 2: b = -0.210, SE = 0.0503, z = -4.177, p < .001), such that participants recalled more Autobiographical Fact details during Early as compared to Delayed testing. There were no significant interactions between condition and test time point in either experiment (Experiment 1: b = 0.00830, SE = 0.0223, z = 0.372, p = .710; Experiment 2: b = 0.0467, SE = 0.0475, z = 0.985, p = .325).

Self-Knowledge: The main effect of condition was not significant in either Experiment 1 (b = -0.0508, SE = 0.0417, z = -1.218, p = .223) or Experiment 2 (b = 0.00719, SE = 0.0700, z = 0.103, p = .918). There was a significant main effect of test time point (Early vs. Delayed) in both experiments (Experiment 1: b = -0.252, SE = 0.0707, z = -3.559, p < .001; Experiment 2: b = -0.118, SE = 0.0322, z = -3.652, p < .001), such that participants recalled more Self-Knowledge details during Early as compared to Delayed testing. There were no significant interactions between condition and test time point in either experiment (Experiment 1: b = -0.0168, SE = 0.0417, z = -0.402, p = .688; Experiment 2: b = 0.0238, SE = 0.0322, z = 0.739, p = .460).

Repeated Events: We found a significant main effect of condition (Replayed vs. Baseline) in Experiment 1 (b = -0.323, SE = 0.122, z = -2.639, p = .00831) and Experiment 2 (b = -0.168, SE = 0.0657, z = -2.561, p = .0104). Across experiments, participants recalled more Repeated Event details for Baseline events compared to Replayed events. The main effect of test time point was not significant in either experiment (Experiment 1: b = -0.182, SE = 0.112, z = -1.626, p = .104; Experiment 2: b = -0.0725, SE = 0.0940, z = -0.771, p = .441). The interaction between condition and test time point was not significant in either experiment (Experiment 1: b = 0.0131, SE = 0.0762, z = 0.172, p = .864; Experiment 2: b = 0.0629, SE = 0.0617, z = 1.019, p = .308).

Repeated Details: We found a significant main effect of condition (Replayed vs. Baseline) in Experiment 1 (b = 0.0807, SE = 0.0321, z = 2.518, p = .0118), with participants providing more Repeated details for Replayed events as compared to Baseline events. The effect of condition was not significant in Experiment 2 (b = 0.0431, SE = 0.0329, z = 1.310, p = .190). The effect of test time point was not significant in Experiment 1 (b = -0.132, SE = 0.0738, z = -1.785, p = .0743) but was in Experiment 2 (b = -0.210, SE =

0.0679, z = -3.097, p = .00196). In Experiment 2, participants repeated more details at Early testing than they did at Delayed testing. The interaction between condition and test time point was not significant in either experiment (Experiment 1: b = 0.0171, SE = 0.0321, z = 0.533, p = .594; Experiment 2: b = 0.0206, SE = 0.0272, z = 0.760, p = .447).

Other Details: We found a significant main effect of condition (Replayed vs. Baseline) in Experiment 1 (b = -0.0567, SE = 0.0239, z = -2.370, p = .0178), which reflected more Other details for Baseline as compared to Replayed events. The effect of condition was not significant in Experiment 2 (b = -0.0146, SE = 0.0434, z = -0.337, p = .736). The effect of test time point was not significant in either experiment (Experiment 1: b = 0.103, SE = 0.0542, z = 1.894, p = .0582; Experiment 2: b = 0.00735, SE = 0.0569, z = 0.129, p = .897). There were also no significant interactions between condition and test time point (Experiment 1: b = 0.0243, SE = 0.0239, z = 1.018, p = .309; Experiment 2: b = -0.0160, SE = 0.0224, z = -0.713, p = .476).

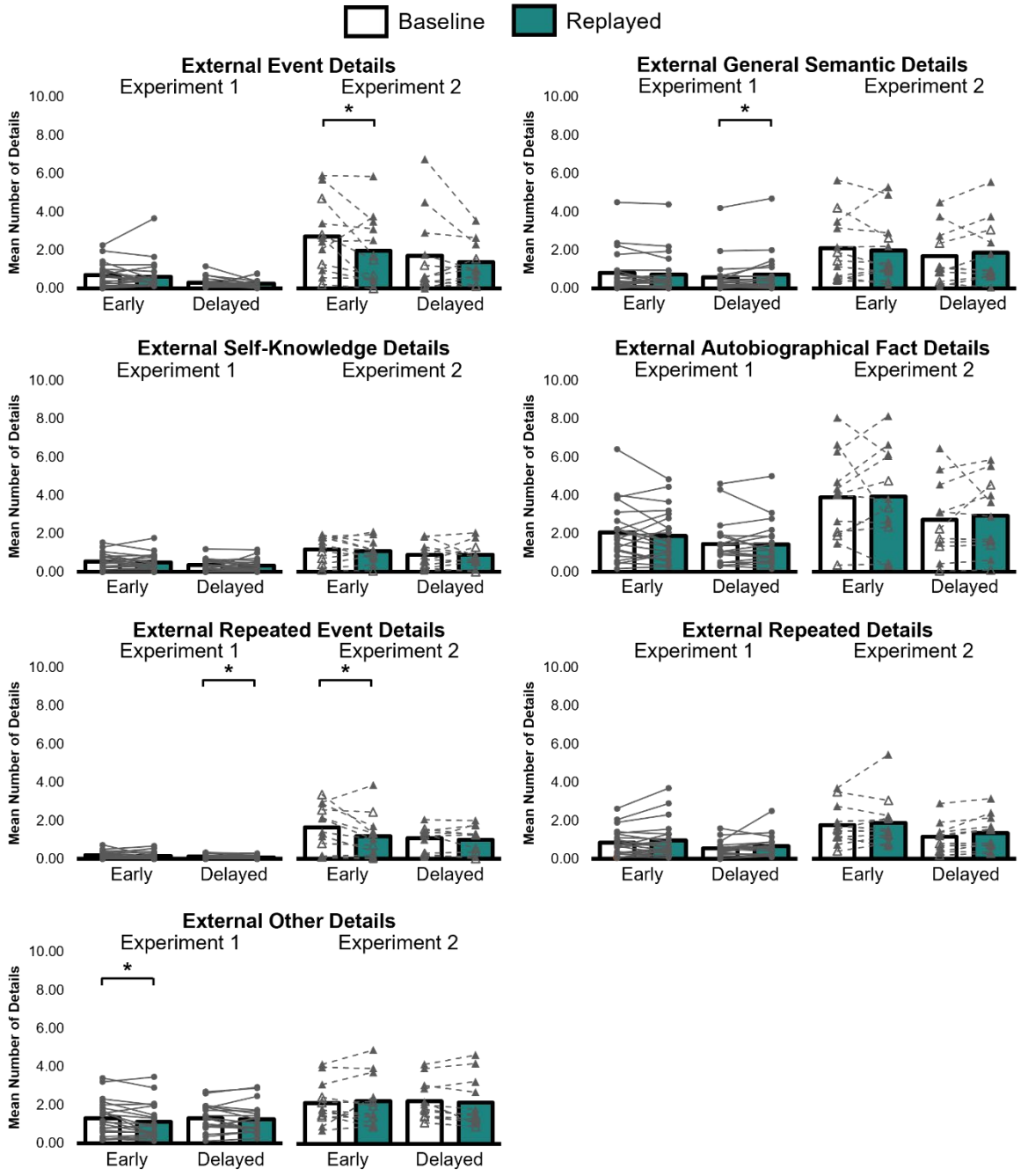

Fig. S3. Behavioral Results – External Details by Subtype. Mean number of External details by subtype for Experiment 1 and Experiment 2. Early testing corresponds to behavioral performance measured during (Experiment 1) and shortly after (Experiment 2) HippoCamera use. Delayed testing corresponds to behavioral performance after a 3-month delay, during which time participants did not have access to their memory cues. Open triangles denote Experiment 2 participants who failed the Montreal Cognitive Assessment. * < .05.

