PNAS
www.pnas.org

Supplementary Information for

Episodic memory enhancement versus impairment is determined by contextual similarity across events

Wouter R. Cox*, Simone Dobbelaar, Martijn Meeter, Merel Kindt, Vanessa A. van Ast*

*Email: W.R.Cox@uva.nl, V.A.vanAst@uva.nl

This PDF file includes:

Supplementary Methods
Supplementary Results
Figures S1 to S4
SI References

Supplementary Methods

Experiment 1: The Direction of Context-Induced Memory Change. *Stimuli.* Words. We obtained 90 nouns from a validated database of Dutch words (1). They consisted of three to ten letters, contained no more than three syllables, and were of neutral valence (range = 3 – 5.5, total scale = 1 – 7). All words were assembled into 30 AB, AC, and BC word pairs (i.e., 30 A words, paired with 30 B and 30 C words) that consisted of relatively low semantic interrelatedness as assessed by Latent Semantic Analysis (2) (range = 0 – 0.3, total scale = -1 – 1). The word pairs were assigned to 3 sets of each 10 AB/AC word pair combinations. Hence, the 3 sets contained 2 subsets of 10 AB and 10 corresponding AC word pairs (6 subsets in total). The average semantic interrelatedness scores of the AB, AC, and BC word pairs within sets were as low as possible (range = 0.04 – 0.06). Since individual word characteristics can severely impact associability among word pairs (e.g., ref. 3), the subsets were also matched on the number of characters, number of syllables, valence, and frequency of word pairs. Accordingly, the average scores on interrelatedness and all individual word characteristics did not differ between the six subsets, as assessed by separate univariate ANOVAs (all $F < 1.277$, all $P > 0.287$).

Experimental task versions. We created six versions of the experimental task that were divided over the participants to counterbalance the assignment of the three word pair sets to the within-subject conditions (same context, different context, original memory control). Every version contained different, randomized context word pair combinations. The AC word pairs of sets that were assigned to the AB control condition were not used in the experiment.

Experiment S1: Eliminating Carry-Over Effects of Original Memory and New Memory Test Phases on the Inference Test. To rule out effects of testing order on the inference scores, we repeated Experiment 1, but reversed the order of memory testing on day 3. Hence, participants completed the BC test (inference) first, followed by AC (new memory), and ended with the AB test (original memory). As the effect of Context on inference scores was large in Experiment 1 (*Cohen's dz* = 2.12), we tested fewer participants ($n = 20$, $1\text{-}\beta > 0.99$). Since we predicted the direction of the effects to replicate those observed in Experiment 1, one-sided tests were performed. All other procedures were the same as in Experiment 1.

Participants. Twenty-three participants consented to take part. One participant met one of the exclusion criteria (dyslexia diagnosis), and two participants did not complete the experiment due to a technical error. Hence, we analyzed data of 20 participants (17 females) with a mean age of 21 y (SD = 2.54, range = 18 – 26).

Experiment 2: The Nature of Context-Induced Memory Change. *Stimuli.* Words. To create a control condition for the new memories, we selected 30 additional words (1), and from scratch assembled all 120 words (i.e., the 90 words that were used before and the newly selected words) into 40 AB and 40 AC word pairs, according to the same criteria as in Experiment 1. Instead of 3 sets, we now created 4 sets that again contained 2 subsets of 10 AB and 10 AC word pairs each. The subsets were matched on the same characteristics as before (all $F < 0.585$, all $P > 0.766$).

Experimental task versions. We created four versions of the experimental task, such that each of the four word pair sets were assigned once to every within-subject condition (same context, different context, original memory control, new memory control). Again, every version of the task contained different, randomized context word pair combinations. The new memory word pairs of sets that were assigned to the original memory control condition and the original word pairs of sets that were assigned to the new memory control condition were not used in the experiment.

