Retention and transformation of internal experiences in autobiographical memory narratives

Haowen Su[1†], Mengting Zhang[1†], Coralie Knight[1], Mengqi Zhao[1], Hongmi Lee[1*]

[1]Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907, USA
*Correspondence: Hongmi Lee (hongmilee@purdue.edu)
[†]These authors contributed equally to this work.

SUPPLEMENTARY METHODS

Effects of content word count on detail retention

We used the word count of each detail as a proxy for informational richness. However, function words (e.g., articles, prepositions) may inflate this measure without contributing substantive content. Therefore, we recomputed word count after removing stop words using the Natural Language Toolkit (NLTK) corpus of English stop words. After stop words were removed, externally oriented details still contained significantly more content words than internally oriented details (externally oriented M = 4.71, SD = 1.17; internally oriented M = 3.98, SD = 1.75; $t(178)$ = 5.34, p < .001, 95% CI = [.46, .99], Cohen's d = .40). We then re-ran the mixed-effects logistic regression using the z-transformed content word count as a fixed effect, with participant as a random intercept. Separate models were fit for internally oriented and externally oriented details. The results remained unchanged: word count (excluding stop words) was not a significant predictor of internally-oriented detail retention ($\chi^2(1)$ = .81, p = .37), but continued to be a robust predictor of externally-oriented detail retention (β = .31, 95% CI = [.23, .39], $\chi^2(1)$ = 62.02, p < .001).

Effects of memory details included in the brief summary

In the Hippocorpus dataset[1], each participant generated a brief summary (two to three sentences) of the recalled event during the first recall session, which was later used as the memory cue in the second session. Because participants were instructed to include key details in this summary, it is possible that details mentioned in the summary could have influenced recall performance during the second recall. To account for this possibility, we conducted additional analyses of the brief summaries to assess and control for their potential effects.

We recruited one additional independent annotator to segment and code the summary responses for all 210 participants included in our study. Following the same instructions used for the main narrative coding, the annotator divided each summary into single-idea segments and categorized each segment into one of the five detail types (externally oriented, internally oriented, inferences about others, general comments, or other details). Participants' brief summaries contained an average of 32.8 words (*SD* = 8.93; range = 16–61) and 3.74 details (*SD* = 1.11). Importantly, the number of summary details was not significantly correlated with the number of details produced in the second recall ($r(208) = .05$, $p = .434$, 95% CI = [–.08, .19]), suggesting that summary length was not a critical factor influencing overall recall performance.

We next examined the distribution of detail categories included in the summaries. On average, participants' brief summaries consisted of 65.8% externally oriented details (*SD* = 27.6), 13.6% internally oriented details (*SD* = 19.8), 1.3% inferences about others (*SD* = 5.8), 18.1% general comments (*SD* = 23.0), and 1.2% other content (*SD* = 6.1). Consistent with the pattern observed in the full recall narratives, participants included significantly more externally oriented than internally oriented details in their summaries ($t(209) = 18.21$, $p < .001$, 95% CI = [.47, .58], Cohen's $d = 1.26$).

To control for the potential influence of differing proportions of externally and internally oriented details in the brief summaries, we re-ran the retention and addition analyses after excluding any memory details that overlapped with those included in the summaries. Because the summary sentences, by definition, often condensed the full narratives, manually identifying specific matching details between summaries and full recalls could be more effortful and more prone to error than matching details across the two full-narrative sessions. Therefore, to apply a more objective criterion, we used semantic similarity scores derived from the SBERT model[2] to identify overlapping content. For each participant, we computed the cosine similarity between the sentence-embedding vector of every summary detail and every detail from both the first and second recalls. Any detail with a similarity score of .40 or higher was excluded from subsequent analyses. In addition, when a first-recall detail exceeded the .40 threshold, the corresponding matching detail(s) in the second recall were also removed, and vice versa. The .40 threshold was chosen based on validation analyses comparing manually matched versus non-matched detail pairs across sessions (Supplementary Figure 1). Across participants, the average semantic similarity of non-matched pairs consistently fell below .40, whereas the majority of manually matched pairs showed similarity values above this level.