Table S4. Fit statistics and fixed-effects parameters for best fitting models for External detail counts by subtype.

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 1	External Event	~ condition × test time point + (test time point | participant)	.313	.032	intercept	-1.116	0.174	-6.405	1.50e-10 ***
					condition	-0.0758	0.0418	-1.816	.0694
					test time point	-0.264	0.0753	-3.514	.000442 ***
					condition × test time point	-0.0252	0.0417	-0.604	.546
Experiment 1	General Semantic	~ condition × test time point + (test time point | participant)	.483	.005	intercept	-0.961	0.230	-4.172	3.02e-05 ***
					condition	0.0293	0.0330	0.888	.374
					test time point	-0.0603	0.0866	-0.696	.487
					condition × test time point	0.0868	0.0330	2.630	.00854 **
Experiment 1	Autobiographical Fact	~ condition × test time point + (condition + test time point | participant)	.502	.017	intercept	0.311	0.155	2.007	.0447 *
					condition	-0.0287	0.0364	-0.786	.432
					test time point	-0.135	0.0412	-3.265	.00109 **
					condition × test time point	0.00830	0.0223	0.372	.710
Experiment 1	Self-Knowledge	~ condition × test time point + (test time point | Participant)	.339	.028	intercept	-1.144	0.188	-6.084	1.17e-09 ***
					condition	-0.0508	0.0417	-1.218	.223
					test time point	-0.252	0.0707	-3.559	.000373 ***
					condition × test time point	-0.0168	0.0417	-0.405	.688

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 1	Repeated Event	~ condition × test time point + (condition + test time point | participant)	.289	.038	intercept	-2.405	0.225	-10.695	< 2e-16 ***
					condition	-0.323	0.122	-2.639	.00831 **
					test time point	-0.182	0.112	-1.626	.104
					condition × test time point	0.0131	0.0762	0.172	.864
Experiment 1	Repeated Detail	~ condition × test time point + (test time point | participant)	.364	.014	intercept	-0.573	0.157	-3.640	.000273 ***
					condition	0.0807	0.0321	2.518	.0118 *
					test time point	-0.132	0.0738	-1.785	.0743
					condition × test time point	0.0171	0.0321	0.533	.594
Experiment 1	Other	~ condition × test time point + (test time point | participant)	.440	.013	intercept	0.0198	0.146	0.136	.892
					condition	-0.0567	0.0239	-2.370	.0179 *
					test time point	0.103	0.0542	1.894	.0582
					condition × test time point	0.0243	0.0239	1.018	.309
Experiment 2	External Event	~ condition × test time point + (condition × test time point | participant)	.703	.056	intercept	0.142	0.309	0.459	.646
					condition	-0.103	0.0643	-1.600	.110
					test time point	-0.246	0.0698	-3.520	.000431 ***
					condition × test time point	0.200	0.0951	2.106	.0352 *

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 2	General semantic	~ condition × test time point + (condition + test time point | participant)	.586	.007	intercept	.0287	0.256	1.120	.263
					condition	0.0133	0.0536	0.248	.804
					test time point	-0.0898	0.0496	-1.812	.070
					condition × test time point	0.0339	0.0238	1.422	.155
Experiment 2	Autobiographical Fact	~ condition × test time point + (condition × test time point | participant)	.877	.041	intercept	0.883	0.276	3.200	.00137 **
					condition	0.0203	0.0453	0.449	.654
					test time point	-0.210	0.0503	-4.177	2.95e-05 ***
					condition × test time point	0.0467	0.0475	0.985	.325
Experiment 2	Self-Knowledge	~ condition × test time point + (condition | participant)	.470	.009	intercept	-0.252	0.242	-1.041	.298
					condition	0.00719	0.0700	0.103	.918
					test time point	-0.118	0.0322	-3.652	.000260 ***
					condition × test time point	0.0238	0.0322	0.739	.460
Experiment 2	Repeated Event	~ condition × test time point + (condition × test time point | participant)	.589	.014	intercept	-0.305	0.346	-0.883	.377
					condition	-0.168	0.0657	-2.561	.01404 *
					test time point	-0.0725	0.0940	-0.771	.441
					condition × test time point	0.0629	0.0617	1.019	.308

Experiment	Detail Type	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	z	*P*
Experiment 2	Repeated Detail	~ condition × test time point + (condition + test time point | participant)	.440	.048	intercept	0.230	0.171	1.347	.178
					condition	0.0431	0.0329	1.310	.190
					test time point	-0.210	0.0679	-3.097	.00196 **
					Replay condition × test time point	0.0206	0.0272	0.760	.447
Experiment 2	Other	~ condition × test time point + (condition + test time point | participant)	.356	.001	intercept	0.661	0.125	5.289	1.23e-07 ***
					condition	-0.0146	0.0434	-0.337	.736
					test time point	0.00735	0.0569	0.129	.897
					condition × test time point	-0.0160	0.0224	-0.713	.476

Legend: R^2_C = conditional coefficient of determination, R^2_M: marginal coefficient of determination, SE: standard error, ** *P* < .01, *** *P* < .001. Note: models specified here are multilevel generalized Poisson models to best account for the count-based nature of the number of details recollected.

2.5 Autobiographical Memory Test Recall Density

In addition to analyzing total detail counts, we also considered detail density, which controls for differences in verbal output across participants (10, 11). Internal and External density scores were derived for each trial by dividing respective detail counts by total word count. Using this approach produced results that converge with those obtained using detail counts as a dependent measure (Table S5).

For Internal density scores in Experiment 1, we found a main effect of condition (b = 0.00575, SE = 0.000941, $t(20) = 6.111$, $p < .001$) and test time point (b = -0.00494, SE = 0.000804, $t(21) = -6.149$, $p < .001$). Participants had higher Internal detail density scores both for Replayed events compared to Baseline events, and at Early testing as compared to Delayed testing. Additionally, there was a significant interaction between condition and test time point (b = -0.00140, SE = 0.000575, $t(1377) = -2.440$, $p = .0148$). Participants had a significantly larger boost in their Internal detail density scores for Replayed events relative to Baseline events at Early testing (b = 0.0143, SE = 0.00216, $t(36) = 6.611$, $p < .001$) than they did at Delayed testing (b = 0.00869, SE = 0.00225, $t(40) = 3.872$, $p < .001$). In Experiment 2, we found a significant main effect of condition (b = 0.00410, SE = 0.000516, $t(929) = 7.950$, $p < .001$). As in Experiment 1, participants had higher Internal detail density scores for Replayed events compared to Baseline events. However, we found neither a significant main effect of test time point (b = -0.000127, SE = 0.000516, $t(929) = -0.246$, $p = .806$), nor a significant interaction between condition and test time point significant (b = -0.000981, SE = 0.000516, $t(929) = -1.901$, $p = .0576$).