Procedure. Source memory test. Upon completion of the modified modified free recall test, participants performed a self-paced source memory test (4). They were informed that they would be presented with the cue word (A), together with the two target words they had recalled. It was stated that trials at which participants had filled in “xx” in both text boxes during recall would not reappear during the source memory test. They were explained that three tick boxes (day 1, day 2, do not know) would be displayed next to every recalled word to indicate on which day they had imagined the associates. To make sure that correct retrieval of words during the source memory test could be recorded, it was pointed out that two additional text boxes for optional use were displayed at every trial as well, such that they had the opportunity to type in associates that they could not remember during the recall test but came to mind while completing source memory testing (4). Participants were instructed to also indicate their source memory of such late recalls. Importantly, if participants had filled in two associates and knew the source of one (e.g., day 2), they could by implication infer the source of the other (day 1). To obtain actual measures of source memory, they were asked to refrain from using such deductive reasoning and to base their responses only on recollections of the day they encoded the recalled words. The order of trial presentation was the same as during the recall test (4).

Data analysis. To gain insight into context-induced changes of source memory, we calculated and analyzed the proportion of trials for which source memory was intact. However, this variable also comprises associative memory accuracy, since source memory can only be correct if the respective word is correctly retrieved. Therefore, we also calculated source memory accuracy proportional to retrieval accuracy. That is, we divided the number of trials with correct source memory (on which, by definition, recall was also correct) by the number of trials on which recall was correct (i.e., including trials with incorrect source memory). By doing so, we thus obtained a metric that solely reflects source memory accuracy.

Note that while completing source memory testing, participants were given the opportunity to fill in words that they could not remember during the earlier recall test. However, trials at which participants had filled in "xx" in both text boxes during recall did not reappear during the source memory test. Therefore, differences in the number of opportunities for these late recalls between conditions could have occurred. To ascertain that the retrieval results do not reflect such unequal response opportunities, we additionally analyzed accuracy scores that are solely based on responding during the initial recall test.

Supplementary Results

Experiment S1: Eliminating Carry-Over Effects of Original Memory and New Memory Test Phases on the Inference Test. Experiment S1 was designed to rule out that a carry-over effect from preceding original and new memory tests explains superior memory inference scores in the same context condition versus the different context condition in Experiment 1. The mean number of learning rounds to reach the criterion of 90% correct recall was 1.65 (SD = 0.59) on day 1, and 1.60 (SD = 0.60) on day 2. No differences in accuracy were found between the conditions during the day 1 or day 2 testing rounds (Fig. S2B). As in Experiment 1, we observed relatively enhanced inferential memory on day 3 when connected episodes took place in same contexts [same versus different, $z = 3.933$, $P < 0.001$, one-sided]. We also again found higher new memory accuracy in the same context versus different context condition [$z = 3.347$, $P = 0.001$, one-sided] (Fig. S1). For original memory accuracy we could not compare all conditions, because now inference tests preceded the original memory tests, and therefore the correct associates of the original memories (B) had already been provided as cues during inference tests (BC) in the same context and different context conditions, invalidating comparisons with control. A comparison between the same context and different context conditions, however, again showed relative enhancement in the same context condition [$z = 3.966$, $P < 0.001$, one-sided]. It must be noted though that performing inference tests can alter responses during subsequent original and new memory tests (5), such that these findings do not necessarily reflect true replications. Also, as Experiment 1 showed large differences in original and new memory accuracy between the same context and different context conditions, and intact original and new memories are required to perform inferences, BC performance does not necessarily reflect independent evidence for integration. However, Experiment S1 successfully ruled out an effect of testing order on inference scores.

Experiment 1, S1, 2, and 3: Immediate Effects of Contextual Stability on New Learning. As contexts are well-known for reinstating memories (6–10) and retrieval-mediated learning results in a strengthening of new memories that becomes apparent quickly (11), we explored whether accuracy of new memories was enhanced during encoding on day 2 in the same context condition in Experiment 1. We also tested the replicability of these findings in Experiment S1, 2, and 3.