After excluding the details that overlapped with the summary content, the remaining counts were 12.60 details (*SD* = 6.85) for the first recall and 8.81 details (*SD* = 5.20) for the second recall. Among these remaining details, 17.8% (*SD* = 19.8) of all detail types reported in the first recall were retained in the second. Importantly, when externally oriented and internally oriented details were analyzed separately, internally oriented details ($M = 11.7\%$, $SD = 22.1$) were again significantly less likely to be retained than externally oriented details ($M = 24.1\%$, $SD = 28.7$), indicating that internally oriented experiences were more likely to be omitted over time ($t(133) = 5.03$, $p < .001$, 95% CI = [7.54, 17.30], Cohen's $d = 0.43$). Similarly, among the details newly introduced in the second recall, internally oriented details comprised a much smaller proportion ($M = 11.3\%$, $SD = 15.7$) than externally oriented details ($M = 52.1\%$, $SD = 27.8$; $t(207) = 16.61$, $p < .001$, 95% CI = [35.95, 45.63], Cohen's $d = 1.15$), again reflecting their lower overall prevalence. These results are qualitatively the same as those obtained in the original analysis conducted without excluding summary details, suggesting that presentation of the summary was not a critical driver of the higher retention and addition rates for externally oriented details.

Effects of delay between the first and second recalls on memory detail retention

Among the 210 participants included in the analysis, the delay between the first and second recall sessions varied substantially (Supplementary Figure 2). Therefore, we examined whether the length of this interval influenced the likelihood that memory details would be retained across sessions.

We first computed Pearson correlations between the delay interval (in days) and the proportion of memory details retained in the second recall, separately for internally oriented and externally oriented details. There was no significant correlation for either internally oriented details ($r(177) = .05$, $p = .50$, 95% CI = [-.10, .20]) or externally oriented details ($r(208) = –.01$, $p = .86$, 95% CI = [-.15, .12]). We also fit mixed-effects logistic regression models in which the delay (*z*-transformed) was included as a fixed-effect predictor and participant was included as a random intercept, predicting whether each individual memory detail was retained in the second recall. Separate models were constructed for internally and externally oriented details and compared with corresponding null models (containing only the participant random effect) using likelihood ratio tests. Again, adding delay did not significantly improve model fit for either internally oriented details ($\chi^2(1) = .14$, $p = .71$) or externally oriented details ($\chi^2(1) = .01$, $p = .92$), indicating that the time interval between sessions had little influence on whether a detail was recalled again.

Moreover, the results remained qualitatively unchanged even after excluding outlier and implausible delay values that were likely due to errors (i.e., delay = 0 days or > 200 days). There was no significant correlation between delay and the proportion of retained internally oriented details ($r(165) = .04$, $p = .60$, 95% CI = [-.11, .19]) or externally oriented details ($r(196) = .002$, $p = .98$, 95% CI = [-.14, .14]). Similarly, mixed-effects logistic regression showed that delay did not significantly predict retention of individual memory details for either internally oriented ($\chi^2(1) = .09$, $p = .77$) or externally oriented details ($\chi^2(1) = .16$, $p = .69$).

SUPPLEMENTARY TABLES

Supplementary Table 1. Hypothetical example of memory detail categorization provided to human annotators.

Detail	Category
I'm a dog person,	0
and I've been living with a lovely dog named Otis since last year.	0
It was a sunny day about a month ago,	1
and I thought it would be cool to have a picnic with my dog.	2
We picked a chill spot in the park,	1
and tossed a blanket under a shady tree.	1
I brought some dog treats,	1
and for me, just some regular human snacks.	1
As we sat there, my dog got all excited about the treats.	1
It was so nice to be there.	2
The sun peeking through the leaves, a light breeze, and my pup.	1
I couldn't help but think how awesome it was to just chill with my dog outside of the usual walks.	2
It felt like a special moment.	2
Definitely one of those days that sticks in your memory.	4
Since then, I've been having a picnic with my dog almost every week.	0

1 = specific objective external elements, 2 = specific subjective internal thoughts and feelings, 3 = speculations or inferences about others, 4 = general comments or evaluations, 0 = all other

Supplementary Table 2. Detail category coding and first-second recall matching of an example participant.