For External density scores in Experiment 1, we found a significant main effect of both condition (b = -0.00265, SE = 0.000521, $t(1382) = -5.084$, $p < .001$) and test time point (b = 0.00296, SE = 0.000778, $t(19) = -3.807$, $p = .00111$). Participants had higher External detail density scores for Baseline events compared to Replayed events, and at Early testing as compared to Delayed testing. The interaction between condition and test time point was not significant (b = 0.000592, SE = 0.000520, $t(1384) = 1.137$, $p = .256$). In Experiment 2, we also found a significant main effect of condition (b = -0.00350, SE = 0.000560, $t(917) = -6.239$, $p < .001$). Participants had higher External detail density scores for Baseline events compared to Replayed events. We did not find a significant

main effect of test time point (*b* = 0.0000822, SE = 0.000715, *t*(10) = -0.115, *p* = .911) nor was the interaction between condition and test time point significant (*b* = 0.000271, SE = 0.000560, *t*(917) = 0.484, *p* = .629).

Overall, results obtained after converting cued recall responses to density scores converge with those obtained using raw detail counts. This outcome indicates that the effect of Replay on episodic recollection of everyday experiences is not driven by particularly loquacious individuals. This interpretation is also consistent with the fact that all participants showed a behavioral advantage in episodic recall for Replayed as compared to Baseline trials.

Table S5. Fit statistics and fixed-effects parameters for best fitting models for density scores.

Experiment	Density Score	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	df	t	p
Experiment 1	Internal	~ condition × test time point + (condition + test time point | participant)	.381	.081	intercept	0.0379	0.00309	21	12.283	4.56e-11 ***
					condition	0.00575	0.000941	20	6.111	4.96e-06 ***
					test time point	-0.00494	0.000804	21	-6.149	4.09e-06 ***
					condition × test time point	-0.00140	0.000575	1377	-2.441	.0148 *
Experiment 1	External	~ condition × test time point + (test time point | participant)	.219	.033	intercept	0.0373	0.00206	21	18.114	2.02e-14 ***
					condition	-0.00265	0.000521	1382	-5.084	4.20e-07 ***
					test time point	0.00296	0.000778	19	3.807	.00111 **
					condition × test time point	0.000592	0.000520	1384	1.137	.0256

Experiment	Density Score	Model Formula	R^2_C	R^2_M	Fixed Effect	Estimate (b)	SE	df	t	p
Experiment 2	Internal	~ condition × test time point + (1 | participant)	.353	.046	intercept	0.0256	0.00319	11	8.008	6.49e-06 ***
					condition	0.00410	0.000516	929	7.950	5.39e-15 ***
					test time point	-0.000127	0.000516	929	-0.246	.806
					condition × test time point	-0.000981	0.000516	929	-1.901	.0576
Experiment 2	External	~ condition × test time point + (test time point | participant)	.276	.030	intercept	0.0349	0.00292	11	11.944	1.21e-07 ***
					condition	-0.00350	0.000560	917	-6.239	6.71e-10 ***
					test time point	-8.217e-05	0.000715	10	-0.115	.911
					condition × test time point	0.000271	0.000560	917	0.484	.629

Legend: R^2_C = conditional coefficient of determination, R^2_M = marginal coefficient of determination, SE = standard error, ** $p < .01$, *** $p < .001$.

2.6 Sentiment Analyses

We generated histograms of sentiment scores for all trials in order to provide a comprehensive picture that captures precisely how replay influence autobiographical sentiment (Fig. S4). Visual inspection suggests that the distribution of Replayed sentiment scores is shifted to be more positive than Baseline sentiment scores, with the most pronounced differences in the tails. Specifically, we find more highly positive outcomes and fewer negative outcomes for Replayed memories. In line with our inferential statistics, this shift is most apparent in Experiment 1 (2 weeks of HippoCamera use) and at Early testing. When considered together with autobiographical recall detail counts, these data suggest that beneficial effect of replay on episodic memory is more durable over time than is the effect of replay on sentiment.

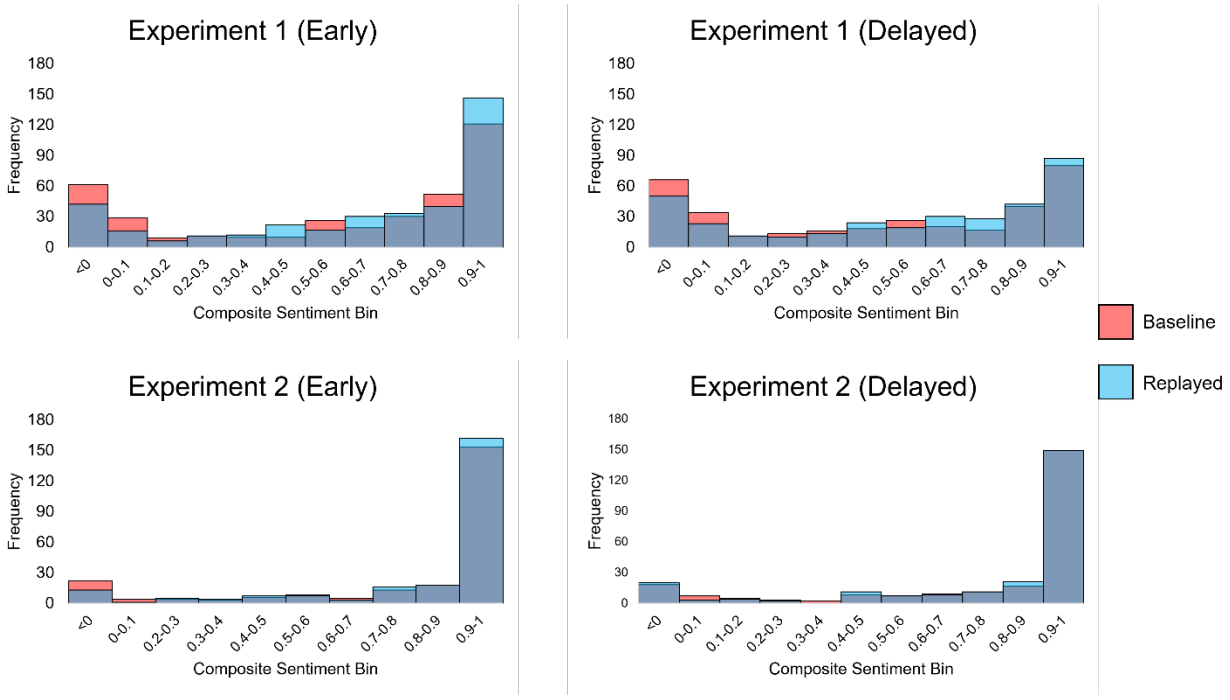

Fig. S4. Behavioral Results – Sentiment Histograms. Frequency distributions for composite sentiment scores for Early and Delayed testing in Experiments 1 and 2. Higher scores reflect more positive sentiment.

3. fMRI Methods

3.1 MRI and fMRI Acquisition

MRI data were recorded on a 3T Siemens Magnetom Prisma system at the Toronto Neuroimaging facility using a 32-channel head coil. High-resolution anatomical images were acquired with a 3D-MPRAGE T1-weighted sequence with oblique axial slices covering the whole brain (TR = 2000 ms, TE = 2.4 ms, flip angle = 9°, voxel size = 1 mm^3, matrix size = 192 x 256 x 160). Functional images were recorded using a gradient echo EPI sequence with 56 oblique axial slices oriented parallel to the hippocampus (TR = 1000 ms, TE = 30 ms, flip angle = 45°, voxel size = 3.4 x 3.4 x 3.0 mm, matrix size = 70 x 70 x 56). The number of functional volumes acquired per run varied across participants (M = 298.25, range = 265-327), reflecting the fact that some participants had more memories tested in the autobiographical interview than others.