Experiment 1. Context did not affect accuracy during the first or second testing round on day 1 [$\chi2_2 = 0.439$, $P = 0.803$; $\chi2_2 = 2.296$, $P = 0.317$, respectively] (Fig. S2A). However, we found significantly higher accuracy in the same context versus different context condition during the first testing round on day 2 [$z = 3.291$, $P = 0.001$]. No differences were observed on the second round of day 2 testing [same versus different, $z = 0.962$, $P = 0.336$]. These findings suggest that encoding new episodes in the same context as a related memory elicits retrieval-mediated learning that results in an immediate enhancement of new memories.

Experiment S1. No effect of Context on accuracy during any day 1 or day 2 round was found [all $P > 0.157$] (Fig. S2B). However, as the difference in accuracy between the same context and different context condition during the first testing round on day 2 was medium size in Experiment 1 (*Cohen's dz* = 0.58), there was relatively limited power to replicate this effect with 20 participants ($1\text{-}\beta = 0.67$).

Experiment 2. As Experiment 2 contained a bigger sample and, crucially, a new memory control condition was included, we could test whether the accuracy of new memories during day 2 testing was enhanced compared to baseline in the same context condition, as predicted by retrieval-mediated learning. Like in Experiment 1 and S1, no effect of Context was found on accuracy during the first and second testing rounds on day 1 [$\chi2_2 = 1.023$, $P = 0.600$; $\chi2_2 = 0.188$, $P = 0.911$, respectively] (Fig. S2C). Importantly, we observed a significant effect of Context on accuracy during the first testing round on day 2 [$\chi2_2 = 24.248$, $P < 0.001$], with a follow-up comparison showing memory enhancement in the same context condition compared to the different context condition again [$z = 3.836$, $P < 0.001$], but also compared to control [$z = 3.032$, $P = 0.002$]. No difference was found between the different context and control condition during the first round of day 2 testing [$z = -1.365$, $P = 0.172$]. Also, again no effect of Context was observed on accuracy of the second testing round on day 2 [$\chi 2_2 = 0.353$, $P = 0.838$]. Therefore, Experiment 2 confirmed that encoding new episodes in the same context as a related existing memory seems to elicit retrieval-mediated learning that leads to a rapid strengthening of new memories. Furthermore, encoding new episodes in a different context, interestingly, did not lead to an enhancement of new memories already during learning on day 2, but did induce strengthened recall one day later (see Main Manuscript), suggesting either that immediate and delayed effects of retrieval-mediated learning are not necessarily identical in strength or that a different process underlies enhancement of new episodes that are encoded in a different context.

Experiment 3. We did not find an effect of Context on accuracy during any day 1 or day 2 testing round [all $P > 0.287$] (Fig. S2D). Hence, despite large power ($1\text{-}\beta = 0.91$), no immediate effect of retrieval-mediated learning was observed in Experiment 3.

Experiment 2: Accuracy Scores Based on Responding during the Recall Test only. In an additional analysis we aimed to rule out that the observed effects on retrieval accuracy are due to differences in recall opportunities (conditions may have differed in the number of times "xx" was filled in both text boxes during recall and hence the number of options to include late recalls during source memory testing). Therefore, we analyzed accuracy scores that are solely based on responding during the recall test. The number of correct late recalls was infrequent (a total of ten trials across all participants). Furthermore, the results were virtually identical to the scores that included correct late recalls, for both AB accuracy [Context, $\chi2_2 = 62.573$, $P < 0.001$; same versus control, $z = 5.662$, $P < 0.001$, one-sided; different versus control, $z = -1.783$, $P = 0.038$, one-sided], and AC accuracy [Context, $\chi2_2 = 61.168$, $P < 0.001$; same versus control, $z = 5.667$, $P < 0.001$; different versus control, $z = 3.752$, $P < 0.001$]. Therefore, the results and interpretations are not influenced by a possible different number of recall opportunities between conditions.