Recall session	Detail number	Memory detail	Category
First recall	1	I was hiking near Lake Ontario with my partner and our two grandkids.	Externally oriented
	2	It was a beautiful sunny day.	Externally oriented
	3	The lake sparkled brilliantly.	Externally oriented
	4	I have a bad knee, so I was struggling with some of the physical activity.	Externally oriented
	5	My partner suggested that I rest a bit on a fallen log.	Externally oriented
	6	I was nervous, because I would not be able to get up by myself,	Internally oriented
	7	but I agreed.	Externally oriented
	8	My partner and my granddaughter wanted to hike further to see the bluffs.	Externally oriented
	9	I said I would be OK for a bit,	Externally oriented
	10	but my sweet grandson insisted on staying with me.	Externally oriented
	11	He said, "I won't let my granny sit in the forest all alone."	Externally oriented
	12	Well it was a good thing.	General comment
	13	My partner and granddaughter didn't return in a reasonable amount of time.	Externally oriented
	14	We got very nervous!	Internally oriented
	15	My grandson is 10 years old, but not strong enough to help me up.	Externally oriented
	16	He searched for a stout walking stick and found one nearby.	Externally oriented
	17	I used it to prop myself up,	Externally oriented
	18	and managed to get my feet underneath me.	Externally oriented
	19	Together we went down the shore to find the rest of our party.	Externally oriented
	20	Luckily all was well,	General comment
	21	but I would truly have been distraught if I had been all alone waiting for so long!	General comment
Second recall	1	I was at a nearby nature center with my grandkids and my partner.	Externally oriented

	Added	This was an outing to enjoy the nice weather with my grandson and granddaughter,	Externally oriented
	Added	as well as a time to show them the beauty of Lake Ontario.	Externally oriented
	4	I am older and have minor mobility issues.	Other
	Added	When we got to the lake shore,	Externally oriented
	Added	the water was very wavy	Externally oriented
	Added	and the level was high	Externally oriented
	Added	so the strip of sand for walking was narrow.	Externally oriented
	5,7	I decided to rest	Externally oriented
	8	while the rest of the party headed up to the bluffs.	Externally oriented
	5	I sat down on a fallen log.	Externally oriented
	Added	My granddaughter was so joyous to see the lake,	Externally oriented
	Added	which she thought was as grand as the ocean.	Externally oriented
	Added	I assumed the three of them would continue on to the bluffs.	Internally oriented
	Added	Then the most surprising thing happened.	Internally oriented
	11	My 10 year old grandson said that he couldn't leave his granny there by herself,	Externally oriented
	10	and he chose to sit with me.	Externally oriented
	Added	We sat for a long time talking.	Externally oriented
	13	Afterwhile, we realized that the others should have returned, but they hadn't.	Externally oriented
	Added	Also my cellphone was dead,	Externally oriented
	6	and I wasn't able to stand up by myself.	Externally oriented
	12	It was so fortuitous that my grandson was there with me.	Internally oriented
	16	He found a stout stick	Externally oriented
	17,18	and helped me arise	Externally oriented
	19	so that we could find out what happened!	Externally oriented

Supplementary Table 3. Memory transformation metrics for Inferences about others, General comments, and Other details categories.

Metric	Inferences about others	General comments	Other
Proportion retained from first recall (%)	37.8 (43.0)	33.0 (37.7)	26.5 (25.4)
Proportion within added details in second recall (%)	2.4 (5.8)	8.2 (10.1)	24.9 (23.7)
Cross-session semantic similarity (cos)	.52 (.17)	.59 (.16)	.57 (.15)

Mean values across participants (SD in parentheses).

Supplementary Table 4. Features of memory details for Inferences about others, General comments, and Other categories generated during the first recall.