3.2 MRI and fMRI Preprocessing

All neuroimaging data were preprocessed using FSL 6.00 (FMRIB software library, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki). Images were skull-stripped using BET (12), and data were corrected for slice-acquisition time, high-pass temporally filtered using a 50-second period cut-off, and motion corrected using MCFLIRT (13). Each participant's functional data were smoothed using a 5mm FWHM kernel and registered to their high-resolution anatomical image using FLIRT boundary-based registration. The resulting data were analyzed using first-level FEAT (14). Parameter estimates of BOLD response amplitude were computed using FILM, with a general linear model that included temporal autocorrelation correction and six motion parameters as nuisance covariates. Each task component of each trial (Watch Cue, Mentally Relive, and Episodic Probe) was modeled with a boxcar function corresponding to the event onset and then convolved with a double-gamma hemodynamic response function. Regressors were orthogonalized with respect to the preceding regressor using FSLs 'orthogonalise' function. Separate t-statistic images were created for each task component on each trial, which were then normalized to MNI space. Memory-specific t-images were then averaged across repetitions, resulting in a single t-image for each memory. These data were used for the purpose of our pattern-based similarity analyses.

3.3 ROI Definition

All fMRI analyses were completed using data that were normalized to MNI space. Accordingly, one hippocampal ROI was created in each hemisphere using an MNI template. For the purpose of our first exploratory analysis, we used the uncal apex as an anatomical marker to create anterior hippocampus ROI that was distinct from a posterior hippocampus ROI, bilaterally (15). We created seven additional ROIs to for the purpose of a second set of exploratory analyses intended to quantify the neuroanatomical specificity of neuroimaging results obtained in the hippocampus. Six of these ROIs have figured prominently in fMRI and behavioral research on autobiographical memory in neurologically healthy populations and individuals with brain damage (vmPFC, ACC, PCC, angular gyrus, precuneus, and PHC; 16, 17). The seventh ROI, postcentral gyrus (i.e., primary somatosensory cortex) served as a control region that has not been systematically implicated in memory research. All exploratory analyses were performed using functional data combined across the left and right cerebral hemispheres. The vmPFC ROI was created using criteria established in previous research focused on memory consolidation (18). This ROI encompassed Brodmann's Area (BA) 14, BA 25, ventral parts of BA 24 and BA 32, the caudal part of BA 10, and the medial part of BA 11. ACC, PCC, angular gyrus, precuneus, and PHC ROIs were defined probabilistically using the Harvard-Oxford Cortical Atlas using 50% as a voxel inclusion threshold.

3.4 Representational Similarity Analysis

Representational similarity analyses were performed using the CoSMoMVPA Matlab toolbox (http://www.cosmomvpa.org/) (19). Analyses focused on activity in the hippocampus, including exploratory analyses that examined the anterior and posterior extent separately, were performed using multivoxel patterns extracted from our ROIs bilaterally.

3.5 Linear Mixed Modeling with fMRI Data

Data were analyzed using multilevel modeling in R 4.1.0 (R Core Team, 2021) using the same tools implemented in our behavioral analyses. We used 2-level multilevel linear models with individual trials nested within participants. Individual models were specified

for the entire hippocampus, the anterior hippocampus, posterior hippocampus, and ventromedial prefrontal cortex (vmPFC). Multilevel modeling was appropriate given the degree of variance explained by individual participants for differentiation scores across all regions of interest, as assessed by their ICCs (Table S6). For all models, we would first always fit the model with the maximal random effects structure, according to Barr et al. (7). To investigate the effect of condition and task component, this entailed estimating fixed effects for condition (Baseline vs. Replayed), task (Episodic Probe vs. Mentally Relive vs. Watch Cue), and their interaction; a random intercept estimated for each participant; and a random slope estimated for each fixed effect. Both condition and task were effect coded (condition: Baseline = -1, Replayed = 1; task: Episodic Probe = 1, Mentally Relive = 0, Watch Cue = -1; for second task component effect code: Episodic Probe = 0, Mentally Relive = 1, Watch Cue = -1). In the situation that the maximal model failed to converge due to overparameterization, we employed the backward-selection heuristic (8). To probe any significant interactions, simple effects tests were conducted with the Tukey adjustment for pairwise comparisons using the emmeans package (9)—adjusted P-values are reported for simple effects. The lmerTest package in R was used to obtain *P*-values corresponding to each fixed effect using likelihood ratio tests with the Satterthwaite approximation for degrees of freedom (20). The best fitting models described using Wilkinson notation, their corresponding model fit statistics, and ANOVA tables assessing the significance of fixed-effects parameters for the above analyses are summarized in Table S6. All models were estimated with an unstructured covariance matrix.

In addition, we performed an exploratory analysis to investigate whether differentiation scores differed across the long axis of the hippocampus. Specifically, we added an additional fixed effect for hippocampal ROI (Anterior vs. Posterior). Hippocampal ROIs were effect coded (Anterior = 1, Posterior = -1). Model specification was otherwise performed in the same fashion as described above.

4. fMRI Results

4.1 Pattern Differentiation in the Anterior versus Posterior Hippocampus

Our main set of fMRI pattern-based similarity analyses focused on the hippocampus in its entirety, which revealed greater differentiation (i.e., pattern dissimilarity) among memories in the Replayed condition than among memories in the Baseline condition (Fig. 4C). A substantial body of evidence suggests that the hippocampus is neither anatomically nor functional homogeneous along its anterior-posterior extent and that the functional distinction may be captured by differences between gist-based and detail-based memory representations (21, 22). Within this framework, the activity in the anterior hippocampus is thought to support gist-based memory, whereas the posterior hippocampus is thought to support detail-based memory. To investigate whether replay-related increases in pattern differentiation differed between the anterior and posterior extent of the hippocampus, we performed an exploratory analysis that used the uncal apex as an anatomical marker separating the anterior from the posterior extent (23).

Differentiation scores are presented in Fig. S5. Model fits are presented in Table S6-7. A linear mixed model revealed a significant main effect of ROI, such that anterior hippocampus showed increased differentiation in activity patterns overall compared to posterior hippocampus (b = 6.143×10^{-3}, SE = 2.855×10^{-3}, $t(4597) = 2.152$, $P = .0351$). However, we did not find evidence for any interactions involving ROI and any other predictor for differentiation in activity patterns (all P's > .05). We did find a significant main effect of condition (Replayed vs. Baseline: b = 2.587×10^{-2}, SE = 2.855×10^{-3}, $t(4597) = 9.060$, $P < .001$) and a significant interaction between condition and task ($F(2, 4597) = 22.337$, $P < .001$). This was driven by increased differentiation in hippocampal activity patterns for Replayed compared to Baseline events during the Episodic Probe task (b = 0.0886, SE = 8.54×10^{-3}, $t(4597) = 10.378$, $P < .001$) and the Mentally Relive task (b = 0.0582, SE = 0.0122, $t(4597) = 4.791$, $P < .001$) tasks. The difference between Replayed and Baseline differentiation during the Watch Cue task was not significant (b = 8.37×10^{-3}, SE = 8.54×10^{-3}, $t(4597) = 0.980$, $P = .327$). Lastly, we did not find evidence for a significant main effect of task ($F(2, 13) = 0.161$, $P = .853$).