Experiment 2: Context-Induced Changes in Source Memory. To uncover the influence of context on source memory accuracy, we explored (i) the proportion of trials that were correctly attributed to the corresponding day of encoding, and (ii) these same scores as a proportion of associate recall, a metric that more accurately reflects source memory (Fig. S3).

Source memory accuracy. We found a significant effect of Context on AB source memory accuracy [$\chi2_2 = 66.463$, $P < 0.001$], with follow-up comparisons showing source memory enhancement in the same context condition [same versus control, $z = 5.665$, $P < 0.001$; same versus different, $z = 5.683$, $P < 0.001$]. No difference between the different context and the control condition was observed [$z = -0.796$, $P = 0.426$]. In addition, we found strengthening of AC source memory in the same context condition [Context, $\chi2_2 = 70.084$, $P < 0.001$; same versus control, $z = 5.729$, $P < 0.001$; same versus different, $z = 5.627$, $P < 0.001$], and to a smaller degree in the different context condition as well [different versus control, $z = 4.116$, $P < 0.001$]. These results of source memory accuracy scores mostly mirror the results of the retrieval test (modified modified free recall). However, in these source memory scores that depend on recall accuracy (source memory can only be correct if recall is correct), no impairment of original memories in the different context condition was present. Hence, it seems that unlike memory recall (i.e., the central variable that we report in the main manuscript), source memory of original memories was not altered in the different context condition. However, as these source memory scores are not independent of memory retrieval, it is important to also analyze source memory accuracy relative to the number of correctly recalled words.

Source memory accuracy as proportion of associative memory. To more directly compare source memory accuracy of the recalled words between conditions, we analyzed ratios of correct source memory relative to the total number of recalled associates. A significant effect of Context on source memory of retrieved AB memories was observed [$\chi 2_2 = 18.860$, $P < 0.001$]. Follow-up comparisons revealed enhancement in the same context condition [same versus control, $z = 3.986$, $P < 0.001$, same versus different, $z = 2.871$, $P = 0.004$], but no difference between the different context and control condition [$z = 1.507$, $P = 0.132$]. Also, source memory of AC memories was significantly affected by Context [$\chi 2_2 = 28.695$, $P < 0.001$]. Again, an enhancement was observed in the same context condition [same versus control, $z = 4.704$, $P < 0.001$, same versus different, $z = 4.672$, $P = 0.004$], but also source memory in the different context condition was marginally strengthened [different versus control, $z = 1.811$, $P = 0.070$]. Interestingly, reminiscent of the asymmetrical pattern of memory misattributions in Experiment 1 (Fig. 3A), we found that the observed enhancement of source memory in the same context condition, compared to the different context condition, was marginally greater for AC memories than AB memories [$z = 1.720$, $P = 0.085$].

Together, these findings suggest that context not only modulates changes of memory retrieval, but also source memory of retrieved episodes. When encoding takes place in same contexts, the source memory of retrieved original and new episodes is markedly enhanced. Also when new episodes are learned in a different context, a slight strengthening of source memory seems to occur. For original memories that are encoded in different contexts than new learning, however, source memory is not altered, unlike retrieval that was impaired in this condition (see Main Manuscript).