Feature	Inferences about others	General comments	Other
Emotional valence	.08 (.40)	.28 (.33)	.10 (.19)
Emotional intensity	.38 (.22)	.41 (.21)	.23 (.15)
Global semantic similarity	.26 (.07)	.25 (.07)	.20 (.05)
Local semantic similarity	.30 (.10)	.29 (.12)	.23 (.08)
Word count	8.6 (3.9)	10.6 (4.3)	10.4 (2.9)

Mean values across participants (SD in parentheses).

Supplementary Table 5. Mixed-effects logistic regression and likelihood ratio test results for predictors of memory detail retention.

Detail category	Feature	β	95% CI	$\chi^2(1)$	p (uncorrected)	p (FDR corrected)
Internally oriented	Emotional valence	.14	[-.03, .31]	2.77	.096	.122
	Emotional intensity	.20	[.03, .37]	5.67	.017	.030
	Global semantic similarity	.24	[.07, .40]	7.78	.005	.014
	Local semantic similarity	.27	[.10, .44]	10.21	.001	.008
	Word count	.09	[-.07, .25]	1.19	.275	.273
Externally oriented	Emotional valence	.02	[-.05, .10]	.40	.528	.527
	Emotional intensity	.09	[.01, .16]	5.23	.022	.028
	Global semantic similarity	.22	[.14, .31]	27.99	< .001	< .001
	Local semantic similarity	.16	[.08, .24]	16.37	< .001	< .001
	Word count	.28	[.20, .36]	49.72	< .001	< .001

Supplementary Table 6. Relationship between subjective memory importance and Inferences about others, General comments, and Other details categories.

Variable pair	Inferences about others		General comments		Other	
	r	*p*	*r*	*p*	*r*	*p*
Importance during first recall & Proportion during first recall	-.03	.640	.02	.789	.12	.074
Importance during second recall & Proportion during second recall	-.14	.042	-.02	.770	.20	.004
Second – first recall importance & Proportion retained from first recall	.11	.392	.01	.865	.05	.481
Second – first recall importance & Proportion within added details in second recall	-.14	.050	-.11	.116	.11	.127
Second – first recall importance & Cross-session semantic similarity	.02	.913	.02	.864	.05	.586

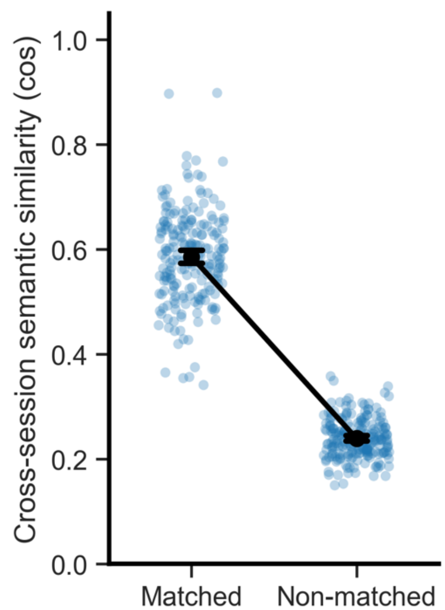
r = Pearson correlation coefficient between the pair of variables.
p = uncorrected two-tailed *p*-value.

Supplementary Table 7. Relationship between changes in subjective memory importance across recalls and the proportion of retained or added externally oriented and internally oriented details.