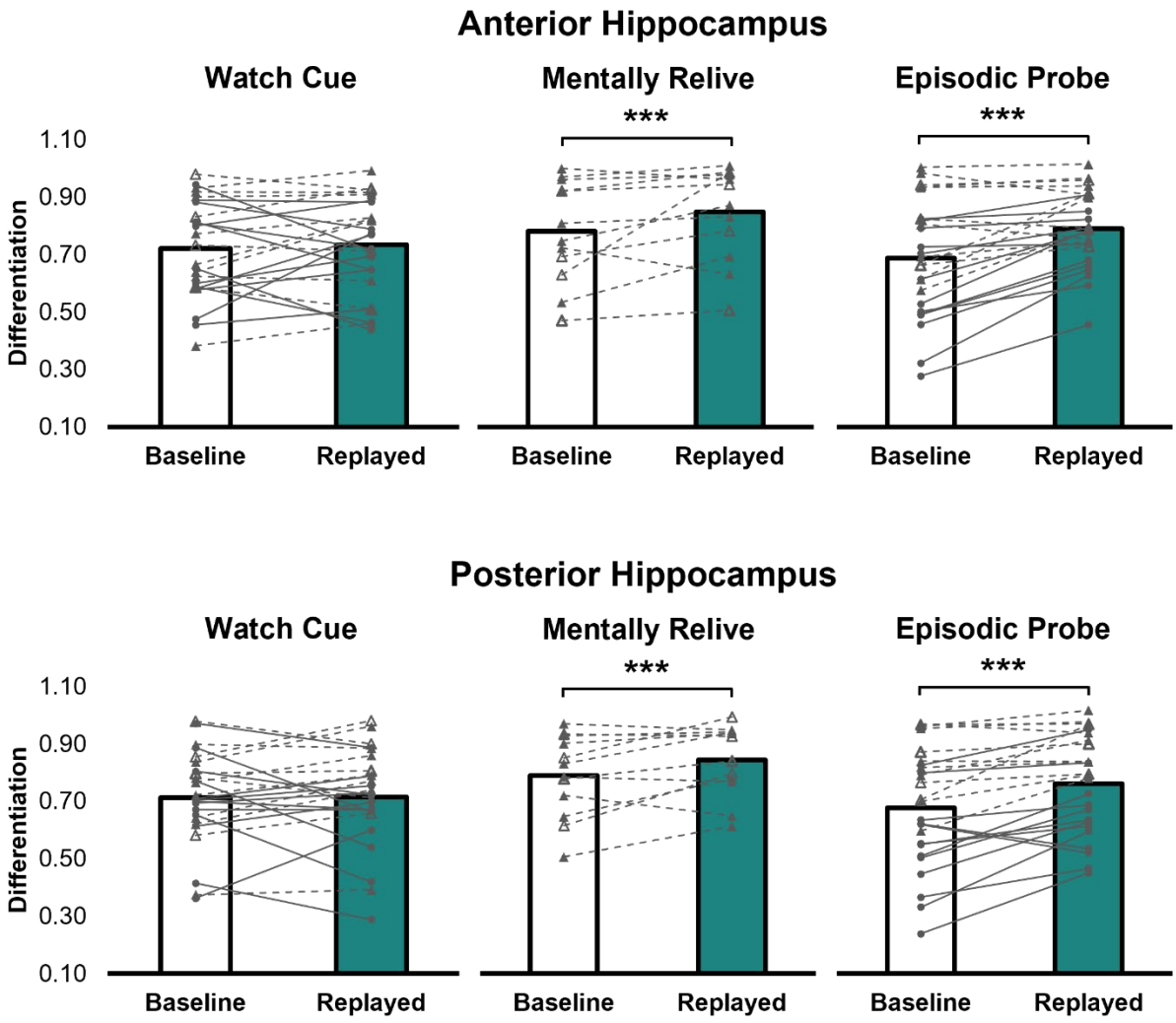

Fig. S5. Differentiation of Anterior and Posterior Hippocampus Activity. We quantified within condition measures of differentiation (1 – Pearson's r) using activity patterns obtained from the anterior and posterior hippocampus during each component of the fMRI task. Solid lines/circle markers depict data from Experiment 1 and dashed lines/triangle markers depict data from Experiment 2. Open triangles denote Experiment 2 participants who failed the Montreal Cognitive Assessment. *** = $P < .001$.

4.2 Pattern Differentiation Beyond the Hippocampus

In addition to the well-established role of the hippocampus in the encoding and retrieval of memory for recent events, previous neuroimaging research has also revealed important contributions of vmPFC, ACC, PCC, angular gyrus, precuneus, and PHC (24-28). We performed exploratory pattern analyses in these regions to characterize the neuroanatomical specificity of the replay-related increased differentiation we obtained

when focusing on the hippocampus. We also included postcentral gyrus as a control ROI that has not figured prominently in memory research. See Fig. S6 for plots with data from individual participants and Table S6 for comprehensive fit statistics from a series of linear mixed models.

The main effect of condition (Replayed vs. Baseline) was not significant in any ROI after correcting for multiple comparisons (Bonferroni corrected threshold = 0.007; vmPFC P = 0.0218; Angular Gyrus P = 0.0684; Anterior Cingulate P = 0.1409; Posterior Cingulate P = 0.0216; Precuneus P = 0.0493; Parahippocampal Cortex P = 0.2547; Postcentral Gyrus P = 0.1772). We note, however, that we did obtain marginal evidence for an effect of replay in vmPFC, Posterior Cingulate, and Precuneus (i.e., significant effects prior to correcting for multiple comparisons). The main effect of task (Watch Cue vs. Episodic Probe vs. Mentally Relive) was significant in Postcentral Gyrus (P = 0.0040), an effect that was driven by greater overall differentiation in the Mentally Relive task. No other ROI showed a main effect of task (Bonferroni corrected threshold = 0.007; vmPFC P = 0.2090; Angular Gyrus P = 0.1093; Anterior Cingulate P = 0.7812; Posterior Cingulate P = 0.3016; Precuneus P = 0.4209; Parahippocampal Cortex P = 0.0775). Lastly, there was a significant interaction between condition and task in Parahippocampal Cortex (P = 0.0006) and Precuneus (P = 0.0003), but not in any other ROI (Bonferroni corrected threshold = 0.007; vmPFC P = 0.0476; Angular Gyrus P = 0.1873; Anterior Cingulate P = 0.0206; Posterior Cingulate P = 0.1497; Postcentral Gyrus P = 0.0052). The interactions in Parahippocampal Cortex and Precuneus were driven by increased differentiation for Replayed events compared to Baseline events in the Episodic Probe task; no differences were revealed between conditions in the Watch Cue or Mentally Relive tasks.