SI Figures

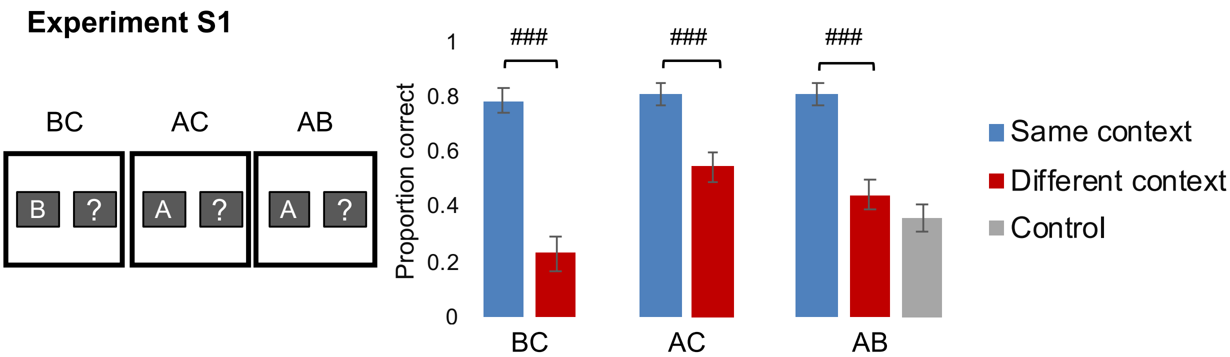

Fig. S1. Example trials of testing procedures (*Left*), and column chart of proportion correct recall per condition (*Right*) in Experiment S1 ($n = 20$). Error bars represent SEM. ###$P < 0.001$, one-sided.