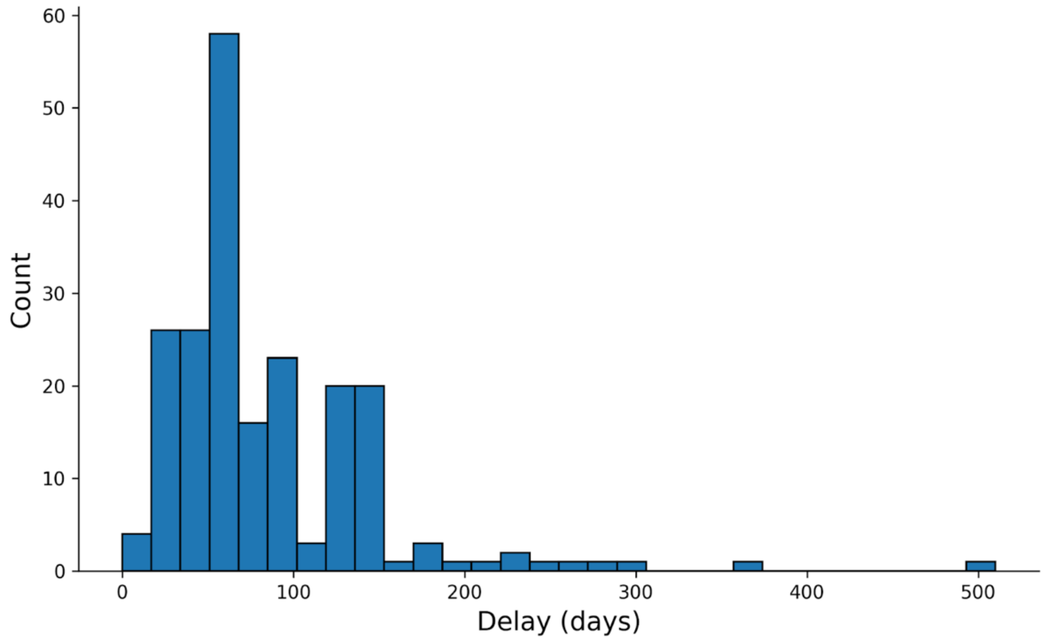
Variable pair	Externally oriented		Internally oriented	
	r	*p*	*r*	*p*
Second – first recall importance & Proportion retained from first recall	-.02	.792	.11	.145
Second – first recall importance & Proportion within added details in second recall	-.03	.708	.00	.963

r = Pearson correlation coefficient between the pair of variables.
p = uncorrected two-tailed *p*-value.

SUPPLEMENTARY FIGURES



Supplementary Figure 1. Mean cross-session semantic similarity between matched and non-matched memory details. For each participant, every memory detail was represented as a 768-dimensional embedding vector using the SBERT model[2]. We then computed cosine similarity between each detail from the first recall and its corresponding detail from the second recall. For matched details (as identified by human annotators), we calculated the mean cosine similarity across all matched pairs. If a single detail from the first recall corresponded to multiple sentences in the second recall, those sentences were concatenated into one string prior to embedding and similarity computation. For non-matched pairs, we exhaustively calculated cosine similarities between all possible first–second recall detail combinations not annotated as matches and averaged these values. The mean similarity was significantly higher for matched pairs (M = .59, SD = .09) than for non-matched pairs (M = .24, SD = .04; two-tailed paired-samples $t(209)$ = 59.59, p < .001, 95% CI = [.34, .36], Cohen's d = 4.11), indicating that the manual matching procedure effectively identified semantically similar memory detail pairs across sessions. Black circles indicate group means, colored dots represent individual participants, and error bars show 95% confidence intervals.



Supplementary Figure 2. Histogram showing the distribution of delays (in days) between the first and the second recall sessions (range = 0–510 days). In each recall session, participants estimated how long ago the reported event had occurred by entering a numerical value and selecting a unit (weeks or months). These responses were converted to days to approximate the event's age at each session. The delay interval was then computed as the difference between the event's estimated age at the first and second recalls. Unlikely values such as 0 days may have resulted from the coarse resolution of the response units or from errors in participants' self-reports.

SUPPLEMENTARY REFERENCES

1. Sap, M., Horvitz, E., Choi, Y., Smith, N. A. & Pennebaker, J. Recollection versus Imagination: Exploring Human Memory and Cognition via Neural Language Models. in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (eds Jurafsky, D., Chai, J., Schluter, N. & Tetreault, J.) 1970–1978 (Association for Computational Linguistics, Online, 2020). doi:10.18653/v1/2020.acl-main.178.
2. Reimers, N. & Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Preprint at https://doi.org/10.48550/arXiv.1908.10084 (2019).