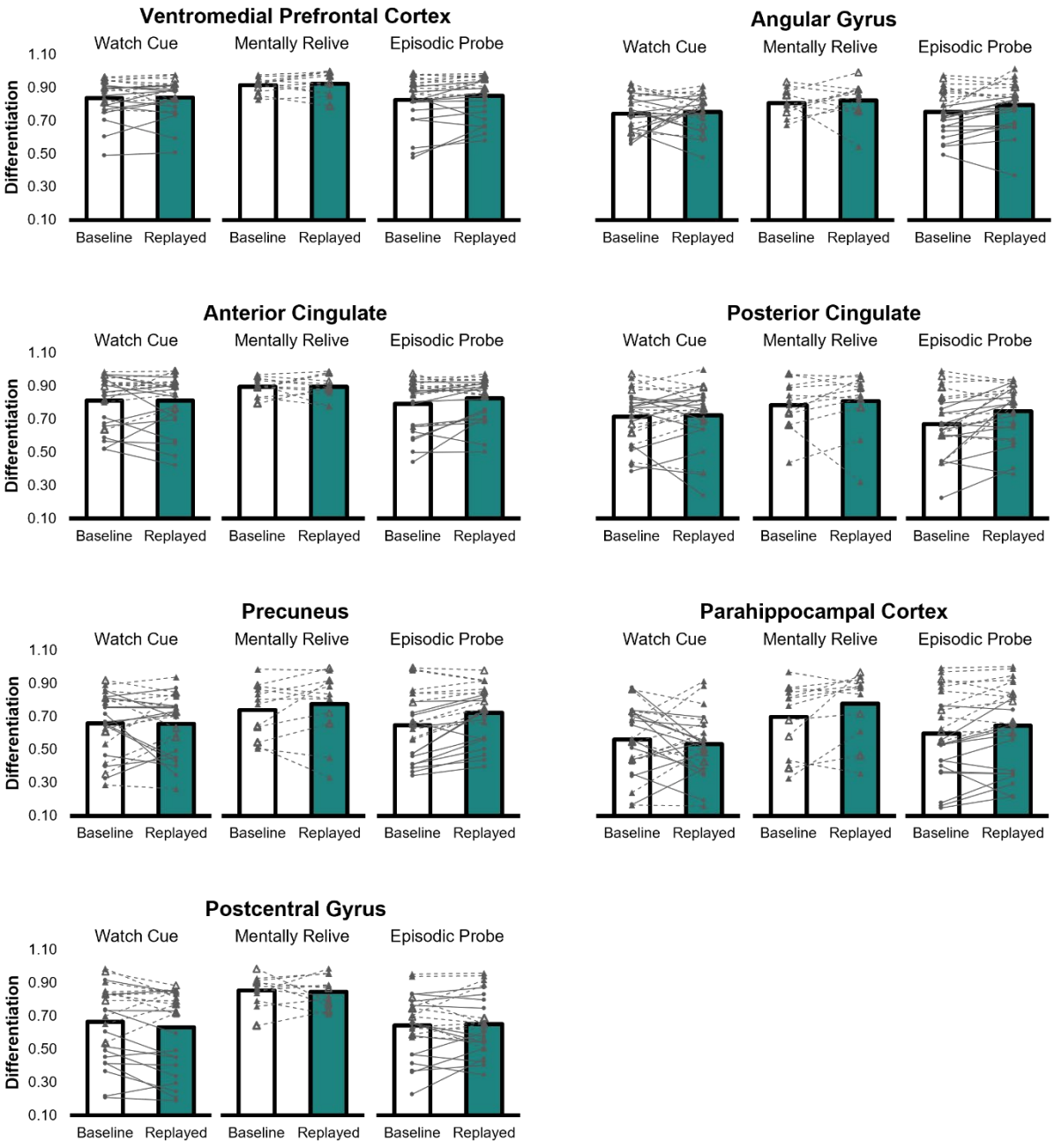

Fig. S6. Pattern differentiation beyond the hippocampus. We quantified differentiation (1 – Pearson's r) using activity patterns obtained from seven exploratory ROIs, including six that have figured prominently in research on autobiographical memory (ventromedial prefrontal cortex, angular gyrus, anterior cingulate, posterior cingulate, precuneus, and parahippocampal cortex) and one control region that has not been systematically linked to memory (postcentral gyrus). Solid lines/circle markers depict data from Experiment 1 and dashed lines/triangle markers depict data from Experiment 2. Open triangles denote Experiment 2 participants who failed the Montreal Cognitive Assessment.

Table S6. Fit statistics and fixed-effects parameters for best fitting models of fMRI differentiation by region of interest

Analysis (ROI)	Model Formula	R^2_C	R^2_M	Fixed Effect	SS	MS	df_N	df_D	F	*P*
Hippocampus	~ condition × task + (condition + task | participant)	.500	.022	condition	0.0479	0.479	1	24	17.508	.000330 ***
				task	0.0838	0.0419	2	11	1.532	.257
				condition × task	0.762	0.381	2	2186	13.933	9.71e-07 ***
Anterior/Posterior Hippocampus	~ condition × task × region of Interest + (task | participant)	.423	.018	condition	2.800	2.800	1	4597	82.076	<2.2e-16 ***
				task	0.0110	0.00549	2	13	0.161	.853
				region of Interest	0.158	0.158	1	4597	4.629	.0315 *
				condition × task	1.524	0.762	2	4597	22.337	2.22e-10 ***
				condition × region of Interest	0.0305	0.0305	1	4597	0.863	.345
				task × region of Interest	0.0645	0.0322	2	4597	0.945	.389
				condition × task × region of Interest	0.00402	0.00201	2	4597	0.0589	.943
ventromedial Prefrontal Cortex	~ condition × task + (1 | participant)	.473	.004	condition	0.0762	0.0762	1	2303	5.272	.0218
				task	0.0452	0.0226	2	2309	1.566	.209
				condition × task	0.0881	0.0440	2	2303	3.048	.0476
Angular Gyrus	~ condition × task + (condition + task | participant)	.294	.010	condition	.1061	.1061	1	25.50	3.621	.0684
				task	.1508	.0754	2	15.19	2.571	.1093
				condition × task	.0983	.0492	2	2135.33	1.676	.1873

Analysis (ROI)	Model Formula	R^2_C	R^2_M	Fixed Effect	SS	MS	df_N	df_D	F	*P*
Anterior Cingulate	~ condition × task + (1 | participant)	.465	.003	condition	.0461	.0461	1	2299.9	2.169	.1409
				task	.0105	.0052	2	2305.4	0.247	.7812
				condition × task	.1653	.0827	2	2299.9	3.89	.0206
Posterior Cingulate	~ condition × task + (condition * task | participant)	.480	.017	condition	.2086	.2086	1	14.456	6.626	.0216
				task	.0800	.0400	2	20.686	1.271	.3016
				condition × task	.1328	.0664	2	18.411	2.109	.1497
Precuneus	~ condition × task + (condition + task | participant)	.496	.011	condition	.1452	.1452	1	25.90	4.255	.0493
				task	.0629	.0315	2	13.69	0.922	.4209
				condition × task	.5665	.2832	2	2199.07	8.302	.0003†
Parahipp. Cortex	~ condition × task + (condition + task | participant)	.547	.017	condition	.0599	.0599	1	25.79	1.357	.2547
				task	.2723	.1361	2	14.12	3.081	.0775
				condition × task	.6599	.3299	2	2213.21	7.467	.0006†
Postcentral Gyrus	~ condition × task + (condition + task | participant)	.582	.013	condition	.0569	.0569	1	26.42	1.923	.1772
				task	.7069	.3535	2	7.94	11.939	.0040
				condition × task	.3121	.1560	2	2174.21	5.271	.0052

Legend: R^2_C = conditional coefficient of determination, R^2_M: marginal coefficient of determination, SS: sum of squares, MS: mean squares, dfN: numerator degrees of freedom, dfD: denominator degrees of freedom, * $P < .05$, ** $P < .01$, *** $P < .001$, † $P < .007$ (Bonferroni corrected significance threshold for exploratory analyses).