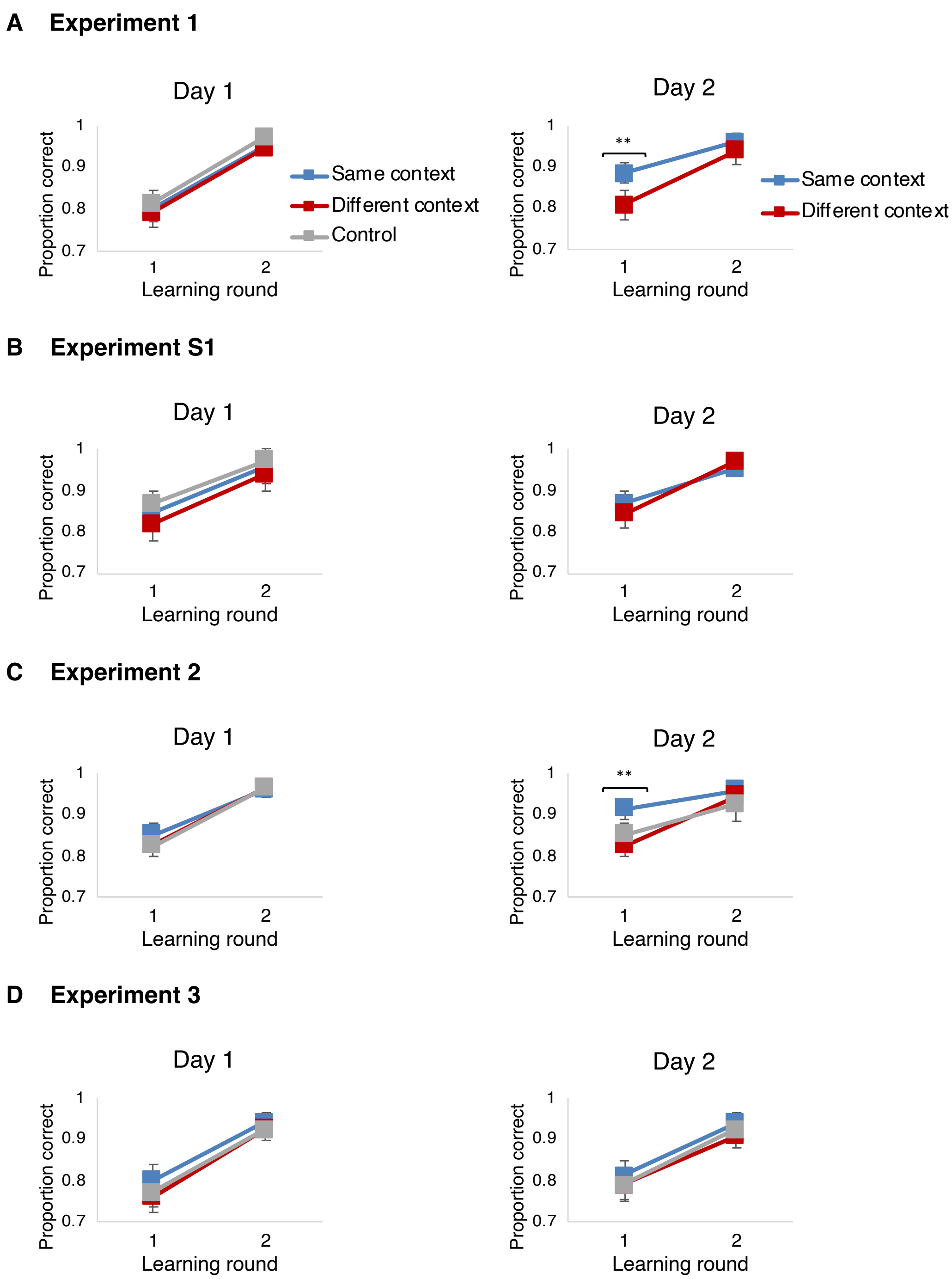

Fig. S2. Line chart of proportion correct recall per condition during day 1 and day 2 testing. (*A*) Accuracy of round 1 (n = 39) and round 2 (n = 23) on day 1, and round 1 (n = 39) and round 2 (n = 20) on day 2 in Experiment 1. (*B*) Accuracy of round 1 (n = 20) and round 2 (n = 12) on day 1, and round 1 (n = 20) and round 2 (n = 11) on day 2 in Experiment S1. (*C*) Accuracy of round 1 (n = 43) and round 2 (n = 29) on day 1, and round 1 (n = 43) and round 2 (n = 15) on day 2 in Experiment 2. (*D*) Accuracy of round 1 (n = 41) and round 2 (n = 25) on day 1, and round 1 (n = 41) and round 2 on day 2 (n = 20) in Experiment 3. Error bars represent SEM. $**P < 0.01$.

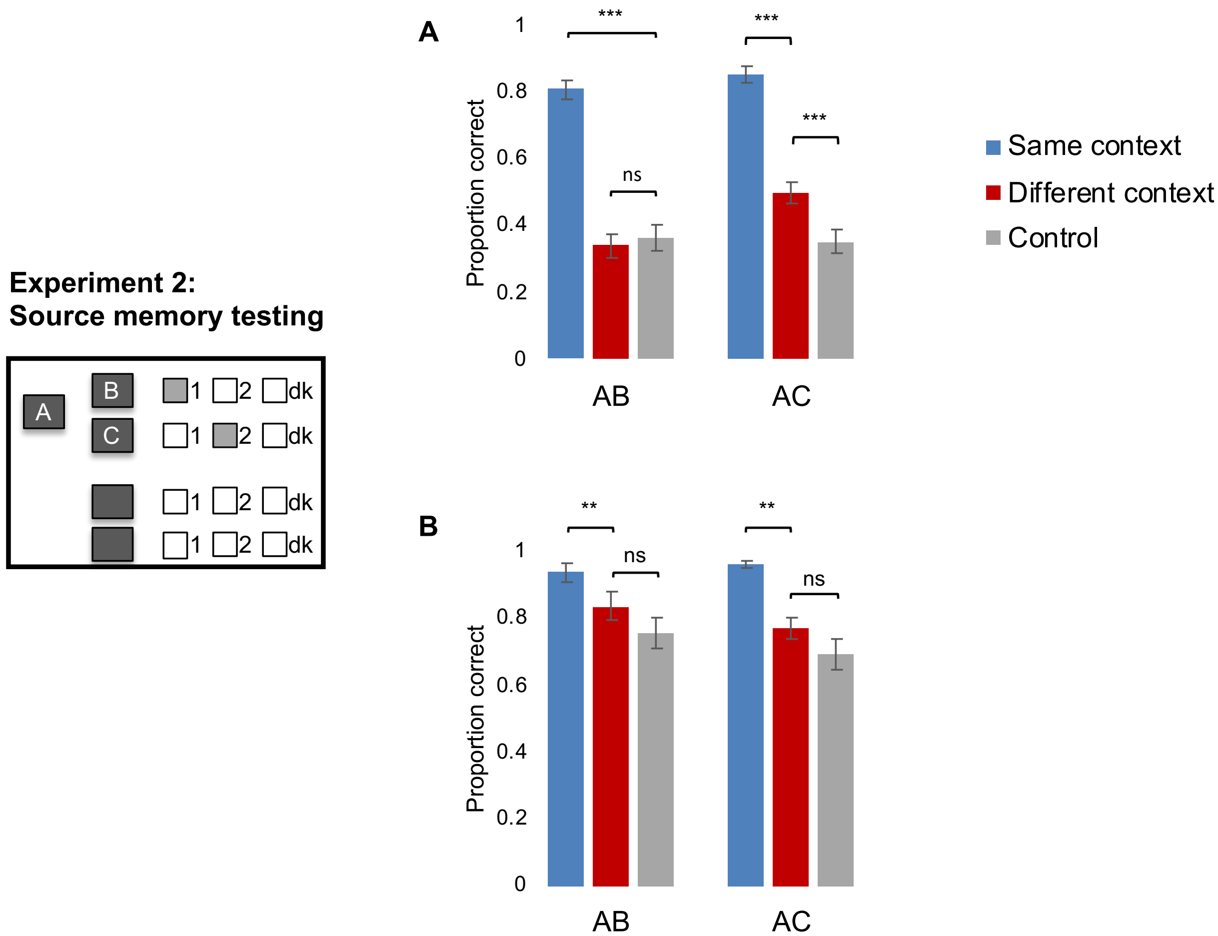

Fig. S3. Example trial of source memory testing in Experiment 2 ($n = 43$) (*Left*). Column charts of proportion correct source memory (*A*), and source memory accuracy as proportion of associative memory (*B*) (*Right*). Participants indicated whether, and if so which day, they remembered having imagined the associates by ticking one of three boxes (day 1, day 2, do not know/dk). Two additional text boxes for optional use were displayed at every trial as well, such that participants had the opportunity to type in associates that they could not remember during recall but came to mind while completing the source memory test (4). Participants were instructed to also indicate their source memory of such late recalls. Error bars represent SEM. ***$P < 0.001$, **$P < 0.01$, ns = not significant.