Table S7. Intraclass correlations for intercept-only models of fMRI differentiation by region of interest

Analysis (ROI)	ICC
Hippocampus	.391
Anterior/Posterior Hippocampus	.349
Ventromedial Prefrontal Cortex	.475
Angular Gyrus	.187
Anterior Cingulate	.464
Posterior Cingulate	.364
Precuneus	.376
Parahippocampal Cortex	.369
Postcentral Gyrus	.524

4.3 Relationship Between Pattern Differentiation in the Hippocampus and Recall of External Details

In a final set of exploratory analyses, we correlated differentiation scores from the hippocampus and number of External details recalled in our Early and Delayed behavioral assessments (Fig. S7). Whereas hippocampal differentiation was significantly correlated with number of Internal details recalled (Fig. 5 in main text), we found limited evidence for any such relationship with External details. A one-sample t-test against a mean of zero revealed that the group averaged correlation values were significantly greater than chance for the Mentally Relive task at Early testing (t(11) = 2.23, $P < .05$). No other correlations were significantly greater than chance (Watch Cue and Early testing ($t(24) = 0.286$, $P = .777$), Watch Cue and Delayed testing ($t(22) = -0.866$, $P = .396$), Mentally Relive and Delayed testing ($t(11) = -0.224$, $P = .827$), Episodic Probe and Early testing ($t(24) = -0.469$, $P = .643$), or Episodic Probe and Delayed testing ($t(22) = 1.59$, $P = .124$).

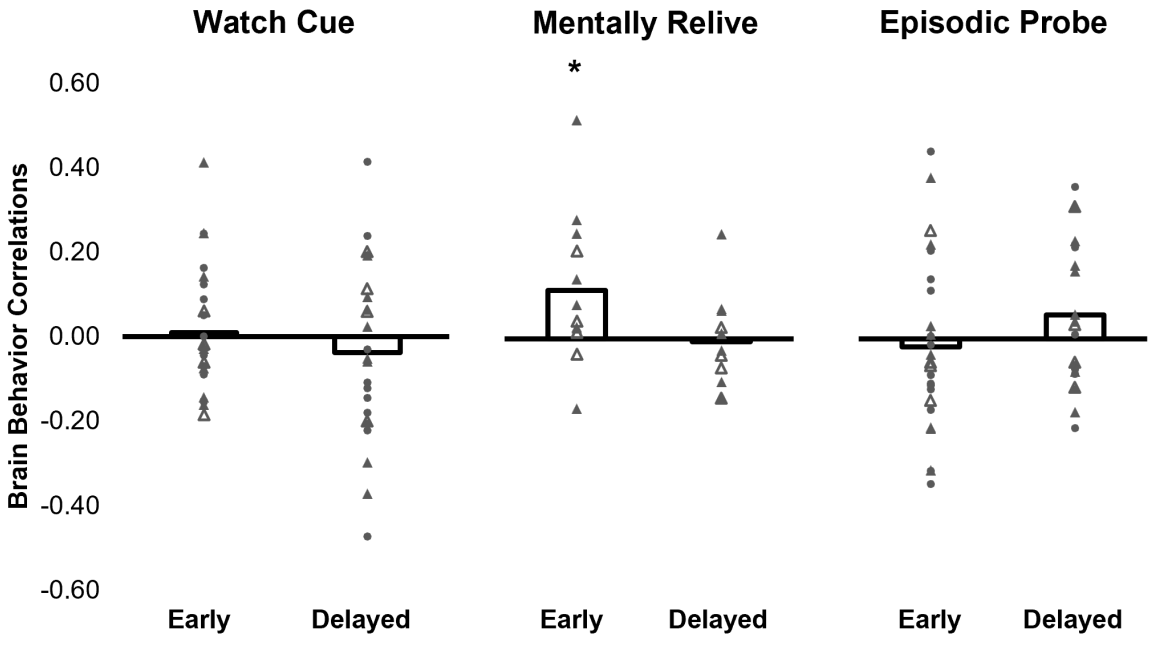

Fig. S7. Correlation Between Differentiation Scores in the Hippocampus and Retrieval of External Details. Pearson's correlation values obtained between hippocampal differentiation and overall number of External details recalled at Early testing and Delayed testing. Circles reflect participants from Experiment 1. Triangles reflect participants from Experiment 2. Open triangles denote Experiment 2 participants who failed the Montreal Cognitive Assessment. * < .05

5. Single-Case Study in Hippocampal Amnesia

Here we report behavioral and neuroimaging data that reveal the potential of HippoCamera as a digital memory aid in older adults, including a subset of participants who are at risk for cognitive decline. An important outstanding question is whether this cognitive behavioral intervention will similarly benefit participants with frank memory disorders. To this end, we complement our primary findings with evidence from an individual who suffers from dense hippocampal amnesia.

5.1 Participant

We tested HippoCamera in a 28 year-old woman (LP) who developed temporal lobe encephalitis due to a viral infection. Review of a T1-weighted MR image revealed near complete ablation of posterior hippocampus bilaterally but some sparing of anterior hippocampus. Further abnormalities are evident in the amygdala and medial temporal lobe cortical areas (i.e., entorhinal cortex, perirhinal cortex, and parahippocampal cortex). She was left profoundly amnesic (e.g., Logical Memory Immediated = 4/25, Delayed = 0/25) and showed considerable evidence of general cognitive decline (e.g., 6/30 on the Montreal Cognitive Assessment).

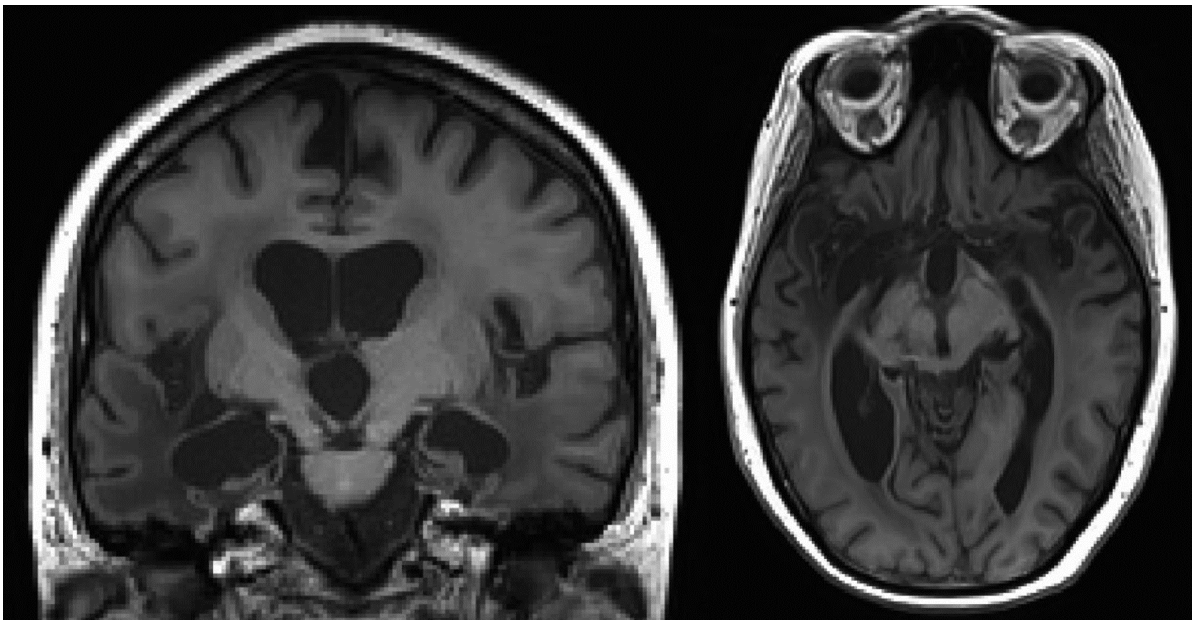

Fig. S8. T1-Weighted MR Image of L.P.'s Bilateral Medial Temporal Lobe Lesion. White arrows highlight extensive tissue loss in the hippocampus, amygdala, entorhinal cortex, perirhinal cortex, and parahippocampal cortex.