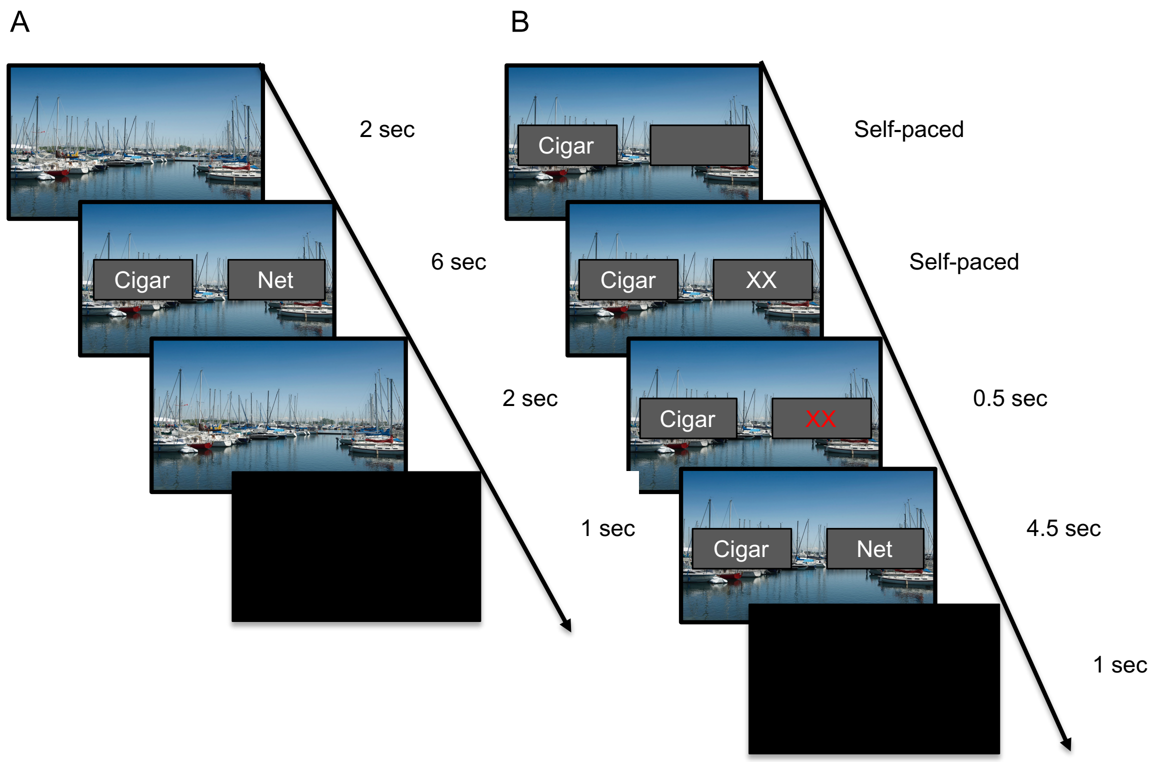

Fig. S4. Timing of learning and subsequent test trials. (*A*) Participants were shown a word combination in a unique context image and were instructed to come up with a vivid scene of the meaning of both words in the image. (*B*) During subsequent memory testing, participants were presented with the cue word on the context image and were asked to type in the correct associate. Participants were instructed to type in "xx" if they could not recall the correct answer. The response turned red when the answer was incorrect (displayed here) and green when the answer was correct. Afterwards, the word pair appeared again for participants to imagine in case of an incorrect answer. The complete list of word pairs was repeatedly tested until 90 percent of the associates were correctly recalled.

SI References

1. A. Moors, *et al.*, Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behav. Res. Methods* **45**, 169–177 (2013).
2. T. K. Landauer, P. W. Foltz, D. Laham, An introduction to latent semantic analysis. *Discourse Process.* **25**, 259–284 (1998).
3. B. H. Pierce, E. A. Kensinger, Effects of emotion on associative recognition: Valence and retention interval matter. *Emotion* **11**, 139–144 (2011).
4. J. M. Barnes, B. J. Underwood, "Fate" of fist-list associations in transfer theory. *J. Exp. Psychol.* **58**, 97–105 (1959).
5. A. C. Carpenter, D. L. Schacter, Flexible Retrieval: When true inferences produce false memories. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 335–349 (2017).
6. S. M. Smith, E. Vela, Environmental context-dependent memory: A review and meta-analysis. *Psychon. Bull. Rev.* **8**, 203–220 (2001).
7. V. A. van Ast, S. Cornelisse, M. Meeter, M. Joëls, M. Kindt, Time-dependent effects of cortisol on the contextualization of emotional memories. *Biol. Psychiatry* **74**, 809–816 (2013).
8. V. A. van Ast, S. Cornelisse, M. Meeter, M. Kindt, Cortisol mediates the effects of stress on the contextual dependency of memories. *Psychoneuroendocrinology* **41**, 97–110 (2014).
9. D. R. Godden, A. D. Baddeley, Context-dependent memory in two natural environments: On land and underwater. *Br. J. Psychol.* **66**, 325–331 (1975).
10. S. M. Smith, Remembering in and out of context. *J. Exp. Psychol. Hum. Learn. Mem.* **5**, 460–471 (1979).
11. M. L. Schlichting, A. R. Preston, Memory reactivation during rest supports upcoming learning of related content. *Proc. Natl. Acad. Sci.* **111**, 15845–15850 (2014).