5.2 Experimental Design and Results

At intake we noticed that LP diligently kept a diary with approximately 10 entries per day. Moreover, she would often review it to make conversation or reminder herself of an event. In addition to comparing memory for Replayed vs. Baseline events captured using HippoCamera, we also used her written notes as a diary control condition. LP recorded 49 events (2.2/day) and viewed 106 Replay sessions (4.8/day) over a 3-week period. HippoCamera memory cues were assigned to the Replayed and Baseline condition randomly in an interleaved manner (i.e., the same procedure used for the purpose of Experiment 1).

At test, we showed LP cues from either her diary (N = 13 trials), the Baseline HippoCamera condition (N = 10 trials), or the Replayed HippoCamera condition (N = 12 trials) and asked her to describe each event. She produced External details for all tested events (mean number per event: Diary = 2.92; Baseline HippoCamera = 2.30; Replayed HippoCamera = 3.08), as is typical of patients with hippocampal damage. She recalled an equal number of Internal details for events in the Diary condition (mean = 0.54) and Baseline HippoCamera condition (mean = 0.50). Critically, however, she produced nearly three times as many Internal details for events replayed with HippoCamera (mean = 1.50) relative to either her Diary or the Baseline HippoCamera condition (Fig. S9). Notably, while she produced no episodic details for 3-week old events recorded with her diary, she produced some details for 3-week old events recorded with HippoCamera. Anecdotally, her mother recounted a “breakthrough moment” (her words) in which the patient described an important event recorded with HippoCamera – a kayak ride at her friend’s cottage. Critically, this memory was produced spontaneously in conversation (i.e., without cues), suggesting that HippoCamera might engender flexible retrieval of events in a manner that journal entries cannot. Lastly, these findings suggest that HippoCamera may have therapeutic value for individuals with memory disorders.

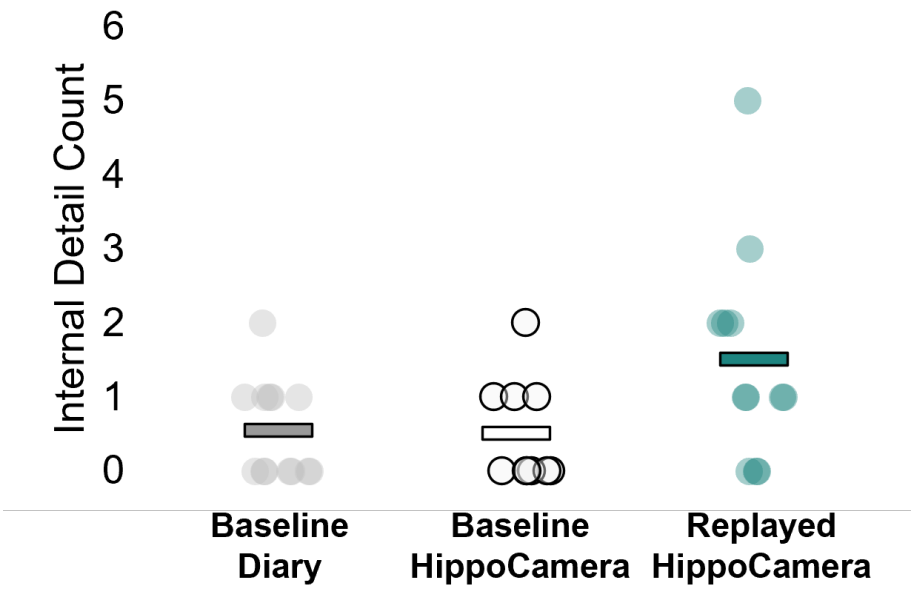

Fig. S9. HippoCamera Replay Improved Recall of Internal Details in LP. Bars reflect mean values for each condition. Individual markers reflect performance on each trial within a given condition.

6. References

1. Levine B, Svoboda E, Hay JF, Winocur G, Moscovitch M (2002) Aging and autobiographical memory: dissociating episodic from semantic retrieval. Psychol Aging 17:677-689.

2. Renoult L, Armson MJ, Diamond NB, Fan CL, Jeyakumar N, Levesque L, et al. (2020) Classification of general and personal semantic details in the Autobiographical Interview. Neuropsychologia, 144: 107501.

3. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw, 67(1).

4. Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D (2021) Performance: An R package for assessment comparison and testing of statistical models. Journal of Open Source Software, 6.

5. Nakagawa S, Johnson PC, Schielzeth H (2017) The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effect models revisited and expanded. J R Soc Interface 14(134):20170213.

6. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24(3):127-135.

7. Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. J Mem Lang 68(3):255-278.

8. Matuschek H, Kliegl R, Vasishth S, Baayen H, Bates D (2017) Balancing type I error and power in linear mixed models. J Mem Lang 94:305-315.

9. Lenth RV (2021) Emmeans: Estimated marginal means, aka least-squares means. R package version 1.6.1.

10. Lockrow AW, Setton R, Spreng KAP, Sheldon S, Turner GR, Spreng RN (2021) Taking stock of the past: a psychometric evaluation of the Autobiographical Interview. bioRxiv https://doi.org/10.1101/2021.12.22.473803

11. Setton R, Sheldon S, Turner GR, Spreng RN (2022) Temporal pole volume is associated with episodic autobiographical memory in healthy older adults. Hippocampus 32(5):373-385.

12. Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17(3):143-155.

13. Jenkinson M, Bannister P, Brady M, Smith S (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17(2):825-841.

14. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62(2):782-790.

15. Poppenk J, Moscovitch M (2011) A hippocampal marker of recollection memory ability among healthy young adults: contributions of posterior and anterior segments. Neuron 72(6):931-937.

16. Svoboda E, McKinnon MC, Levine B (2006) The functional neuroanatomy of autobiographical memory: a meta-analysis. Neuropsychologia, 44(12):2189-2208.

17. Schacter DL, Addis DR, Buckner RL (2007) Remembering the past to imagine the future: the prospective brain. Nat Rev Neurosci 8(9):657-661.

18. Bonnici HM et al (2012) Detecting representations of recent and remote autobiographical memories in vmPFC and hippocampus. J Neurosci 32(47):16982-16991.
19. Oosterhof NN, Connolly AC, Haxby JV (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Front Neuroinform 10:27.
20. Kuznetsova A, Brockhoff PB, Christensen RH (2017) lmerTest package: Tests in linear mixed effects models. J Stat Softw 82(1):1-26.
21. Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013) Long-axis specialization of the human hippocampus. Trends Cogn Sci 17(5):230-240.
22. Brunec IK et al (2018) Multiple scales of representation along the hippocampal anteroposterior axis in humans. Curr Biol 28(13):2129-2135.
23. Poppenk J, Moscovitch M (2011) A hippocampal marker of recollection memory ability among healthy young adults: contributions of posterior and anterior segments. Neuron 72(6):931-937.
24. Gilboa A (2004) Autobiographical and episodic memory—one and the same?: Evidence from prefrontal activation in neuroimaging studies. Neuropsychologia 42(10):1336-1349.
25. Cabeza R, St. Jacques P (2007) Functional neuroimaging of autobiographical memory. Trends Cogn Sci 11(5):219-227.
26. Bonnici HM et al (2012) Detecting representations of recent and remote autobiographical memories in vmPFC and hippocampus. J Neurosci 32(47):16982-16991.
27. Svoboda E, McKinnon MC, Levine B (2006) The functional neuroanatomy of autobiographical memory: a meta-analysis. Neuropsychologia, 44(12):2189-2208.
28. Schacter DL, Addis DR, Buckner RL (2007) Remembering the past to imagine the future: the prospective brain. Nat Rev Neurosci 8(9):657-661.